UQAC

Université du Québec
à Chicoutimi

ROBUST AND EFFICIENT MESHFREE SOLID THERMO-MECHANICS

SIMULATION OF FRICTION STIR WELDING

Kirk Fraser, Eng.

Université du Québec à Chicoutimi, UQAC

Thesis presented to l'Université du Québec à Chicoutimi towards the grade of Doctor of Philosophy in Mechanical Engineering

Quebec, Canada

SUMMARY

Friction stir welding, FSW, is a solid-state joining method that is ideally suited for welding aluminum alloys. Welding of the aluminum is accomplished by way of a hardened steel tool that rotates and is pushed with great force into the work pieces. Friction between the tool and the aluminum causes heat to be generated, which softens the aluminum, rendering it easy to deform plastically. In recent years, the FSW process has steadily gained interest in various fabrication industries. However, wide spread acceptance has not yet been attained. Some of the main reasons for this are due to the complexity of the process and the capital cost to procure the required welding equipment and infrastructure. To date, little attention has been paid towards finding optimal process parameters that will increase the economic viability of the FSW process, thus offsetting the high initial investment most. In this research project, a robust and efficient numerical simulation code called SPHriction-3D is developed that can be used to find optimal FSW process parameters. The numerical method is meshfree, allowing for all of the phases of the FSW process to be simulated with a phenomenological approach. The dissertation starts with a focus on the current state of art. Next an in-depth development of the proposed meshfree formulation is presented. Then, the emphasis turns towards the presentation of various test cases along with experimental validation (the focus is on temperature, defects, and tool forces). The remainder of the thesis is dedicated to the development of a robust approach to find the optimal weld quality, and the associated tool rpm and advancing speed. The presented results are of engineering precision and are obtained with low calculation times (hours as opposed to days or weeks). This is possible, since the meshfree code is developed to run in parallel entirely on the GPU. The overall outcome is a cutting edge simulation approach for the entire FSW process.

Résumé

Le soudage par friction malaxage, SFM, est une méthode idéale pour relier ensemble des pièces en aluminium. Lors du procédé, un outil en acier très dur tourne à haute vitesse et est presser dans les plaques avec beaucoup de force. L'outil frotte sur les plaques et génère la chaleur, ce qui ramollie l'aluminium, ceci le rendant plus facile à déformé mécaniquement. Récemment, le SFM a connu une croissance de reconnaissance important, par contre, l'industrie ne l'as pas encore adopté unilatéralement. Il existe encore beaucoup de terrain à défricher avant de bien comprendre comment les paramètres du procédé font effet sur la qualité de la soudure. Dans ce travail, on présente une approche de simulation numérique sans maillage pour le SFM. Le code développé est capable de prendre en considération des grandes déformations plastiques, le ramollissement de l'aluminium avec la température, et la condition de frottement complexe. Cette méthode permet de simulé tous les phases du procédé SFM dans une seule modèle. La thèse commence avec un mis en contexte de l'état actuel de la simulation numérique du SFM. Une fois la méthodologie de simulation sans maillage présenté, la thèse concentre sur différents cas de vérification et validation. Finalement, un travail d'optimisation des paramètres du procédé est réalisé avec le code numérique. La méthode de simulation présentée s'agit d'une approche efficace et robuste, ce qui le rend un outil de conception valable pour les ingénieurs qui travaille dans le domaine de SFM.

This work was only possible due to the support and love of my wife, Laurie, my daughter, Kaïla and my son, Alexander.

They are my reason to live, love and continually improve

I dedicate this work to them

Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do.

Pele

Companies have too many experts who block innovation. True innovation really comes from perpendicular thinking.

Peter Diamandis

Effort only fully releases its reward after a person refuses to quit.

Napoleon Hill

ACKNOWLEDGEMENTS

Finally, and most importantly, I would like to thank my best friend for her boundless and unconditional support and love. Her incredible gift for mathematics and strategy were major driving forces to complete my undergraduate degree and to take on the challenges of a Ph.D. She is a beautiful person, an incredible mother, and the love of my life.

Saguenay, Quebec, April , 2017

Kirk Fraser, Eng.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

xvii

xxi

xxiii

xxvii

LIST OF SYMBOLS - ALPHANUMERIC

Symbol	Units	Meaning
$\bar{F}_N$	[N]	Normal contact force
$\bar{F}_T$	[N]	Tangential contact force
$\bar{F}_{ext}$	[N]	External force vector
$\bar{F}_{thread}$	[N]	Thread contact force
$\bar{Q}_j$	[m]	Contact point location on FEM element
$\tilde{V}_{SF}$	[-]	Velocity scaling factor
$\hat{n}_{FEM}$	[-]	FEM free surface element normal vector
$\hat{n}_{SPH}$	[-]	SPH free surface particle normal vector
$\bar{x}_{com}$	[m]	Center of ith particle cluster
h_{conv}	[Wm^{-2}K^{-1}]	Convection coefficient of heat transfer
h_{scale}	[-]	Smoothing length scaling factor
A_{JC}	[Nm^{-2}]	Initial yield (reference state) for Johnson-cook material model
A_c	[m^2]	Contact area
A_s	[m^{-2}]	Surface area
B_{JC}	[Nm^{-2}]	Strain hardening coefficient for Johnson-cook material model
C_{JC}	[Nm^{-2}]	Strain rate coefficient for Johnson-cook material model
C_p	[Jkg^{-1}K^{-1}]	Specific heat capacity
D_{CDF}	[-]	Cumulative damage factor for CDF model
E_T	[Nm^{-2}]	Tangent modulus

E_p	[Nm⁻²]	Plastic hardening modulus
$\bar{\bar{J}}$	[-]	Jacobian matrix
$\bar{\bar{M}}$	[-]	Kernel gradient modifying tensor
N_i	[-]	Number of j particles in the neighbourhood of i
$\bar{\bar{S}}$	[Nm⁻²]	Deviatoric stress tensor
T_∞	[K]	Ambient temperature
T_R	[K]	Room temperature
T_{melt}	[K]	Solidus (melt) temperature
T_{opt}	[K]	Optimal maximum welding temperature
T_{surr}	[K]	Surrounding temperature
W_{ij}	[m⁻³]	Smoothing function
a_1	[Nm⁻²]	Initial (reference) yield strength for FKS material model
a_2	[Nm⁻²]	First strain hardening coefficient for FKS material model
a_3	[Nm⁻²]	Second strain hardening coefficient for FKS material model
$\bar{b}$	[Nm⁻²]	Body force vector
b_1	[-]	First strain rate coefficient for FKS material model
b_2	[-]	Second strain rate coefficient for FKS material model
b_3	[-]	Third strain rate coefficient for FKS material model
c_1	[-]	First thermal softening coefficient for FKS material model
c_2	[-]	Second thermal softening coefficient for FKS material model
d_{wear}	[m]	Wear depth

m_{JC}	[-]	Thermal softening exponent for Johnson-Cook material model
n_{JC}	[-]	Strain hardening exponent for Johnson-Cook model
p_{thread}	[thread m^{-1}]	Thread pitch
$\bar{q}$	[Js^{-1}m^{-2}]	Heat flux vector
$\dot{q}$	[Js^{-1}m^{-3}]	Volumetric heat rate (heat source)
$\bar{v}$	[ms^{-1}]	Velocity
$\tilde{v}$	[ms^{-1}]	Relative velocity (ALE)
$\tilde{V}_{SF}$	[-]	Velocity scaling factor
$\breve{v}$	[ms^{-1}]	XSPH velocity average
$\bar{x}$	[m]	Position
$\tilde{x}$	[m]	Averages position using XSPH method
Δs	[m]	Initial particle spacing
h	[m]	Smoothing length
E	[Nm^{-2}]	Modulus of elasticity
G	[Nm^{-2}]	Shear modulus (second Lame parameter)
K	[Nm^{-2}]	Bulk modulus (isothermal)
L	[-]	Lagrangian
Q	[Js^{-1}]	Heat rate
R	[-]	Dimensionless smoothing distance
T	[K]	Temperature (in Kelvin for radiation)
V	[m³]	Volume
c	[ms^{-1}]	Speed of sound
f	[-]	General function

i	[-]	i^{th} particle
j	[-]	j^{th} particle
k	[Js^{-1}m^{-1}°C^{-1}]	Thermal conductivity
m	[kg]	Mass
p	[Nm^{-2}]	Hydrostatic pressure
r	[m]	Radial distance between two particles
t	[s]	Time

LIST OF SYMBOLS - GREEK

Symbol	Unit	Meaning
$\dot{\bar{\bar{\varepsilon}}}$	[s^{-1}]	Strain rate tensor
$\bar{\bar{\Omega}}$	[s^{-1}]	Spin tensor
α_d	[-]	Dimension specific constant for smoothing function
$\bar{\bar{\delta}}$	[-]	Kronecker delta
$\dot{\delta}$	[ms^{-1}]	Penetration rate in contact algorithm
δ_{CDF}	[-]	Stick-slip ratio for cumulative damage friction model
δ_{SR}	[-]	Stick-slip ratio
$\dot{\varepsilon}$	[s^{-1}]	Effective strain rate
$\bar{\bar{\varepsilon}}$	[-]	Strain tensor
ε_p	[-]	Effective plastic strain
μ_{comp}	[-]	Compression ratio
μ_{eff}	[Nm^{-2}s]	Effective non-Newtonian viscosity
$\bar{\bar{\sigma}}$	[Nm^{-2}]	Cauchy (total) stress tensor
σ_{SB}	[Js^{-1}K^{-4}]	Stefan-Boltzmann constant

σ_{eff}	[Nm-2]	Effective stress (flow stress) used in fluid type material models
σ^{trial}	[Nm-2]	Trial stress in radial return algorithm
σ_y	[Nm-2]	Yield stress
τ_y	[Nm-2]	Shear yield
τ_φ	[Nm-2]	Contact shear stress
$\psi_{T_{var}}$	[-]	Moving thermal couple temperature variation metric
ψ_{Tmax}	[-]	Maximum weld temperature metric
ψ_{defect}	[-]	Defect metric
ψ_{mixing}	[-]	Mixing metric
ψ_{wear}	[-]	Wear metric
$\psi_{weld\ quality}$	[-]	Overall weld quality metric
ϵ_r	[-]	Emissivity
α	[-]	Einstein notation index
β	[-]	Einstein notation index
δ	[m]	Penetration depth in contact algorithm
ζ	[-]	Natural element coordinates (FEM)
η	[-]	Natural element coordinates (FEM)
μ	[-]	Coefficient of friction
ν	[-]	Poisson's ratio
ξ	[-]	Natural element coordinates (FEM)
ρ	[kgm-3]	Density
γ_0	[-]	Gruneisen gamma

LIST OF ABBREVIATIONS

Abbreviations	Meaning
ALE	Arbitrary Lagrangian Eulerian method
BTFSW	Bobbin tool FSW
CFL	Courant-Friedrichs-Lewy number
CNRC-NRC	National research council
CPU	Central processing unit
CSM	Computational solid mechanics
CUDA	Compute unified device architecture
DRAM	Direct random access memory
DRFSW	Dual rotation FSW
EFG	Element free Galerkin
FDM	Finite difference method
FEM	Finite element method
FPM	Finite point method
FSSW	Friction stir spot welding
FSW	Friction stir welding
FVM	Finite volume method
GPU	Graphics processing unit
HAZ	Heat affected zone
Thermal camera	Infrared camera
JC	Johnson-Cook material model
LFW	Linear friction welding

MLPG	Meshless local Petrov-Galerkin
MPI	Message passing interface
MPM	Material point method
MTC	Moving thermo-couple
NEM	Natural element method
PFEM	Particle finite element method
PIC	Point in cell
RFSSW	Refill friction stir spot welding
RFW	Rotary friction welding
SMP	Shared memory parallel
SPH	Smoothed particle hydrodynamics
SSFSW	Stationary shoulder FSW
TC	Thermocouple
TMAZ	Thermo-mechanically affected zone
TWI	The Welding Institute (UK)
XSPH	Extended smoothed particle hydrodynamics

LIST OF APPENDICES

1 INTRODUCTION

Joining of two components of similar or dissimilar material is of fundamental importance for human society. One of the earliest documented forms of joining is "hafting" [1]. The process joins a blade to a
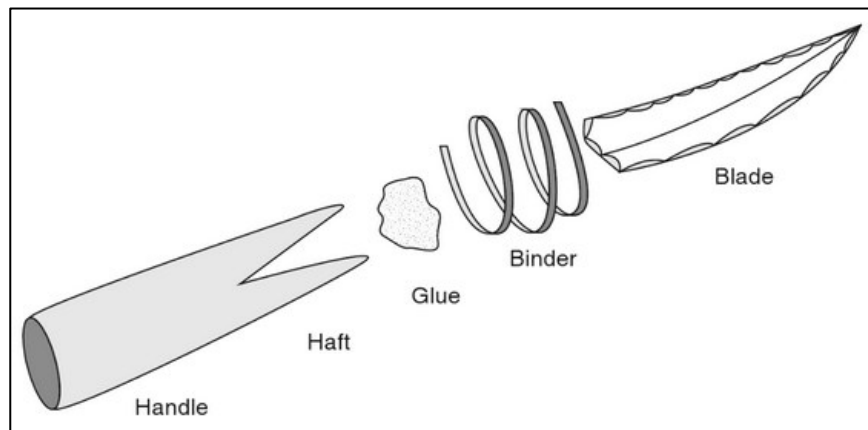
Figure 1-1 – A hafted tool [1]

handle with the use of a binder and adhesive as shown in Figure 1-1. Prior to that, tools would have been made from a single component such as a rock or stick. The advent of hafting ~500,000 years ago marked the start of the additive manufacturing industry. Since then, the variety of methods available to join one object to another has grown drastically. The main categories include fusion welding, fastening, adhesive bonding, and solid-state joining.

FSW (FSW) is a relatively new joining method (The Welding Institute, 1991) that has been steadily gaining appreciation in the automotive, aeronautical, and structural industries. The process is capable of producing full penetration welds in aluminum plates in a fraction of the time and with fewer defects compared to conventional MIG and TIG welding.

A hardened steel welding tool is used to form the weld as shown in Figure 1-2. The tool rotates and is forced (~1-10 tons of force) into the aluminum plates to be welded. Friction and plastic deformation cause the aluminum to heat up, making the material highly plastic and easy to deform. Once the material is hot enough (about 80% of the melting temperature), the tool starts to advance and joins the plates together.

A MACHINE-DRIVEN DIE PUSHES THE ROTARY TIP, OR PROBE, INTO THE METAL.

THE DIE'S ROTATIONAL AND LINEAR SPEEDS ARE CRITICAL TO THE PROCESS AND VARY DEPENDING ON THE METALS INVOLVED.

FRICTION SOFTENS THE TWO METALS ALONG THE WELD PATH INTO A DOUGHY STATE WITHOUT MELTING THEM.

THE SPINNING PROBE GENERATES HEAT AND MIXES THE METALS TOGETHER.
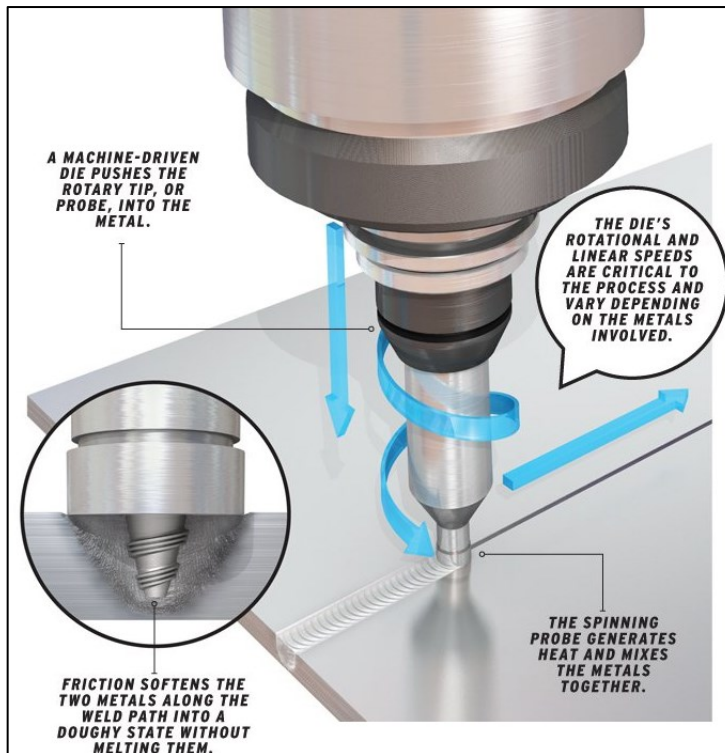
Figure 1-2 – FSW process [2]

Great care must be taken when selecting the process parameters. If the rpm of the tool is too large compared to the advancing speed, the weld temperature will surpass the melting temperature of the aluminum. This will cause unwanted defects and loss of weld strength. Certainly, the opposite is true as well, if the material does not heat up enough, the weld will not form properly due to insufficient material flow.

Choosing the optimal process parameters is a laborious undertaking. The most common approach is to perform welds at different rpm and advancing speeds to determine the best parameters. This approach is time consuming and costly. Since the process is highly dependent on the temperature in the weld zone, the ideal process parameters will differ for different plate thickness, material properties, and joint layout (butt, fillet, or lap).

Due to the computational complexity of the problem, the current state of art in numerical simulations of a single weld (one set of process parameters) typically requires many days or even weeks. This requires the use of specialized computing hardware such as large (very expensive) computing clusters. This technology is typically accessible to research institutions or companies that can afford (very) large simulation budgets with a dedicated team of simulation engineers.

1.1 PROBLEM STATEMENT

Optimization of FSW can be accomplished through many different methods. Currently, there is no fast, precise, and low cost procedure. In order to make FSW a more economically viable process for industries such as the structure industry, a method is needed to find optimal process parameters. The parameters should be chosen such that the weld can be completed as fast as possible (speed of tool advance) without causing any defects in the completed weld track.

The numerical model must be able to accurately predict temperature distribution, weld defects, and stress distributions during welding as well as residual stresses. For this reason, only a computational solid mechanics formulation is appropriate.

1.2 RESEARCH PROJECT OBJECTIVES

The general goal of the proposed research project is to develop a numerical approach to optimize the FSW process parameters. Optimized parameters will be found to provide the fastest rate of advance without causing weld defects. We are working towards providing a highly efficient and robust fully coupled thermo-mechanical simulation approach for FSW. Our numerical modelling tactic will use the GPUs (GPU) as the main computational components. The GPU is an inherently parallel processing unit that is economically feasible for any sized company. A typical GPU has thousands of processing cores for a minimal expenditure of around one to five thousand dollars. An equivalent CPU cluster would be in the range of half a million dollars.

The developed FSW simulation code will use a parallel implementation of a meshfree method known as smoothed particle hydrodynamics (SPH) for computational solid mechanics. The code will be written to function on any CUDA enabled NVIDIA GPUs. This will allow for finding optimal process parameters within a 24-hour period. No other research group could accomplish such a feat with the current state of art in simulation. In order to provide a phenomenological simulation tool for the entire FSW process, the following key objectives must be accomplished:

- Develop a meshfree computational solid code for the entire FSW process
- Parallelize the code on the GPU using CUDA Fortran

- Simulate FSW and validate the simulations from experiments

- Numerically predict defects, temperature, and residual stresses

- Develop a robust numerical simulation based optimization approach to determine the fastest welding speed without causing defects

The successful completion of these objectives will bring about a new and powerful way to analyze the FSW process. The work will provide invaluable insight into the physical mechanisms that cause certain defects. Only through knowledge of the cause, can we move towards cost-effective and viable solutions.

1.3 MAJOR DEVELOPMENTS IN THIS PROJECT

This research project has led to a significant number of important developments in the field of numerical simulation of FSW. However, the work has been carried out in a manner that ensures that the new approaches and algorithms are applicable to a wide range of engineering problems, not just to the FSW process. The following is a list of the major developments that have been completed:

1- Solid mechanics meshfree simulation code of the entire FSW process

2- Parallel programming implementation on the GPU makes the code highly efficient

3- New flow stress model for large plastic deformation processes with strain hardening, strain rate effects, and thermal softening

4- New friction model that uses a cumulative damage approach to dynamically change the interface friction

5- Full implicit smoothed particle method implementation for the cooling phase to improve performance. The implicit formulation solves a system of linear equations using either the conjugate gradient or bi-conjugate gradient stabilized methods. Both the mechanical and thermal formulation have been developed

6- Hybrid thermal-mechanical contact algorithm allows heat to be transferred efficiently from one body to another through the contact interface

7- Adaptive neighbor search specifically for large plastic deformation meshfree solid mechanics simulations

8- Adaptive thermal boundary conditions to improve the precision of the FSW simulations

9- Free surface extraction and triangulation algorithm for improved visualization of the simulation results. Surface triangulation is important for meshfree methods to be able to show the continuous contours

10- Phenomenological surface and internal defect prediction

11- Process parameter optimization using the fully coupled meshfree code along with response surface methodology and a novel constrained steepest descent approach

12- Extensive validation of the simulation code with experimental work

A great number of the aforementioned developments have led to improved precision in the simulation code. In every case, great care was taken to ensure that the developed algorithms are efficient and, in general, optimized for deployment on the GPU. Prior to the research project, the simulation of all the phases of the FSW process with a solid mechanics formulation in 3D was not feasible. Such a calculation would take many weeks or months to complete using a high performance-computing cluster. Now, such a model can be completed in a fraction of the time, leading to calculation times in the order of hours.

Certainly, a great number of minor developments were completed during the project. In some cases, these developments are discussed within the body of the thesis. However, a great number will not be described in order to keep this document concise and to the point.

1.4 THESIS LAYOUT

The thesis is presented in a sequential fashion; each chapter builds upon the previous chapters to provide a concise treatment of the methodology. In some cases, the reader will be referred to a latter section or an entry in the appendices. However, this will not be standard procedure. Every attempt will be made to ensure that all the required theory will have previously been established in earlier sections in order to prevent un-necessary "page turning".

Chapter 2 - Current State of the Art will discuss many different aspects of numerical simulation of the FSW process. The current state of art of numerical methods, material and friction modeling, optimization, as well as many other aspects will be discussed in this chapter

Chapter 3 focuses on laying out the framework for a meshfree simulation approach for FSW. First, the set of continuum mechanics equations that will be able to simulate the underlying physics of the FSW process will be outlined and explained. Next, the underlying principles of the meshfree method, smoothed particle hydrodynamics, will be introduced, and applied to discretize the continuum mechanics equations. The chapter will also outline a new flow stress and cumulative damage model specifically developed to simulate the entire FSW process. The implementation of the meshfree code will ultimately lead to a phenomenological approach to simulate the entire FSW process.

Chapter 4 *- Parallel Programming* concentrates on the use of the graphics-processing unit to improve drastically the performance of the meshfree code. The GPU is ideally suited for parallel programming of the SPH method. With this framework, the entire FSW process can be simulated in a fraction of the time than is possible in an equivalent CPU implementation.

Chapter 5 *- Modelling, Validation and Test Cases* will deal with the verification and validation of the meshfree code. Various test cases with experimental data will be used to validate results such as temperature history and profiles, spindle torque, forging force, and residual stresses, as well as internal and surface defects.

Chapter 6 - Process Parameter Optimization concentrates on minimizing defects, maximizing the weld quality, and finding the fastest advancing speed while maintaining an acceptable weld quality. This chapter represents the culmination of the research work. It is expected to be of significant interest to design engineers that are faced with determining the ideal process parameters for their application. The optimization approach is straightforward and presented in a parametric fashion that allows for the greatest level of applicability for the FSW community.

Chapter 7 - Conclusions and Outlook will wrap up the concepts and developments in the research project. The outlook for numerical simulation of FSW will be discussed along with the ultimate role of meshfree methods in the optimization of the welding process.

1.5 A NOTE ON NOTATION

This section is designed to be a brief explanation of the tensor notation used in this document. It is not intended to be an exhaustive presentation of tensor analysis. Continuous functions are written using tensor notation. For example, the equation of motion would be written as:

$$\frac{D\bar{v}}{Dt} = \frac{1}{\rho} \nabla \cdot \bar{\bar{\sigma}}$$

The material derivative is used since the numerical approach will be described from a Lagrangian frame of reference (for more details on the material derivative see Chapter 9 - Appendices). $\bar{v}$ is a velocity vector assumed to have Cartesian components $\bar{v} = (v_1, v_2, v_3)$ in three dimensional space. $\bar{\bar{\sigma}}$ is a stress tensor assumed to have components:

$$\bar{\bar{\sigma}} = \begin{bmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} \\ \sigma^{21} & \sigma^{22} & \sigma^{23} \\ \sigma^{31} & \sigma^{32} & \sigma^{33} \end{bmatrix}$$

Scalars are shown without any overbar (such as ρ in this example). This notation form is used to make an obvious distinction between a tensor of rank zero (scalar with no overbar), a tensor of rank one (vector, denoted with a single overbar), and a tensor of rank two (tensor, denoted with two overbars). This also provides a way to distinguish between continuum mechanics formulations (continuous field equations) and the discretized SPH formulations.

The indicial notation system will be used for all SPH equations. The system allows to more directly visualize the discrete interactions between the particle pairs. For example, the computational solid mechanics, CSM, SPH equation of motion would be written as:

$$\frac{dv_i{}^{\alpha}}{dt} = \sum_{j=1}^{N_i} m_j \left(\frac{\sigma_i{}^{\alpha\beta}}{\rho_i{}^2} + \frac{\sigma_j{}^{\alpha\beta}}{\rho_j{}^2} \right) \frac{\partial W_{ij}}{\partial x_i{}^{\beta}}$$

The subscript signifies that the variable exists at a discrete location; in SPH, these locations are thought of as particles. For example, the $v_i{}^{\alpha}$ field variable indicates that this is the velocity vector for the ith particle. The superscript follows the Einstein notation convention. For example, $\sigma_i{}^{\alpha\beta}$ is the stress tensor for the ith particle with components $\sigma_i{}^{\alpha\beta}$. In this notation system, scalars are shown without any

superscript, for example, density for the ith particle is written as ρ_i. The subscript letters i and j are strictly reserved for indicating particles. The superscript Greek letters α, β, and γ will be used for the indicial notation.

Often we will need to perform vector and tensor operations. A rather common operation in SPH is the composition of two tensors. For example in two dimensions tensor A can be composed with tensor B to give C, which is a tensor of rank two,

$$C^{\alpha\beta} = A^{\alpha\gamma}B^{\gamma\beta} \quad \text{is the same as} \quad \bar{\bar{C}} = \bar{\bar{A}} \cdot \bar{\bar{B}}$$

which, when written out fully is

$$\begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix} = \begin{bmatrix} A^{11}B^{11} + A^{12}B^{21} & A^{11}B^{12} + A^{12}B^{22} \\ A^{21}B^{11} + A^{22}B^{21} & A^{21}B^{12} + A^{22}B^{22} \end{bmatrix}$$

Another common operation is the inner product of tensors resulting in a scalar. For example, if we take the inner product of two rank-two tensors, say A and B, then we will get d, which is a scalar (tensor of rank zero):

$$d = A^{\alpha\beta}B^{\beta\alpha} \quad \text{is the same as} \quad d = \bar{\bar{A}} : \bar{\bar{B}}$$

Which, when written out in full for a two-dimension case, is:

$$d = A^{11}B^{11} + A^{12}B^{12} + A^{21}B^{21} + A^{22}B^{22}$$

if A and B are both symmetric, d is then:

$$d = A_{11}B_{11} + A_{22}B_{22} + 2(A_{12}B_{12})$$

In some situations, both the overbar and indicial notation are used. In all cases, every attempt is made to present the SPH equations with only the indicial notation. Many resources provide very in-depth presentations of vector and tensor analysis. The reader is directed towards Brannon [5, 6], Bonet and Wood [7], Fung [8], and Spencer [9] among a seemingly never-ending list.

2 CURRENT STATE OF THE ART

A number of different approaches can be used to join two parts. In general, the primary condition required to make a sound joint is to provide favorable conditions for the atoms to bond. To accomplish this, any impurities, dirt, or oxide must be forced out of the interface and the parts brought together with sufficient pressure to form a strong connection. In this chapter, the state of the art in solid joining, and specifically FSW, will be discussed. The various sections are:

Section 2.1 will present and discuss various solid state joining methods.

Section 2.2 introduces the FSW process in general, including an outline of different tools that are commonly used. The various phases will be discussed with a focus on the underlying physics.

Section 2.3 discusses variations on the FSW processes that have been developed to address specific issues such as defects, process efficiency, and robustness.

Section 2.4 introduces some of the common defects that can occur in a friction stir welded joint.

Section 2.5 focuses on the differences between a solid and a fluid formulation when treating the physics of the FSW process.

Section 2.6 will discuss the merit of various numerical methods for the simulation of the FSW process. Particular attention is paid to the three main classes of methods; grid based, meshfree requiring a background mesh, and meshfree not requiring a background mesh.

Section 2.7 explains a number of different material models that have commonly been used for FSW simulation. Details will be provided giving clear and concise reasoning as to why each method is not ideally suited for FSW.

Section 2.8 discusses one of the most important aspects of the FSW process; the friction model. In this section, a review of different friction modelling approaches will be presented and deliberated.

Section 2.9 will outline various optimization approaches that have been used to find best process parameters for a certain situation.

Section 2.10 provides a critical explanation as to why mass scaling should be avoided for meshfree methods.

2.1 SOLID-STATE JOINING PROCESSES

Solid-state joining processes form an important class of methods to join metallic and non-metallic parts. The main criterion for a process to be solid state is that the material does not melt (most commonly the solidus temperature is not surpassed). FSW is one of many other solid-state processes. In fact, FSW is one of the youngest and least mature methods. Linear and rotary friction welding date back to the 1950s. There are a great number of solid-state processes. In this section, linear and rotary friction welding as well as explosion welding will be discussed.

2.1.1 LINEAR FRICTION WELDING

In linear friction welding, LFW, the joint is made by pressing two parts together with a large force and rubbing them together with a reciprocating motion at high frequency. Figure 2-1 shows the various phases of the process. The initial phase involves the initial contact of the two parts; friction causes the parts to heat up and become easier to deform plastically. Once the part starts to get hot enough,
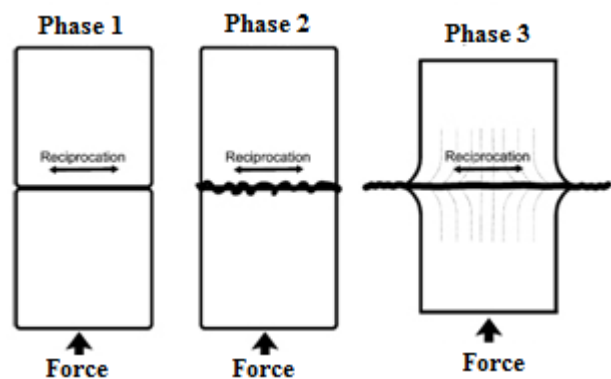
Figure 2-1 – Linear friction welding phases [4]

an increase in the forge force causes material from the contact interface to be expelled to the side of the joint. This effectively ejects all the surface contaminants and impurities to allow a strong solid-state joint to form.

Figure 2-2 – Turbine assembled by linear friction welding [4]

A number of different types of parts are joined using this process. It is a popular approach for the fabrication of bladed disks for turbine engines (Figure 2-2). In general, any shape can be joined as long as one of the parts can be held fixed and the other can be reciprocated. Because of the energy requirements and structural consideration for the welding machine, the reciprocating part should be the smaller of the two to be joined; creating a limit on the size of the parts that can be welded. Moving a large part at a high frequency would require a very powerful and large machine.

Numerical simulation of the LFW process, in many ways, follows closely the underlying physics of the FSW process. Various authors [10-14] have developed models capable of simulating the process using different approaches with focuses ranging from evaluating the material flow to the development of residual stresses.

2.1.2   ROTARY FRICTION WELDING
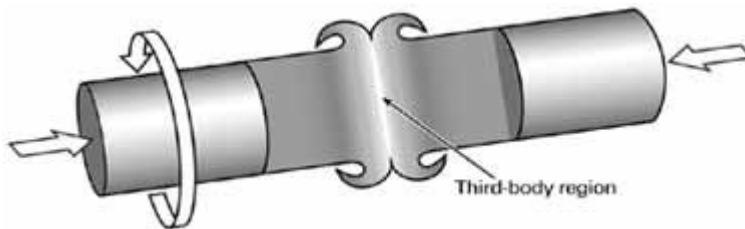
Figure 2-3 – Rotary friction welding [3]

Rotary friction welding, RFW, is akin to linear friction welding. The main difference is that instead of a reciprocating motion, the part is rotated at high rpm. Figure 2-3 gives an understanding of the process. One part is held fixed, while the other part is rotated. As a rule of thumb, the rotated part should have an axis of symmetry. There is no strict requirement that the fixed part be cylindrical (although this is commonly the case). RFW can be used to join many different types of parts; it is most commonly used for solid and hollow core shafts.

Schmicker *et al.* [15-17] have shown that the rotary friction welding process can be efficiently modeled using a 2D (actually 2-1/2D since twist degree of freedom is included) adaptive re-meshing approach. They use a fluid formulation with a modified Carreau constitutive law. The developed code uses a higher

order axisymmetric element formulation with mixed integration (isochoric and deviatoric stress integrated with different quadrature schemes). They are able to predict precisely the flash formation, process forces and torques, temperatures and the final dimensions of the welded part.

2.1.3 EXPLOSION WELDING

Explosion welding is an impressive process that involves forcing two plates together via the detonation of an explosive charge. The detonation forces the plates together with enormous pressure, causing a strong solid-state bond. Explosion welding is a robust process to join many different types of dissimilar materials that commonly cannot be joined (or at least, cannot be joined easily) by other methods. Figure 2-4 explains the process by which two plates are forced together. On the right side of the image is an example of the typical wave pattern that forms at the joint line between the two materials.
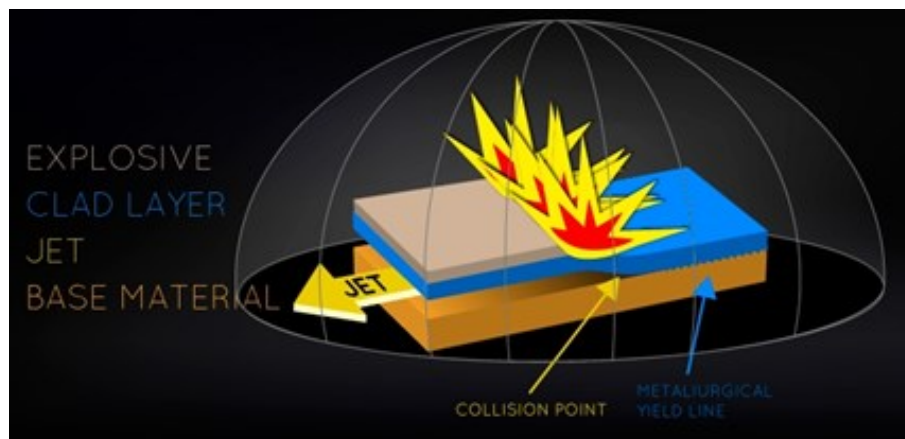
Figure 2-4 – Explosion welding process (adapted from [18])

Tanaka [19] has simulated the explosion welding process using smoothed particle hydrodynamics. He has shown that the simulation is able to capture precisely the wave formation phenomenon at the joint line. Other authors [20-23] have used other methods such as the material point method and finite element methods to simulate the process.

2.2 FSW PROCESS

Over the recent years, FSW has steadily grown in popularity. In fact, since the original patent was filed in 1991 by The Welding Institute, TWI, over 3000 related patents [24] have been filed, covering new tool

designs, process variations, and general improvements. One of the main roadblocks to widespread acceptance and use of FSW has ironically been due to the existence of the original patent. Because of this, the capital investment cost was very high and prevented small to medium sized companies from making the move to FSW. With the end of TWI's patent in 2015, one expects the process to continue to grow in popularity as the need for a company to hold a FSW licence is no longer in vigor. The FSW process is able to join many different alloys such as steel, titanium, aluminum, copper, magnesium, as well as certain non-metallic materials such as plastics, just to name a few. The range of materials that has been successfully joined with FSW is still growing and many exotic alloys can now be welded.

A schematic of the FSW process for a common butt joint is shown on the left side of Figure 2-5. The image shows a cylindrical hardened steel tool with a protruding pin. The underside of the tool (between the pin and the outer diameter of the tool) is called the tool shoulder. The two plates to be joined together are called the work pieces (WP). The two plates are classed according to the relative tool rotation. If the tool rotates clockwise (as indicated in Figure 2-5), the left and right plate will be on the advancing and retreating sides respectively.
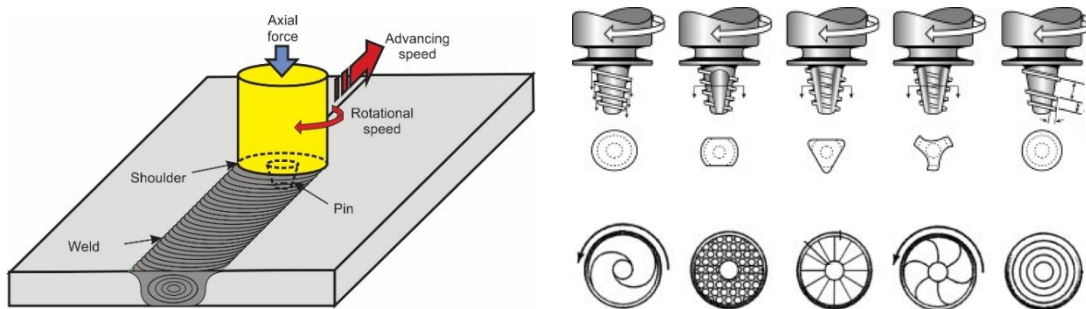
Figure 2-5 – FSW process parameters and tool geometry [25, 26]

The design of the tool is a diverse field and has a strong impact on the final weld quality. The shape of the pin can be a straight cylinder, tapered, triangular, or even three sided (as shown in the upper right image in Figure 2-5). Various features such as threads on the pin surface, scrolls, knurls, and protrusions on the shoulder surface can be added to the tool as depicted in the lower right image of Figure 2-5. The shape of the shoulder is also important, depending on the application; a flat, convex, or concave shoulder can be beneficial. Generally, the main parameters of the tool design are the shoulder diameter, $D_{shoulder}$,

the major, D_{pin}, and minor, d_{pin}, pin diameters (in the case of a tapered pin), and the pin length, l_{pin}. Lin *et al.* [27] have shown the effect of different pin thread designs. Bilici *et al.* [28] investigated parameters such as the pin length, pin angle, and dwell time in polyethylene sheets. Their results can generally be extended to metallic alloys.

The various phases of the process are shown in Figure 2-6. The main phases shown in the image are; 1-plunge, 2-dwell, 3-advance, and 4-retraction. From a general perspective, other phases can be identified prior to and following the four main phases.

Legend:
1 : plunge phase
2 : dwell phase
3 : welding phase
4 : retracting phase
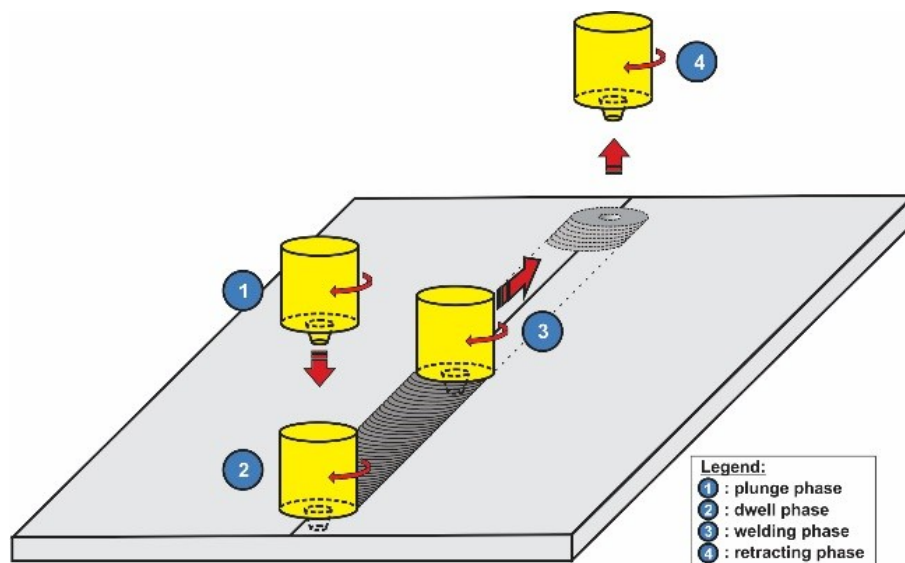
Figure 2-6 – FSW process phases [29]

The full chronology of a typical FSW joining process is then:

0a - Clamping

0b - Tool approach

0c - Initial tool contact

1 - Plunge

2 - Dwell

3 - Advance

4a - Tool retraction

4b - Un-clamping

4c - Cooling

The process history throughout all the phases is of importance for determining the residual stresses as well as the occurrence of defects. Certainly, the residual stresses will be strongly affected by the method by which the work pieces are supported and clamped [30]. In addition, the cooling method and, ultimately, the rate will have an important effect on the locked in stresses [31-35]. Ji *et al.* [32] investigated the effects of high cooling rates on the residual stresses and found that this was a viable option for reducing the residual stress in the joint. A description of the various phases will be provided in the subsequent paragraphs.

2.2.1   0A – CLAMPING

As the name suggests, the clamping phase involves securing the plates to prevent any unwanted movement. The manner in which the WPs are supported is of high importance. As the tool plunges and advances, very large forces and torques are transferred to the WPs. These forces must be resisted by an appropriate supporting structure. At the very least, a backing plate, also commonly called the anvil, is required (assuming the components to be welded are part of a larger assembly that is sufficiently restrained in the lateral and transverse directions). More commonly, clamps will be required to restrain the WPs to guard against the rotation or translation of the plates to be joined. Some authors [36, 37] have investigated self-supporting FSW processes and have shown the effect on the joint quality and the residual stresses.

2.2.2   0B - TOOL APPROACH

Once the plates are adequately supported, the tool is moved into position. This phase ensures that the weld will be started at the desired location. The rate and manner in which this phase is performed is essentially irrelevant to the overall finished weld quality.

2.2.3   0C - INITIAL TOOL CONTACT

After the tool is moved into position, the tool begins to descend into the WPs. Very high forces and torque will be developed at the onset of contact of the tool with the plates. Although the duration of this phase is very short, the generated forces can have a critical effect of the FSW machine and tool. If the initial contact is made too quickly, the forces can cause the process to stop. This is because most machines will have a load limit; once the limit is surpassed, the machine will stop to prevent unwanted damage to the welding apparatus.

2.2.4   1 - PLUNGE

Immediately following the initial contact, the important plunge phase is initialized. This phase is very important in the grand scheme of the process. The tool rotates and normal and tangential forces (friction) develop at the interface between the pin and the WPs. At the start of the plunge phase, the plates are cool. As the tool continues to interact with the plates, the material will start to heat up due to the conversion of friction work into heat. As the tool continues to plunge, the temperature will continue to rise. A drastic increase in the heating rate is commonly observed once the tool shoulder is exposed to the surface of the WPs. Various authors [38-42] have commented on the role of the pin and shoulder in relation to their respective heat input. Although the various research groups report different heat input ratios for the tool shoulder and pin, the consensus is that the shoulder is responsible for the majority of the heat input. Using the analytical heat generation equations of Durdanovic *et al.* [43], the heat generation ratios can be approximated for the shoulder, $\chi_{shoulder}$, the pin tip, $\chi_{pin\ tip}$, and the pin side surface, $\chi_{pin\ side}$. Commonly, the tool dimensions are such that the shoulder radius, $R_{shoulder}$, is twice the radius of the pin, r_{pin}, and the pin length, l_{pin}, is equal to half the shoulder radius (assuming a straight cylinder pin). This means that $R_{shoulder} = 2r_{pin} = 2l_{pin}$. For such a case, the heat input ratios are:

$$Q_{total} = Q_{shoulder} + Q_{pin\ tip} + Q_{pin\ side} \qquad \text{(2-1)}$$

$$\chi_{shoulder} = \frac{Q_{shoulder}}{Q_{total}} = \frac{14/3\ r_{pin}{}^3}{22/3\ r_{pin}{}^3} = 64\% \qquad \text{(2-2)}$$

$$\chi_{pin\ tip} = \frac{Q_{pin\ tip}}{Q_{total}} = \frac{2/3\ r_{pin}{}^3}{22/3\ r_{pin}{}^3} = 9\%$$

(2-3)

$$\chi_{pin\ side} = \frac{Q_{pin\ side}}{Q_{total}} = \frac{6/3\ r_{pin}{}^3}{22/3\ r_{pin}{}^3} = 27\%$$

(2-4)

Although the aforementioned analysis is basic, it nevertheless, provides an understanding of the heat input ratios. In practice, the effects of plastic deformation energy also contribute to the heat input. Furthermore, the influence of shoulder scrolls and pin threads will transport the material in a complex manner, shifting ratios depending on the situation.

The maximum force and torque is commonly attained [44] during the end of the plunge phase upon the instant that the shoulder makes full contact with the plates.

2.2.5 2 - DWELL

The plunge phase ends once the shoulder has made full contact with the WPs and the desired plunge depth is attained. During this phase, the tool remains stationary while continuing to rotate. The goal of this phase is to increase the temperature in the weld zone. The dwell phase should be long enough to attain a satisfactory temperature to allow the WPs to mix mechanically one within the other. Qian *et al.* [45] developed an analytical model to determine the optimal welding temperature for different alloys. For 6061-T6, the optimal temperature is given as ~0.84 of the solidus (443 °C). The model is based on providing a balance between the material flow and the maximum weld zone temperature. In this sense, the dwell phase should continue until this optimal temperature is attained or slightly surpassed.

The tool force and torque typically will decrease or at least stabilize during this phase since the tool is no longer actively pushing into the joint. In addition, since frictional heating continues, the material continues to soften, providing less resistance to the tool and machine.

2.2.6 3 - ADVANCE

The advance phase is where all the action occurs. The material has become hot enough and the tool starts to move forward (advance). In this phase, a common target temperature range will be 0.8-0.95 of the solidus to ensure that the WPs are soft and malleable. As long at the material is hot enough, the tool

will cause physical mixing of the two plate materials, forming a solid-state bond. An important parameter during this phase is the ratio of advancing speed to rpm. A general rule of thumb is that the ratio should be in the range of 0.2 to 0.8 mm/revolution [46]; however, this is not a hard rule. If the ratio is overly large, the weld will commonly contain defects. On the other hand, when the ratio is too small, the weld zone temperature will be too high, leading to a poor surface finish and an overly large, course grained microstructure [47-49]. Although the optimal ratio will be joint specific, FSW machines are becoming more sophisticated; some systems are equipped with temperature control units [50-52] that will alter the rpm and/or advance speed to attempt to keep an optimal temperature.

2.2.7 4 - TOOL RETRACTION, UN-CLAMPING, AND COOLING

When the FSW tool reaches the end of the weld, the tool must now be extracted from the joint. This process will leave a hole in the finished weld at the retraction site. In some cases, it may be beneficial to use a retractable pin tool (RPT) to mitigate the presence of the hole [53, 54]. The choice to un-clamp before or after cooling will depend on the application. However, the residual stresses are commonly greater when the plate cools in the clamped configuration. During the cooling phase, the stresses and distortions built up during the welding process will relax and re-distribute. The stresses during the welding process will be bound by the yield of the material at the specific strain rate and temperature. Upon full cooling, the final distortions and locked in stresses will be present. The stresses are primarily tensile in the weld zone and compressive in the rest of the plate (outside the heat-affected zone). For this reason, a FSW joint will mainly fail within (or on the edge of) the weld zone.

2.3 FSW VARIANTS

A number of variations on the standard FSW process have been proposed and successfully used over the past decades. As previously mentioned, in the standard FSW process, the tool is pressed into the work piece with enormous downward force (forging force); the material heats up due to plastic deformation and friction heating. The standard process cannot, however, be used to weld hollow core sections because a support is required underneath the work pieces. Bobbin tool FSW [55] consist of using a double-sided tool that squeezes the two work pieces (see Figure 2-7). The squeezing creates the required forging force in the material to cause plastic deformation. Since the tool is double sided, the

force at each side cancels out and the machine essentially feels a zero force along the axis of the tool rotation. The process is interesting since it allows hollow core sections to be welded with relative ease.
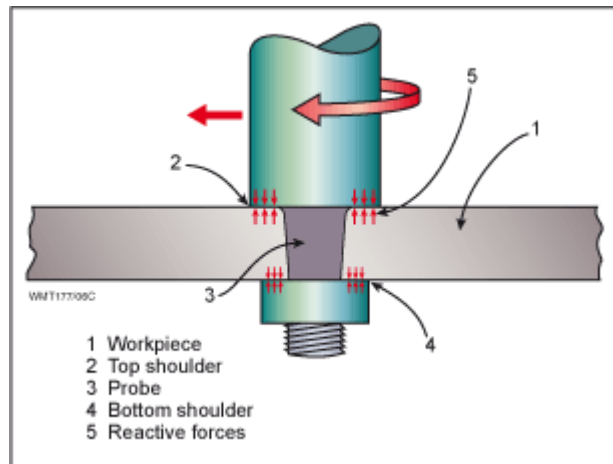
Figure 2-7 – Bobbin tool process [55]

Spot welding with the FSW process, commonly referred to as friction stir spot welding (FSSW) has also seen a lot of interest. The method involves only the plunge and dwell phases. A number of researchers [56-62] have focused on this variant to determine important process parameters such as plunge rate, rpm, and dwell time. The FSSW process will leave a hole when the tool is withdrawn from the WPs as shown in Figure 2-8. A slight variation of FSSW is the stitch process [63-65] where the tool is traversed a short distance to improve the overall strength of the joint. Another slight modification to the stitch process is the sweep process whereby the stitch is performed by causing the tool to arc out of the WPs. The process is interesting from an industrialization point of view due to the ease of implementation with a robot.
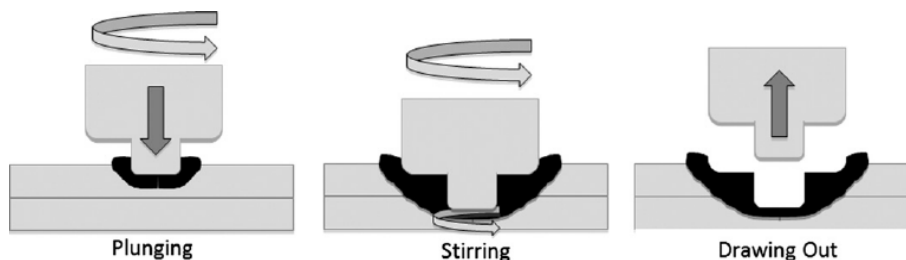
Figure 2-8 – FSSW process [66]

Recently, the refill FSSW (RFSSW) process was developed as an improvement to FSSW. In this approach, a composite tool is used with a retractable pin and collar as shown in Figure 2-9. Initially the plunge phase is as in FSSW and then the pin is retracted. This will allow material to flow into the recess created. Next, the collar is retracted while the pin is pushed back down. This will fill in the hole, leaving a stronger and improved fatigue-resistant joint. The process is relatively new and has been the object of study of a number of research groups [67-74].
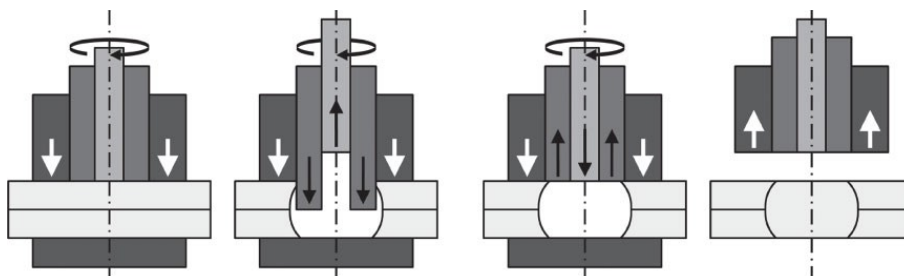
Figure 2-9 – Refill FSW process [75]

For applications where the surface finish is critical, the stationary shoulder FSW process (SSFSW) process has been developed. As the name suggests, the shoulder is prevented from rotating. In this sense, the pin will generate all (or almost all) of the heat required to bring the material to the plastic state. The process has significant differences in terms of the thermal profile; and as such, the resulting microstructure [76-82]. In order to allow the pin to rotate free of the shoulder, a complex tool is required with a system of thrust and radial bearings. A support method is required to prevent the shoulder from turning.

The dual-rotation FSW (DRFSW) process has been investigated by TWI [83]. Here, the pin and shoulder can rotate at different rates or even in opposing directions. The goal of the process is to have a more uniform temperature distribution in the weld zone, leading to improved welding conditions and ultimately improved joint strength. The process is very interesting, although the tool design and machine requirements lead to a more costly and complicated process than conventional FSW. The material motion is complex during the process and has been studied numerically by a number of authors [84-88]. These authors focus on the opposing rotation case.

A process that was developed to reduce the overall force on the FSW tool and improve the weld quality is Twin Stir[TM]. The method uses two or more tools with the same or opposing rotation. Different variations of this process have been successfully used to mitigate or eliminate various defects commonly [89] found in the FSW process. Hook formation, plate thinning, oxide clumping, and macroscopic porosity defects due to insufficient contact between the plates in the weld zone can be avoided.
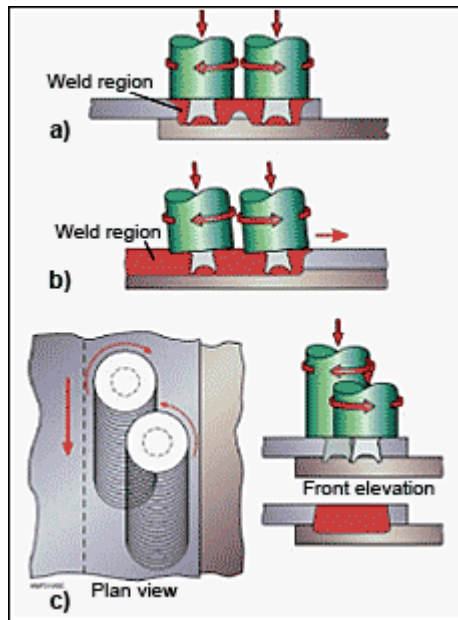
Figure 2-10 – Twin Stir[TM] process [89]

2.4 COMMON FSW DEFECTS

When the welding parameters are not within a certain operating range, weld defects will result. A common defect in FSW is an incomplete penetration defect. This is most often due to a pin depth that is not sufficient to deform plastically the lower region of the work piece joint. This can also be due to a misalignment of the tool center relative to the weld joint. A few of the more common defects for linear butt welds are shown in Figure 2-11 (Leonard and Lochyer [90]). In the upper left corner is a lazy S or kissing bond defect. This is commonly due to incomplete mixing of the two workpieces. In the upper right corner, a weld with excessive flash defect is shown. This results when too much axial (downward) pressure is applied to the tool. The material plastically flows past the shoulder of the tool and forms a periodic wave like pattern.

This upper right image also shows the telltale FSW groove pattern and particle deposits left behind in the weld track. As the tool rotates, it 'extrudes' a layer of material; the extruded material then forms a series of bunched up ridges (grooves). The grooves contain smaller striations that are due to the formation of work-hardened particles that are scraped along between the underside of the FSW tool and the surface of the work pieces. Figure 2-12 shows a macrograph of a FSW weld. The primary grooves and secondary striations are evident in the image.

The image on the bottom left shows the formation of a wormhole (incomplete weld). This can be caused by a number of problems such as insufficient downward pressure, a gap between the work pieces, or a sub-optimal ratio of the advancing speed to the rpm.

The bottom right image shows work pieces that we not properly butted up together (leaving a gap between the two plates). This is situation arises if the support structure and clamps are not well designed for the welding process.
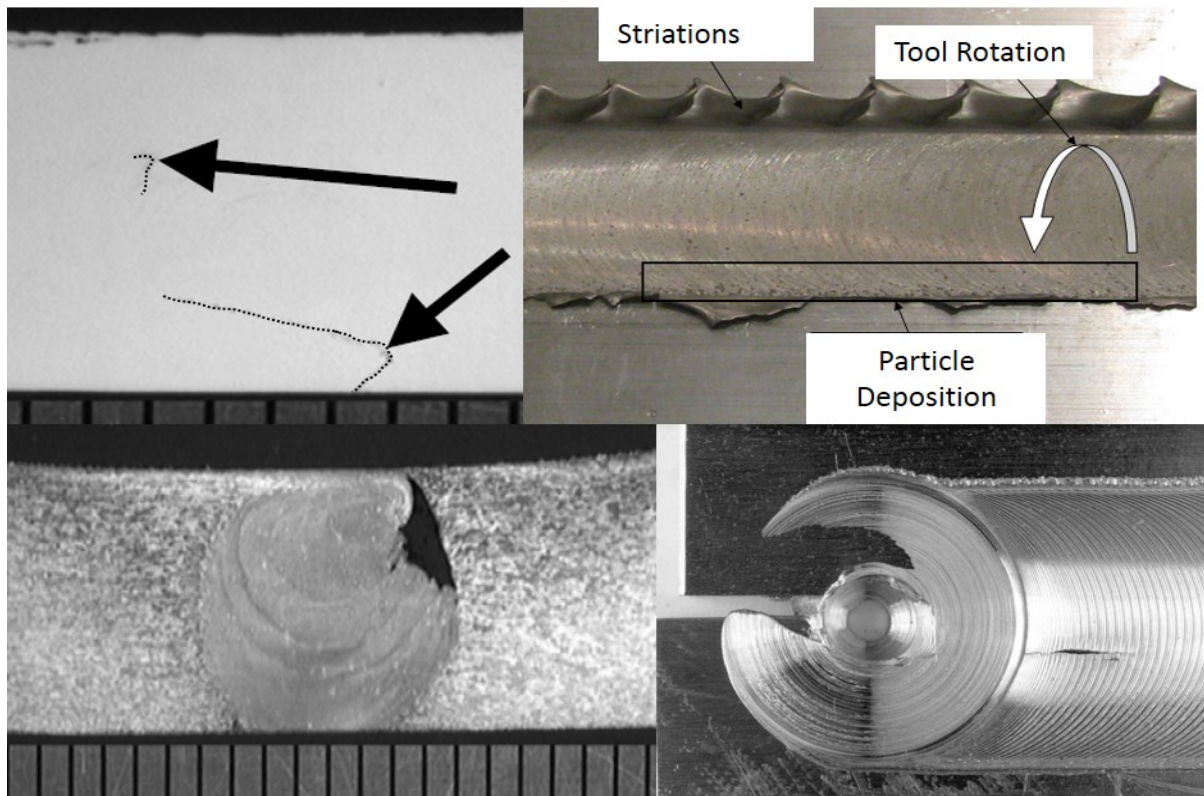
Figure 2-11 – Common defects in FSW welds – Top left: Lazy S or kissing bond; top right: Flash and particle deposits; bottom left: Wormhole; bottom right: Surface defect
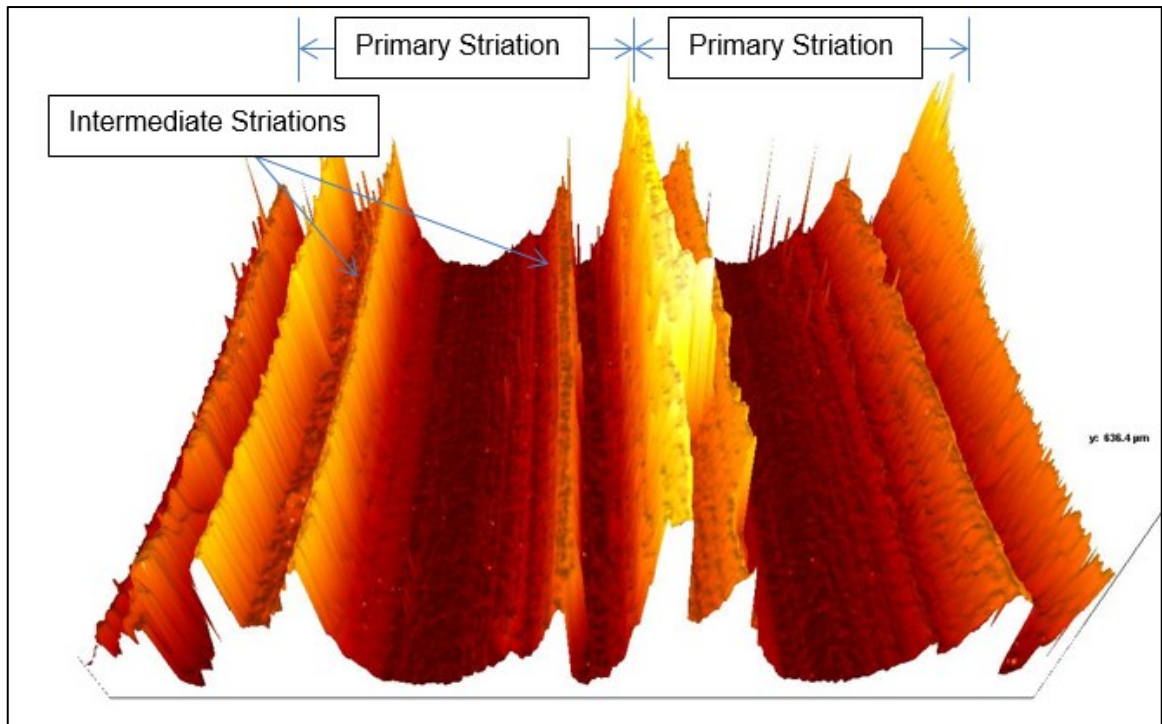
Figure 2-12 – Macrograph of groove pattern in FSW weld track

2.5   FLUID OR SOLID FORMULATION

Because of the large plastic deformation that occurs during FSW, many research groups [91-99] have used a fluid approach to simulate the process. One of the main assumptions becomes that the elastic strains are negligible. Most often, the work pieces will be modeled as non-Newtonian fluids using an effective viscosity that is a function of temperature and strain rate. Using a fluid approach is less computationally expensive than a solid approach since the limit on the time step size can be many orders of magnitude higher (with an explicit temporal integration scheme). In the solid approach, the time step is governed by the speed of propagation of a stress wave across an element. The time step must be small enough to resolve the stress wave within each element. If the time step is overly large, the simulation will become unstable and will eventually diverge. On the other hand, the solid mechanics formulation permits the resolution of elastic stresses, leading to an ability to predict defects, residual distortions and stresses.

Bussetta *et al.* [100] compare the fluid and solid approach for simulating the FSW process. They use the arbitrary Lagrangian Eulerian method. They find that the results are very similar between the two

approaches. They did not endeavor to evaluate the residual stresses or distortions, nor was their model able to predict the formation of defects. Buffa *et al.* [101] use a hybrid approach where they simulate the welding phase using a fluid model and the cooling phase using a solid model. They propose that the residual stresses can be found based on the final state of the weld phase. Certainly, the history of the material deformation cannot be captured during the weld phase. This does not lead to a robust approach to determine the residual stresses. Riahi and Nazari used a decoupled approach where the temperature field is calculated based on analytical heat source models, and then the temperature field at the end of the simulation is used to drive a cooling calculation in a FEM model. They obtain good correlation to experimental results for the parameters studied. However, if they were to attempt parameters likely to lead to internal or surface defects, their approach would not be able to predict the residual stresses accurately. Carlone *et al.* [102] use a novel coupling between FEM and the dual boundary element method to predict the residual stress distribution and subsequent crack propagation under mechanical loading of the joint. They also use an analytical heat source model during the welding phase. They impose the crack geometry in the model leading to a non-physical (not phenomenological) approach.

To date, very few researchers have attempted a fully coupled phenomenological simulation approach to model the entire FSW process. Such a modelling procedure would be robust as it could predict the final weld quality for a wide range of process parameters, even those that lead to internal or surface defects. Clearly, such a model is lacking in the FSW community. For this reason, such a model is one of the focuses of this research work.

2.6 NUMERICAL METHODS

Solving the set of partial differential equations needed to define the underlying physics of FSW is only possible using an appropriate numerical method. The use of analytical solutions may be possible for very simple situations. However, for industrial applications, the use of an analytical approach is not likely in the foreseeable future. Numerical methods considered here will be grouped into mesh-based methods, meshfree methods requiring a background mesh, and true meshfree methods.

2.6.1   MESH (GRID) BASED METHODS

Mesh (or grid) based methods form the most popular collection of approaches to solve field equations. The three most common methods are the finite element method (FEM, possibly originated by Hrennikoff in 1941 [103], although it is difficult to assess the true origin), the finite difference method (FDM), and the finite volume method (FVM). The underlying principle is the same in each of the methods: represent a continuous body (or domain) by a collection of smaller domains, more commonly called elements. Both FDM and FVM are developed based on the Eulerian frame of reference (without advanced treatment), whereas FEM can be formulated either from an Eulerian or Lagrangian reference.

FDM has been used by various research groups to simulate the temperature field during the FSW process. Certainly, FDM could be used for material flow prediction as well; however, this is less common since the boundary conditions are more complicated to represent the work piece free surface and the interaction with the tool. FVM is more commonly used for FSW simulation including material flow. More often, researchers employing the FVM [93, 95, 98, 99], focus on how the material is transported and on the resulting temperature distribution. Hasan *et al.* [94] have studied the effect of worn tool threads on the material flow patterns. Li *et al.* [96] have focused on the optimal design of threads. Although computational fluids approaches using the Eulerian frame abound, this approach is less suited for predicting the location and nature of defects within the weld zone. Since the mesh is fixed in such approaches, the evolution of the free surface cannot be tracked.

Since FEM can be formulated from either frame of reference, this approach is very popular for simulating FSW. If a Lagrangian frame is to be used, complex adaptive re-meshing or moving mesh approaches are required. The reason for such advanced tactics can be explained by the mapping from the geometric to the computational domain. In FEM, volume integration using a hexahedral element of a function, f, can be performed using Gaussian quadrature (among other schemes) according to:

$$\int_V f dV = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} f|\bar{\bar{J}}|d\xi d\eta d\zeta \tag{2-5}$$

As a finite element starts to distort, the Jacobian, $\bar{\bar{J}} = J^{\alpha\beta}$, of the element is affected:

$$J^{\alpha\beta} = \frac{\partial x^\beta}{\partial \xi^\alpha} , \xi^1 = \xi, \; \xi^2 = \eta, \xi^3 = \zeta \qquad (2\text{-}6)$$

As the element undergoes deformation, the determinant of the Jacobian will deviate from 1.0, leading to numerical errors in the quadrature. For example, a perfectly cuboid hexahedral element will have $|\bar{\bar{J}}| = 1.0$ and will give precise results in the volume integration. The general goal of adaptive re-meshing is then to have well shaped elements that will provide improved mapping from the geometry to the calculation domain (maximizing the value of the determinant of the Jacobian).

Schmicker *et al.* [15-17] have developed an advanced computational framework for simulating the RFW process. Their computational method uses higher order finite elements with an implicit geometry re-meshing approach. Due to the inherent axial symmetry in RFW, they are able to solve a so-called "two and a half" dimensional model. The discretization is only in a plane and significantly improves the calculation time of this computationally heavy method. Certain researchers [16, 17, 104-106] have worked on simulating the FSW process in 2D using adaptive meshing. However, 2D models are limited in their applicability to FSW. Other research groups have worked on moving mesh approaches that combine the Eulerian and Lagrangian frames into the numerical method. Such moving mesh methods fall into the category of arbitrary-Lagrangian-Eulerian (ALE) approach. The underlying principle of ALE is to use a relative velocity in the conservation equations to take into account the velocity of the moving mesh, $\bar{v}_{mesh}$. For example the ALE form of conservation of momentum would be:

$$\frac{\partial \bar{v}}{\partial t} + (\tilde{v} \cdot \nabla)\bar{v} = \frac{1}{\rho}\nabla \cdot \bar{\bar{\sigma}} \qquad (2\text{-}7)$$

The convective derivative now involves the relative velocity, $\tilde{v} = \bar{v} - \bar{v}_{mesh}$. By this means, large plastic deformation can be accounted for.

Various groups [101, 105-110] have used moving mesh techniques to simulate the various phases of FSW. One of the major drawbacks of this approach is the very high computational cost. In addition, the method makes it difficult to obtain good performance from common parallel programming models. Not all authors provide information on the required calculation time; however, Grujicic *et al.* [111] have noted that their fully coupled thermomechanical ALE implementation takes on the order of weeks to solve a

single model. One of the other issues with ALE is that the method is highly diffusive because of the means in which material is "advected" from one element to another. As long as the material flow is perpendicular to an element face, the method works well; however, when the material flow is through the corner, the advection is taken as a weighting of the flow in the direction normal to the associated element faces. This can lead to a loss of material and is the source of loss of precision.

2.6.2   MESHFREE METHODS REQUIRING A BACKGROUND MESH

Meshfree methods continue to grow in popularity among the simulation community due to their ability to simulate a wide range of engineering problems. Typically, in a mesh based method (such as the finite element method, FEM) a great amount of the simulation process is spent by the analyst creating a high quality mesh. However, meshfree methods remove the often cumbersome meshing task, leaving more time for the engineer to optimize their design. There are two main types of meshfree methods. In the first type, the set of governing partial differential equations (PDE) are weakened to form a set of discrete system equations. This type requires the use of a background mesh to perform the spatial integration. Methods that fit into this category are element free Galerkin (EFG) originally developed by Belytschko *et al.* [112] in 1994, the particle finite element method (PFEM) by Onate *et al.* [113] in 2004, the point interpolation method (PIM, Liu and Gu [114]), the natural element method (NEM) by Traversoni [115], the particle-in-cell method by the Las Alamos National Laboratory (Harlow [116]), and the material point method (MPM, Sulsky et al [117]) which is an extension of PIC. With these types of simulation methods, the problem domain is typically discretized by a field of nodes; the code must then create a background mesh using a triangulation approach such as Delaunay tessellation. Since the background mesh is generated by the solver, the user is alleviated of the burdensome meshing procedure. However, the background mesh requirement adds programming and computational complexity.

Although such methods are very computationally burdensome, with recent improvements in computer power, some researchers have successfully simulated the FSW process. Alfaro *et al.* [118, 119] have used NEM to simulate various large deformation processes such as FSW and aluminum extrusion. Because of the computational complexity of the method, they limit themselves to a 2D simulation of the weld zone only. Wu *et al.* [120] have used an adaptive EFG approach to simulate the plunge and advance

phase of FSW. They use time scaling techniques (all velocity terms in the formulation are scaled appropriately) to improve the computational time. Both research groups were mainly focused on showing the potential of the NEM and EFG method for FSW, they did not compare their simulation results with experiments. Fagan *et al.* [121] use MPM to simulate the FSW process for copper alloys. They used a mass scaling approach to reduce the calculation time to a minimum. Even with mass scaling, the simulation time was ~49 days using an 8 core Intel Xeon E5-2650. By any standard, a simulation that takes over a month is simply too long. Miyasaka *et al.* [122] used the EFG method to investigate the material movement during the FSW process.

2.6.3 MESHFREE METHODS NOT REQUIRING A BACKGROUND MESH (TRUE MESHFREE)

The second type of meshfree methods does not require any background mesh and encompasses collocation type methods (solution of the strong form of the PDEs). Some examples are the finite point method (FPM, by Onate *et al.* [123]) as well as the meshless local Petrov-Galerkin method (MLPG, Atluri and Zhu [124]). The smoothed particle hydrodynamics (SPH) method is one of the older and more mature meshless methods. The method was first introduced by Gingold and Monaghan [125] and by Lucy [126] in 1977. SPH was introduced for the simulation of strength of materials problems by Libersky and Petschek [127] in 1991. SPH has enjoyed a great deal of development over the past decades. It has been successfully used for many short-to-medium duration transient solid mechanics problems. However, the standard SPH approach is not well suited to long-duration strength of material problems. The main difficulty lies in that the standard SPH formulation uses an explicit (forward difference) time integration scheme that is conditionally stable. For solid mechanics problems, the stability criterion leads to very small time steps (inversely proportional to the speed of stress wave propagation in the solid). This leads to unreasonably long computation times for longer duration simulations.

To date, there is no evidence of research groups using methods such as FPM or MLPG to simulate FSW. SPH on the other hand has been used by a number [122, 128-134] of groups successfully. The meshfree nature of SPH lends itself to formulation of the field equation using a Lagrangian frame of reference, which makes SPH ideal for the large plastic deformation found in the FSW process. A three dimensional implementation of SPH can be expected to be rather computationally onerous without an

advanced parallel programming strategy. There is no doubt that an efficient parallel implementation of SPH will be warmly welcomed by the FSW simulation community. Full details of the parallel programming strategy adopted in the current research project are presented in section 4.

2.7 MATERIAL MODELS

One of the more important modeling considerations in a continuum mechanics based simulation of FSW is the material behavior. An over or under estimate of a single parameter can significantly change the simulation results of the process. Most researchers [94, 95, 104, 107, 108, 111, 119, 135-137] use constant thermal properties such as the heat capacity and thermal conductivity in the numerical model. Using constant thermal properties can be acceptable for fluid models simulations with weakly or decoupled treatment of the thermal and flow problems. However, the variation of thermal properties with temperature should be considered for solid-mechanics models where the flow stress (which directly affects the stiffness of the aluminum) is strongly related to the temperature field. Very little work has been undertaken regarding temperature dependent thermo-physical properties among researchers performing current state of art FSW simulations.

Wide ranges of material models are available for modeling aluminum. The main attribute that the models have in common is an attempt to form a link between the flow stress, strain, strain rate, and temperature. At the very least, a material model used for FSW must include the effects of thermal softening. Specifically for AA6061-T6, the effects of strain hardening and strain rate have been shown to be of less importance than thermal softening.

Numerical simulation of FSW has been undertaken by a myriad of groups using a wide variety of material models. Chiumenti *et al.* [138] use the non-Newtonian Norton-Hoff fluid based law:

$$\sigma_{eff} = K(T)\sqrt{3}\left(\sqrt{3}\dot{\varepsilon}\right)^{m(T)} \tag{2-8}$$

where $\dot{\varepsilon}$ is effective strain rate. Simple inspection of their implementation shows no dependence on temperature since they use constant coefficients for $K(T)$ and $m(T)$ (material specific constants). They have chosen the parameters to suit the average temperature in the weld zone. Certainly, such a model

suffers from a lack of robustness. As such, this approach cannot be used to evaluate the effect of different process parameters.

Dialami *et al.* [107] also use the Norton-Hoff model. They incorporate thermal softening by using temperature dependent functions for $K(T)$ and $m(T)$ of the form:

$$K(T) = \begin{cases} A + bT, & T < T_{melt} \\ 0, & T \geq T_{melt} \end{cases} , \tag{2-9}$$

$$m(T) = \begin{cases} C1 + d1T, & 20 \leq T < 500°C \\ C2 + d2T, & 500°C \leq T < T_{melt} \\ 0, & T \geq T_{melt} \end{cases}$$
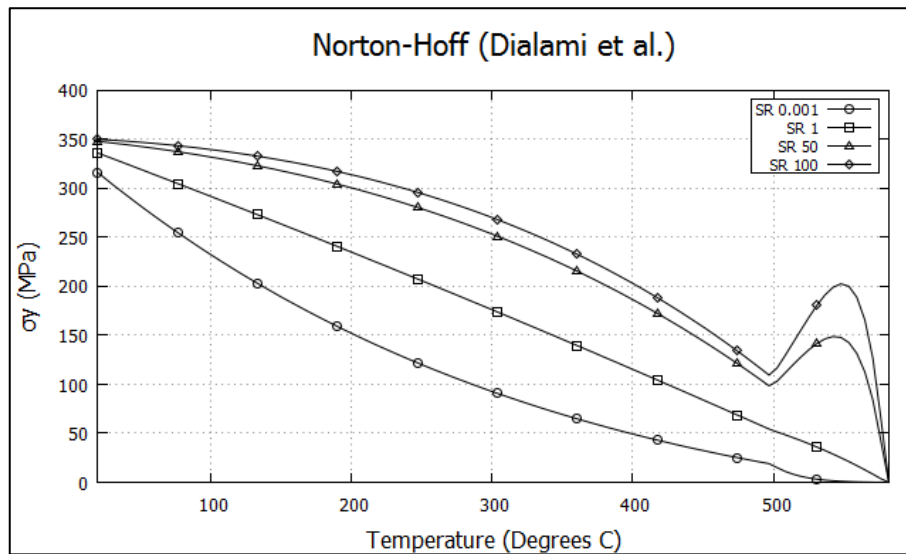
Figure 2-13 – Flow stress model used by Dialami *et al.* [107]

where T_{melt} *is the melt temperature,* A, b, $C1$, $C2$, $d1$, *and* $d2$ are material specific constants. This approach has merit in the sense that thermal softening can now be incorporated. However, the actual values of the constants that they use in their study leads to erratic changes in the flow stress near 500 °C at high strain rates. It is common knowledge that the strain rate in the stir zone is on the order of 100 to 1000 s^{-1}. With this in mind, the flow stress at different strain rates as a function of temperature obtained with their implementation is shown in Figure 2-13 (using the values for A, b, $C1$, $C2$, and $d2$ from Dialami *et al.* [107] to create the graph). The graph clearly shows the erratic behavior.

Timesli *et al.* [134] use a temperature dependent viscosity fluid model in their work to simulate FSW with SPH. They use the Johnson-Cook temperature ansatz to decrease the materials resistance to flow with increasing temperature. They compare their results to a CFD model employing the same assumptions and material behavior. They do not provide validation of the model against experimental work.

Assidi *et al.* [139] have used the Hansel-Spittel model for their simulation work of FSW with AA6061-T6:

$$\sigma_{eff} = A e^{m_1 T} \varepsilon^{m_2} \dot{\varepsilon}^{m_3} e^{m_4/\varepsilon_p} \qquad (2\text{-}10)$$

The model takes into account thermal softening, strain hardening, and strain rate sensitivity. The constants A, m_1 to m_4 are material specific. They note that their implementation of the model results in drastic over prediction of the temperature in the weld zone. This can be explained by the fact that the model predicts a strain rate dependence that decreases with increasing temperature as shown in Figure 2-14 (the material constants for A, m_1 to m_4 are taken from Assidi *et al.* [139] to create the graph); this is the inverse of the common behavior of AA6061-T6. In fact, a recent study by Dorbane *et al.* [140] showed that this material is entirely strain rate insensitive at room temperature, but that a significant increase in flow stress occurs at increasing strain rates when the material is at higher temperatures. The other issue with the Hansel-Spittel model is that strain hardening is considered with a power law. This leads to an over prediction of the flow stress at high plastic strain values (typical in FSW).
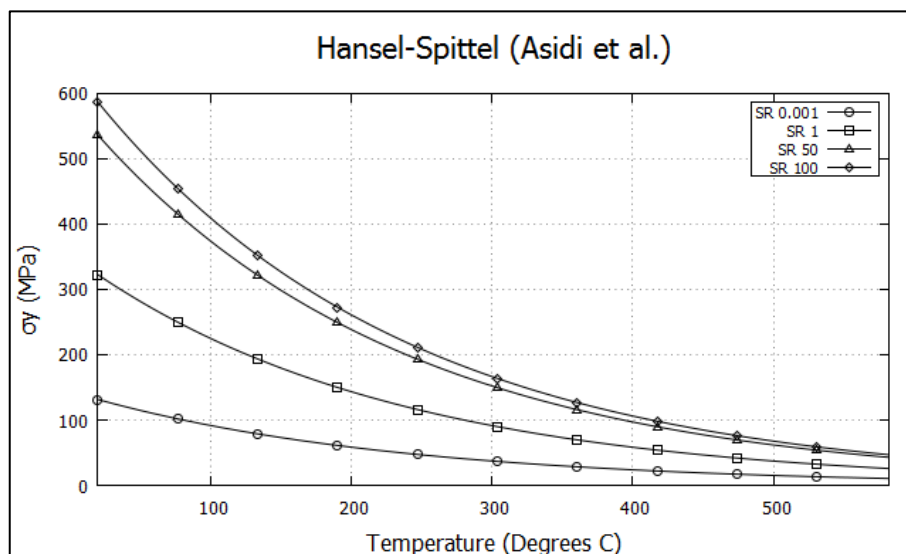
Figure 2-14 – Hansel-Spittel model for AA6061-T6 – Assidi *et al.* [139]

Kuykendall *et al.* [141] compare the Sellars-Tegart and Johnson-Cook model in their study of the FSW process. The Johnson-Cook model has the form:

$$\sigma_{eff} = \left(A_{JC} + B_{JC}\varepsilon^{p^{n_{JC}}} \right)\left(1 + C_{JC}\ln\dot{\varepsilon} \right)\left(1 - T^{*^{m_{JC}}} \right)$$ (2-11)

where T^* is the homologous temperature, ε_p is effective plastic stain, and A_{JC}, B_{JC}, n_{JC}, and m_{JC} *are material specific constants.* They note that the Johnson-Cook model provides an underestimate of the peak temperature and an overestimate of the plastic strain in the weld zone. The Johnson-Cook model is well known to provide flow stress curves that are inaccurate at the level of plastic strain found in FSW. The behavior of the model is shown in Figure 2-15 (JC constants taken from Schwer [142] to make the graph); the main issue is that the flow stress increases exponentially without bound as plastic strain increases.
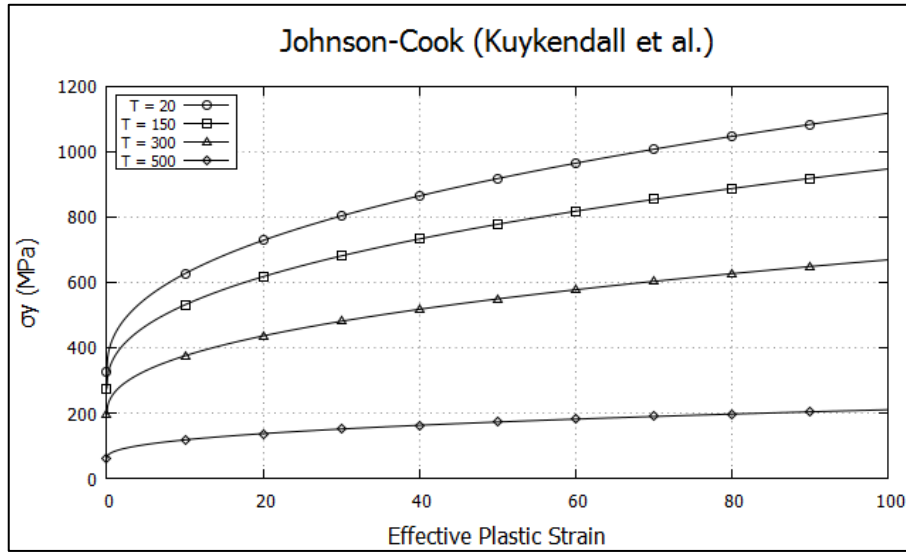
Figure 2-15 – Johnson-Cook model for AA6061-T6 – Kuykendall *et al.* [143]

In FSW, we expect to see plastic strain levels approximately 100-300. As we can see in the graph, the predicted flow stress at room temperature and effective plastic strain of 100 is over 1100 MPa; this is a gross over prediction and leads to overly stiff behavior in the numerical models. Furthermore, the material model predicts a non-existent strain rate dependency.

The Johnson-Cook model incorporates a power law thermal softening through the homologous temperature. Figure 2-16 shows a comparison of the variation of the initial yield as a function of temperature as found from MatWeb [144] and by the Johnson-Cook model [142, 145].
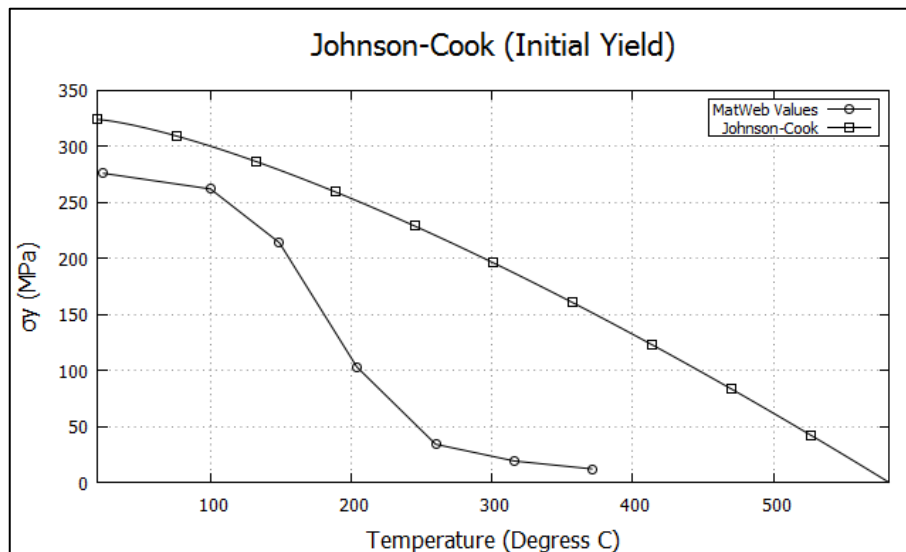
Figure 2-16 – Temperature dependent initial yield

We can clearly see that the Johnson-Cook model significantly over predicts the value of yield across the majority of the temperature range. This will contribute to an increase in the stiffness of the numerical model. Evidently, an appropriate material model to predict the flow stress during the FSW process is currently lacking. In this work, a newly developed flow stress model that is based on compression tests and the work of Dorbane *et al.* [140] will be presented. With the new model, improved material behaviour of the aluminum alloy during the entire FSW process can be accounted for. Details of the new flow stress model are provided in Section 3.10.

2.8   FRICTION AND CONTACT BEHAVIOUR

As the FSW tool mechanically joins the two plates together, heat is generated due to two main sources:

- Friction at the interface between the work pieces and the tool, primarily under the shoulder
- Mechanical work due to the work piece material being plastically deformed

The stick condition (velocity of the contacted body is constrained to be the same as the contacting body) is rather common in fluid like simulations of FSW. The general approach is to require that the contacted nodal points have the same velocity as the contacting object. Pan *et al.* [132, 146, 147] use a stick contact model in their work simulating the FSW process with smoothed particle hydrodynamics. They compare their results to those published by another group and note a slight over prediction of the temperature in the material.

Wang *et al.* [148] provide an insightful review of various stick and slip friction modelling approaches. They report that the use of a stick model tends to distribute the heat generated across a wider volume compared to the slip model. This is because the slip model will simply provide heat input at the interface between the two contacting bodies. One of the major drawbacks of the stick contact model is that the heat generated in such a model can only be due to plastic work. Since the material sticks to the tool, heat generation due to friction is not possible.

The Coulomb model provides a way to model static and dynamic friction between two bodies in contact. The approach is generally formulated using a penalty method. In this method, any penetration between contacting bodies is resisted by a force that is proportional to the contact stiffness, k_{ij}, and the penetration depth, δ. Tangential forces develop so long as there is relative motion, v_{rel}, between the contacting bodies. The tangential force, F_T, is a function of the normal force, F_N, and the coefficient of friction, μ. The force normal at the contact interface is given by:

$$F_N = \begin{cases} 0, & \delta \le 0 \\ k_{ij}\delta, & \delta > 0 \end{cases} \qquad (2\text{-}12)$$

While the relative motion between the two bodies is null, the tangential force increases linearly until the static friction threshold is attained. After this point, relative motion is present, and the tangential force is $F_T = \mu F_N$.

Typically the mathematical model employed to represent friction between the tool and the work piece is the Coulomb friction model with a constant friction coefficient, many authors have produced good results with this approach [16, 105-108, 120, 130, 149]. Zhang [150] studies the differences between the Coulomb and the modified Coulomb friction models. In the modified model, a limit is put on the shear

stress than can build up at the interface. The friction model is also referred to as the Coulomb friction law with shear stress confinement. They show that the result with the modified approach provides closer results to experimental work.

The Norton model is provides a viscous friction approach. The model incorporates a modification to the friction coefficient, μ, as a function of temperature, T, and strain rate, $\dot{\varepsilon}$. The tangential force then becomes:

$$F_T = \begin{cases} 0, & v_{rel} = 0 \\ \mu(T, \dot{\varepsilon})F_N, & Otherwise \end{cases} \qquad (2\text{-}13)$$

Assidi *et al.* [139] study the Coulomb and Norton friction models. They note that the Coulomb law is not well suited for strain rate and temperature dependent friction behavior. They feel that the Norton model provides better representation of the slip mechanism between the tool and the work pieces.

A combination of the stick and slip friction models have also been used for FSW simulation. The formulation was introduced by Schmidt *et al.* [38]:

$$F_T = \delta_{SR} \frac{\sigma_y}{\sqrt{3}} A_c + (1 - \delta_{SR})\, \mu F_N \qquad (2\text{-}14)$$

where δ_{SR} is a slip ratio that weighs the friction behaviour more towards sticking ($\delta_{SR} = 1$) or more towards slipping ($\delta_{SR} = 0$). With this approach, a constant value for the stick parameter is often chosen to provide the best correlation to experimental results.

Schmicker [15] has proposed an exponential decay transition law for the use in simulation of the RFW process. The developed model bears close resemblance to the Coulomb friction law with shear stress confinement; however, regularization terms are added to smooth out the discontinuities associated with the transition from dry to viscous friction. The main goal of the proposed model is to improve convergence in their adaptive FEM code. Since this research is focused on a meshfree explicit code, the developed model is not expected to provide improvement since the meshfree code deals with discontinuities with relative ease.

Clearly, there is room for improvement and development of more precise friction models that take into account the underlying physics at the contact interface. In this work, a new phenomenological friction model that uses a cumulative damage approach is developed. Full details are presented in section 3.12.

2.9 FSW PROCESS PARAMETER OPTIMIZATION

The ultimate strength and fatigue resistance of a welded aluminum joint is of critical importance to the aeronautical industry. In recent years, many authors have focused on optimizing the ultimate strength of a friction stir welded joint using experimental approaches. Very little work has been undertaking to find optimal parameters using a fully coupled 3D thermomechanical simulation approach. Fraser *et al.* [151] found optimized process parameters for 1/8 inch AA6061-T6 plates. They used a simplified heat transfer finite difference model that neglected the material flow. Their approach was validated against the experimental work of Wanjara *et al.* [46], they showed excellent agreement with the experimental results in spite of using a thermal only model. Ammouri *et al.* [152] use a thermomechanical FEM model to predict the optimal process parameters for friction stir processing of AZ31b magnesium alloy. The objective of the optimization is to minimize the grain size of the friction stir processed plates.

The response surface method (RSM) is a very popular approach used by many authors [153-158] to optimize the strength of the weld of dissimilar 6061-T6 and 7975-T6 aluminum butt joint welds. The different research groups follow the general procedure of performing a set of experiments with different process parameters. They then form a response surface by using an appropriate surface fitting algorithm (most commonly least squares). Once the response surface is determined, the optimal process parameters can then be found using an appropriate optimization approach (such as Lagrange multipliers, steepest descent, conjugate gradient, etc.).

Laksihminarayaranan and Balasubramanian [159] use the Taguchi method to determine the most important process parameters for a high strength welded joint. They investigate the RDE-40 aluminum material in their study. Their method shows that the advancing speed, rotational speed, and the axial force on the tool are the most important process parameters. They found that the optimum tensile strength obtainable for the RDE-40 material is ~303 MPa. Their results were verified experimentally. The

Taguchi method is a popular approach and has been used by a number of authors [154, 155, 160-167] for the FSW process.

Zhang and Lui [168] investigate optimization of the underwater FSW (UWFSW) process. The UWFSW is a method that involves performing the FS weld in a closed circulation-welding chamber filled with water. The method allows for very high strength welds. They experimentally determined a set of data points at various welding speeds, and then fit a response surface with the least squares method. They use the constructed response surface to optimize the weld strength. They show that the tensile strength of the welded joint can be increased by 6% with the UWFSW method as compared to the conventional FSW process. A maximum tensile strength of 360 MPa can be obtained with an advance speed of 223 mm/min and a rotational speed of 983 rpm.

Heidarzadeh *et al.* [169] developed a four-parameter model for the tensile strength of pure copper butt welds. They show that the tensile strength of the weld is optimized with an advance speed of 84mm/min and a rotational speed of 942 rpm. The RSM method is used to fit a least squares regression model to the experimentally determined data points.

On the other hand, there has been relatively little research into the optimization of the advance and rotational speed of the FSW tool in order to provide welds that are strong enough (not necessarily optimal strength) and do not have defects. Maximizing the welding advance speed is of great importance for industries, such as structure and bridge fabrication in order to be profitable. Therefore, the optimization approach used in this work is focused on optimizing the speed of advance as opposed to the weld strength. Constraints are applied to the optimization problem to limit the speed of advance, the maximum welding temperature, and the speed of rotation.

2.10 MASS AND TIME SCALING

The time step is very small when simulating the FSW process using a fully coupled thermomechanical solid formulation with explicit time integration. Because of this, the simulation time can be excessively long. A few techniques can be employed to improve the calculation time such as mass scaling or time

scaling. The time step will be proportional to the characteristic element size, h, and inversely proportional to the bulk speed of sound, c, in the material:

$$\Delta t_{CFL} = CFL\left(\frac{h}{c + \|\bar{v}\|}\right) \tag{2-15}$$

This makes the time step approximately 1-100 nano-seconds (CFL can range from 0.1 to over 1.0 depending on the problem) for common metallic alloys. Mass scaling seeks to increase the minimum time step by artificially increasing the mass of an element. This leads to an increase in density and ultimately a decrease in the isothermal speed of sound in the material, $c = \sqrt{dp/d\rho|_T}$. However, for a simulation code using a weakly compressible approach, artificially decreasing the speed of sound will lead to an incorrect speed of the stress wave propagation. This will ultimately lead to inaccuracies in the elastic strain calculation. Although mass scaling is commonplace in explicit dynamic structural analyses [170-175], it should be used with great caution for hydrodynamic applications.

Velocity scaling (or time scaling) on the other hand does not affect the time step size; instead, it allows more simulation to be accomplished in a shorter period. The general idea is to scale all the terms with units of s^{-1} by a velocity scaling factor, $\tilde{V}_{SF}$. The rpm, velocity, thermal conductivity, convection coefficient, etc., are scaled in this approach. The analyst must then ensure that the dynamics of the system have not been unreasonably altered. Speed-up factors of 10-50 are conventional depending on the problem. During the simulation, one must keep track of the ratio of internal energy to kinetic energy. Limido *et al.* [176] as well as Villumsen *et al.* [177] use a time scaling factor of 10 in their work on high speed metal cutting. In their work, they also investigate a mass scaling ratio of 1.2 (an element with a mass of 1.0 kg is increases to 1.2 kg) and note that the force results are highly imprecise even with a small mass scaling factor. Schmidt used a mass scaling factor of 1.0×10^4 in his work on simulating FSW with the material point method (MPM). He notes that this scaling factor leads to significant error and ultimately decides to use no mass scaling.

3 METHODOLOGY

Numerical simulations of the FSW process can bring about an improved understanding of the effect of varying process parameters on the final weld quality. Ideally, the numerical model will predict (without previous knowledge of the outcome) the behaviour of the material during welding. Such an approach precludes the use of empirical models based on experimental results. Indeed, the predictions should come from the underlying physics of the process leading to a phenomenological approach. To this end, the FSW process can be described using the notions of continuum mechanics. First off, a set of conservation equations will be introduced that are applicable to the specifics of the FSW process. Following that, the discretization of these equations will be outlined using the meshfree Lagrangian method called smoothed particle hydrodynamics. The chapter is organized as follows:

Section 3.1 will discuss the underlying physics of the FSW process and the associated continuum mechanics formulation.

Section 3.2 introduces the smoothed particle method along with the important notions needed to write the "weak like" form of the conservation equations.

Section 3.3 provides the notions required to cast the SPH method for solid bodies.

Section 3.4 will focus on the discretization ("weak like" form) of the conservation equations using the SPH method.

Section 3.8 introduces some of the common deficiencies of the SPH method.

Section 3.9 will layout the various formulations required to describe the evolution of the temperature field in the body.

Section 3.10 presents a new flow stress model developed to better simulate the material behaviour during the entire FSW process.

Section 3.11 describes the contact approach used to model the interaction of the tool and/or supporting structure with the work pieces.

Section 3.12 introduces a new friction modelling approach using a cumulative damage approach.

Section 3.13 will provide an efficient means to model the threads and scroll on the tool.

Section 3.14 concentrates on the tool wear evaluation approach adopted in this research project.

Section 3.15 delivers a description of the solution procedure adopted in the developed simulation code.

3.1 CONTINUUM MECHANICS FOR FSW

On a microscopic level, all matter (or physical bodies) can be said to be discontinuous, made up of subatomic particles. Ideally, one could study the effect of forces and displacements on a body from a quantum mechanics (microscopic point of view). This is not practical with our current knowledge of the sub-atomic structure and the power of today's computers. For the purpose of this research work, in order to study how matter (such as fluids and solids) will react to forces and deformation, it is necessary to adapt a macroscopic frame of reference. In this manner, bodies can be treated as being made up of a continuum. We can thus say that the properties of the body are continuous, such as its mass. The premise follows the idea that although microscopically the body is known to be made of discrete particles, these particles are uniformly distributed to permit the treatment as a continuum on a macroscopic scale.

In the continuum description as shown in Figure 3-1, the focus is to determine the evolution of the location, $\bar{X}$ and temperature, T, of a set of material points in an arbitrary shaped body of finite volume, V, and bound by a continuous surface, S. A frame of reference that follows the material point(s) should be adopted since the ultimate objective of the research project is to be able to predict the evolution of defects within a FSW joint. This type of reference frame is called Lagrangian and relates the position of the points within the body at a point in time, t, to a fixed reference frame, $\bar{x}(\bar{X}, t)$ as well as the temperature field, $T(\bar{X}, t)$. The behaviour of the body can be described by ensuring that mass, momentum, and energy are not created or destroyed; simply conserved within the body.
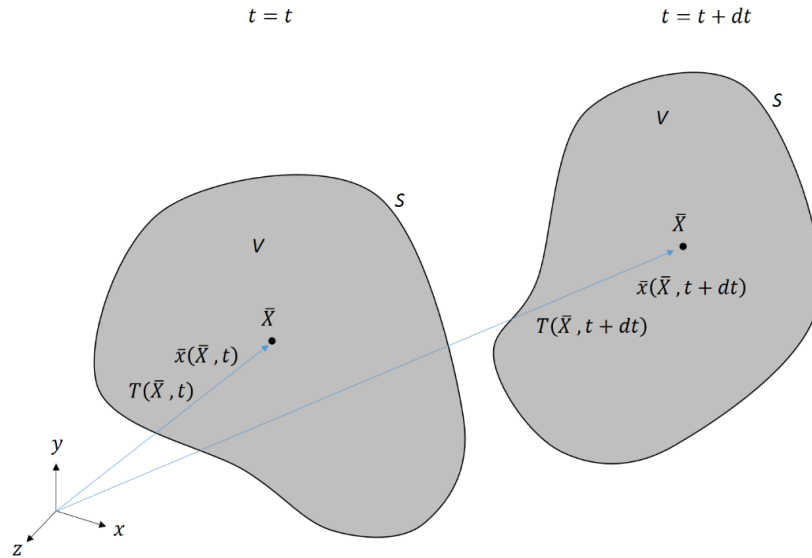
Figure 3-1 – Continuum description

The first conservation equation states that the mass of a material point must remain the same. A convenient measure of the quantity of mass within a finite volume is the material density, ρ. Imagine an infinitesimal volume, dV, bound by an infinitesimal surface, dS, the rate of change of the mass within the finite volume will be related to the material density and the change in volume:

$$\frac{Dm}{Dt} = \frac{D}{Dt} \int_V \rho dV = 0 \tag{3-1}$$

From equation (3-1), evidently, if mass is to remain constant, there must be a balance between ρ and dV. A body that can undergo change in density is said to be compressible. This leads to the notion that the density can change at a material point as time evolves, and as such, to the concept of continuity of the body, where:

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \bar{v} = 0 \tag{3-2}$$

Equation (3-2) introduces the concept that the time rate of change of density must be balanced by the divergence of the velocity field at that point (velocity, $\bar{v}$, as usual is the time rate of change of position, dx/dt). Note the use of the material derivative; for more information see Chapter 9 - Appendices. To put this Concept into context, imagine a rigid box filled with a fluid and with an inlet and an outlet; now

imagine that more fluid flows in than flows out. The only way that this could occur is for the fluid to be compressed, and as such increasing the quantity of fluid in the fixed volume leads to an increase in the density of the fluid. Commonly, engineering materials are considered incompressible, which is convenient but not strictly true. In this work, material is considered compressible as this better represents the underlying physics of the process.

The next concept says that a body in motion should remain in motion without the presence of resistance to the motion. In this sense, the linear momentum, $\bar{P}$, of the body must be conserved:

$$\frac{D\bar{P}}{Dt} = \frac{D}{Dt}\int_V \bar{v}\rho dV = 0$$

(3-3)

However, in the presence of external effects, the linear momentum can change. By introducing a mass specific body force vector, $\bar{b}$, the differential form of conservation of linear momentum is:

$$\frac{D\bar{v}}{Dt} = \frac{1}{\rho}\nabla \cdot \bar{\bar{\sigma}} + \bar{b}$$

(3-4)

The body force can be due to different sources, such as gravity and electromagnetic forces. In the FSW process, the body force is insignificant in comparison to the other forces and can be neglected without any loss of precision. The momentum equation for FSW is then:

$$\frac{D\bar{v}}{Dt} = \frac{1}{\rho}\nabla \cdot \bar{\bar{\sigma}}$$

(3-5)

which is the same as equation (3-4), without the body force term. The idea of divergence of the stress field, $\nabla \cdot \bar{\bar{\sigma}}$, has been included. This notion quantifies the internal force in the body. An extension of linear momentum to angular momentum, $\bar{H}$, conservation can be made by taking the vector product of the forces about an axis:

$$\frac{D\bar{H}}{Dt} = \frac{D}{Dt}\int_V \bar{x} \times \bar{v}\rho dV = 0$$

(3-6)

This conservation law is not directly used in the continuum model; however, it does lead to the proof of the symmetry of the total stress tensor. By taking moments with respect to each of the individual axis, one can show that:

$$\bar{\bar{\sigma}} = \bar{\bar{\sigma}}^T \tag{3-7}$$

The final conservation principle states that the total energy, U, is neither created nor destroyed; simply conserved. The differential form of the energy conservation equation is:

$$\rho \frac{DU}{Dt} = \bar{\bar{\sigma}}(\nabla \cdot \bar{v}) - \nabla \cdot \bar{q} + \dot{q} \tag{3-8}$$

where $\bar{q}$ is the heat flux and $\dot{q}$ is a source term. The total energy in the system should consider the full contribution of kinetic and potential energy. Energy conservation for the remainder of the potential portion (specifically, the thermal portion) of the total energy must be subsequently developed. More will be stated concerning the complete form of conservation of energy in section 3.9. Note that kinetic energy is insignificant (in comparison to thermal and internal energy) in the FSW process and will be neglected.

This then forms a basic set of conservation equations that can be used to model the underlying physics of the FSW process. In the subsequent sections of this chapter, increasing complexity will be built onto the basic equations to improve the approximation of the material behaviour during the FSW process.

3.2 THE SMOOTHED PARTICLE METHOD
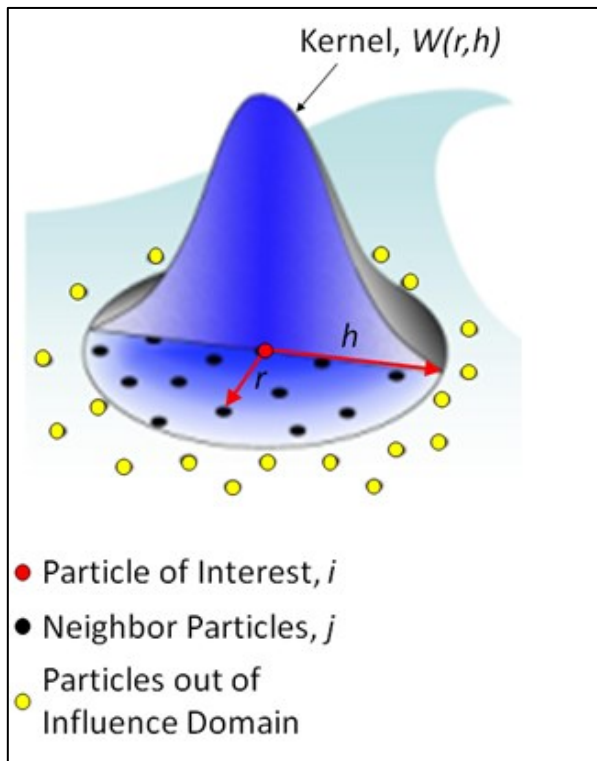
Figure 3-2 – Interpolation in SPH method

Smoothed particle hydrodynamics (SPH) is a collocation meshfree method for a wide variety of dynamic problems. Because of its meshfree nature, it can easily be formulated from a Lagrangian frame of reference. SPH has strong roots in fields such as astrophysics, magneto-hydrodynamics, computational fluid dynamics, and computational solid mechanics. Indeed, the list is by far from being exhaustive, and SPH is constantly being applied to new and sophisticated problems (lava flow, hydrodynamics of a swimming dolphin, tsunamis impact, high-speed impact, fractures mechanics, etc). The main attraction to this method is that no mesh is required. This provides a powerful framework for solving large plastic deformation problems that typically cannot be solved by conventional meshed based methods. For these reasons, SPH is ideally suited for numerical simulation of the entire FSW process.

The method was originally proposed by two independent research groups within the same year, Gingold and Monaghan [125] showed that the method could be used to simulate non-spherical stars and Lucy [126] used the method to test the theory of fission for rotating protostars. Interestingly enough, both groups based their methods on statistical approaches to estimate probability densities. The original SPH schemes proposed in 1977 were developed for applications to astrophysics problems and did not conserve linear and angular momentum [178]. Later in 1982, Gingold and Monaghan modified the formulation to be conservative by using the Lagrangian to describe the conservation laws for a compressible dissipative fluid [179].

The SPH method is considered a mesh-free method because the set of field equation (conservation equations for a solid body in this case) are solved by interpolating from a set of j particles that are within the influence domain of particle i. The size of the influence domain is determined by the smoothing length, h. The smoothing length is commonly taken as the initial (uniform) particle spacing, Δs, multiplied by the smoothing length scale factor, h_{scale}:

$$h = h_{scale}\,\Delta s \tag{3-9}$$

A value of 1.1 is used for h_{scale} throughout this work. This value provides good calculation performance and precision (more details in section 5.5.1). Figure 3-2 gives a graphical representation of the SPH interpolation concept. In this numerical method, the strong form of the field equations can be 'weakened' (this term is coined from the finite element method where by the 'weak form' is used to solve otherwise unmanageable equations [180]) into a set of discrete ordinary differential equations). Note that Liu [181] calls the discretization arising from the SPH method a "weak like" form. SPH is similar to the FDM in the sense that they are collocation schemes. The resulting numerical method is truly meshfree since no background mesh is required. The meshfree nature of SPH makes it ideally suited for numerical simulation of the FSW process.

3.2.1   DISCRETE INTERPOLATION

The basic premise of the SPH method is to reduce a set of PDEs to a set of ordinary differential equations by an approximate interpolation formulation. A continuous function is approximated by a convolution integral:

$$f(\bar{x}) = \int f(\bar{x}')\,W(\bar{x} - \bar{x}', h)\,d\bar{x}' \tag{3-10}$$

where $W(\bar{x} - \bar{x}', h)$ is called the kernel function and has units of inverse of volume. The kernel function is also commonly referred to as the smoothing function. It is a function of the spatial distance between the point at which the function is to be calculated (calculation point, $\bar{x}$) and the interpolation location ($\bar{x}'$), and h is the smoothing length. The kernel is the key to the true meshfree nature of the SPH method. The continuous SPH interpolation equation can then be written for a set of discrete material points:

$$\langle f(x_i{}^\alpha)\rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) W(r,h) \tag{3-11}$$

here x_i is the spatial location vector for particle i and x_j for the jth particle, h is the smoothing length that determines the size of the influence domain of the jth particles on a particle i. m_j, ρ_j are the mass and density of a jth particle, and $r = \left| x_i{}^\alpha - x_j{}^\alpha \right|$. $W(r,h)$ is again the interpolation kernel; it will be written as W_{ij} in the future. The sum is taken over the total number (N_i) of j particles within the *influence domain of i; these are termed the neighbors of the ith particle. Determining the neighbors list* is a major part of the computational time in the SPH method (typically around 30% of a cycle).

The gradient of the interpolation kernel can also be written as:

$$\nabla_i W_{ij} = \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\beta} = \frac{dW(r,h)}{dR}\left(\frac{1}{h}\right)\left(\frac{1}{r}\right)(\bar{x}_i - \bar{x}_j) \tag{3-12}$$

The gradient of a function can be approximated by the value of the function at the neighboring particles multiplied by the gradient of the smoothing function:

$$\langle \nabla f(x_i{}^\alpha)\rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) \nabla_i W_{ij} = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\alpha} \tag{3-13}$$

Equation (3-13) makes the evaluation of first and second derivatives relatively simple in the SPH method. A similar approach can be used to write an approximation of the divergence of a vector:

$$\langle \nabla \cdot f(x_i{}^\alpha)\rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\beta) \cdot \nabla_i W_{ij} = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\beta) \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\beta} \tag{3-14}$$

or for the curl of a vector:

$$\langle \nabla \times f(x_i{}^\alpha)\rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\beta) \times \nabla_i W_{ij} \tag{3-15}$$

and the dyadic product:

$$\langle \nabla \otimes f(x_i{}^\alpha) \rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) \otimes \nabla_i W_{ij} = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\beta} \qquad (3\text{-}16)$$

This lays the foundation for the SPH method that will be used to solve the set of conservation equations to describe the deformation of a general solid body. As of yet we have not made any statements as to the requirements of the kernel function; this will be explained in detail in the next section.

3.2.2 SMOOTHING FUNCTIONS

The smoothing function is the heart of the SPH method and as such has been an area of acute interest for many research groups. Liu *et al.* [182, 183] give details of the seven requirements for constructing a smoothing function:

1. Unity Condition – The smoothing function is required to be normalized over the support domain such that:

$$\int W(\bar{x} - \bar{x}', h) d\bar{x}' = 1 \qquad (3\text{-}17)$$

2. Compact Support – The value of the smoothing function should be zero outside of the support domain:

$$W(\bar{x} - \bar{x}', h) = 0, for \ |\bar{x} - \bar{x}'| > h_{scale} \Delta s \qquad (3\text{-}18)$$

where h_{scale} is a scaling factor on the smoothing length that is used to include more or fewer particles in the discrete sums.

3. Positivity Condition – The value of the smoothing function must be greater than zero for any point within the support domain:

$$W(\bar{x} - \bar{x}', h) > 0, for \ |\bar{x} - \bar{x}'| < h_{scale} \Delta s \qquad (3\text{-}19)$$

4. Decay Condition – The smoothing function should be a monotonically decreasing function of the distance between the calculation point and the interpolation point.

5. Delta Function Property – As the smoothing length approaches zero, the smoothing function should be equal to the Dirac delta function:

$$\lim_{h \to 0} W(\bar{x} - \bar{x}', h) > \delta(\bar{x} - \bar{x}') \tag{3-20}$$

6. Symmetric Property – The smoothing function should be symmetric, this requires that the function be even. This property ensures that linear and angular momentum are conserved, because particle forces become action-reaction pairs.

7. Smoothness Condition – The smoothing function should be smooth without sharp variations of the gradient of the smoothing function.

A myriad of smoothing functions have been proposed, and depending on the simulation type, the different smoothing functions have been successfully used in certain types of problems. The Gaussian function was one of the first functions proposed; this approach gives a smoothing function that is infinitely differentiable. The main drawback is that the influence domain is not compact (compact support) because the value of the function goes to zero at $\pm\infty$. This means that every particle must be included in the evaluation of the discrete conservation equations. The Gaussian smoothing function takes the form:

$$W(R, h) = \alpha_d e^{-R^2} \tag{3-21}$$

where α_d is a dimension specific constant (values given for many kernels in Table 3-1) and $R = |x_i{}^\alpha - x_j{}^\alpha|/h = r/h$. The Gaussian kernel can be used with a truncated support domain to satisfy the compact support condition, but this does not satisfy the unity condition. Other authors have proposed Gaussian kernels that have compact support and satisfy the unity condition, but they are not widely used. The cubic B-spline function as proposed by Gingold and Monaghan [179, 184] is the most popular:

$$W(R, h) = \alpha_d \begin{cases} \dfrac{2}{3} - R^2 + \dfrac{1}{2}R^3 & for\ 0 \leq R < 1 \\ \dfrac{1}{6}(2 - R)^3 & for\ 1 \leq R < 2 \\ 0 & for\ R \geq 2 \end{cases} \tag{3-22}$$

The popularity of this kernel comes from the fact that it satisfies all the requirements, is relatively easy to implement in common programming languages (although the piecewise nature requires the use of conditional statements), and closely resembles the Gaussian kernel. The first derivative is continuously smooth, whereas the second derivative is piecewise linear. This is one of the root causes of the tensile

instability problem in SPH. In spite of the tensile instability, many authors [127, 176, 185-190] note that the B-spline kernel provides an excellent balance between performance and precision.

Quartic and Quintic functions have been proposed by other authors. Higher order kernels are said to be good for an organized orthogonal mesh, but not when the mesh becomes disordered. This is because they require a cancelation of positive and negative contributions that is not possible when the particles are not equally spaced. A higher order smoothing function is the Wendland kernel [191]:

$$W(R,h) = \alpha_d \left(1 - \frac{R}{2}\right)^4 (2R + 1) \ for \ 0 \leq R < 2 \tag{3-23}$$

This is a very attractive kernel from a programming standpoint since no conditional statements are required to implement it. Another quintic kernel proposed by Morris gives improved precision at the expense of increased computational cost:

$$W(R,h) = \alpha_d \begin{cases} (3-R)^5 - 6(2-R)^5 + 15(1-R)^5 & for \ 0 \leq R < 1 \\ (3-R)^5 - 6(2-R)^5 & for \ 1 \leq R < 2 \\ (3-R)^5 & for \ 2 \leq R < 3 \\ 0 & for \ R \geq 3 \end{cases} \tag{3-24}$$

With a support domain of $3R$, this kernel extends further than the others do. This means that the interpolation involves a large number of neighbors in the discrete sums. Because of this, the kernel drastically slows down the calculations (about twice as long as the cubic B-spline).

A rather interesting kernel that has compact support is the modified mollifier proposed by Ismail and Reddy [192]. This approach provides an infinitely differentiable compactly supported smoothing function. The function takes the form:

$$W(R,h) = \alpha_d \begin{cases} e^{\frac{-1}{(1-R^2)}}(1-R^2)^8 & for \ 0 \leq R < 1 \\ 0 & for \ R \geq 1 \end{cases} \tag{3-25}$$

This kernel is not likely to gain widespread acceptance due to the computational complexity in a typical programming language. The evaluation of the exponential is somewhat costly, and the $(1 - R^2)^8$ term is inherently expensive. On the other hand, the support domain extends only to R as compared with the B-

spline that extends to $2R$, as such; fewer particles will be included in the interpolation, requiring less computation effort per cycle.

Johnson and Beissel [193] use a quadratic smoothing function that has a discontinuous first derivative that is always increasing:

$$W(R,h) = \alpha_d \begin{cases} \dfrac{3}{16}R^2 - \dfrac{3}{4}R + \dfrac{3}{4} & for\ 0 \leq R < 2 \\ 0 & for\ R \geq 2 \end{cases}$$

(3-26)

They mention that this type of kernel is ideal for high-speed impact simulations (mainly compressive stress fields). According to Monaghan [178] this smoothing function is not well suited to disordered particle arrangements. For this reason, the function is not expected to give good results for the FSW simulations.

In this work, the smoothing function used is called the hyperbolic spline developed by Yang *et al.* [194]. This function tends to provide improved performance for problems that are mainly treating compressive stress fields with disordered particles. The smoothing function is defined as:

$$W(R,h) = \alpha_d \begin{cases} R^3 - 6R + 6 & for\ 0 \leq R < 1 \\ (2 - R)^3 & for\ 1 \leq R < 2 \\ 0 & for\ R \geq 2 \end{cases}$$

(3-27)
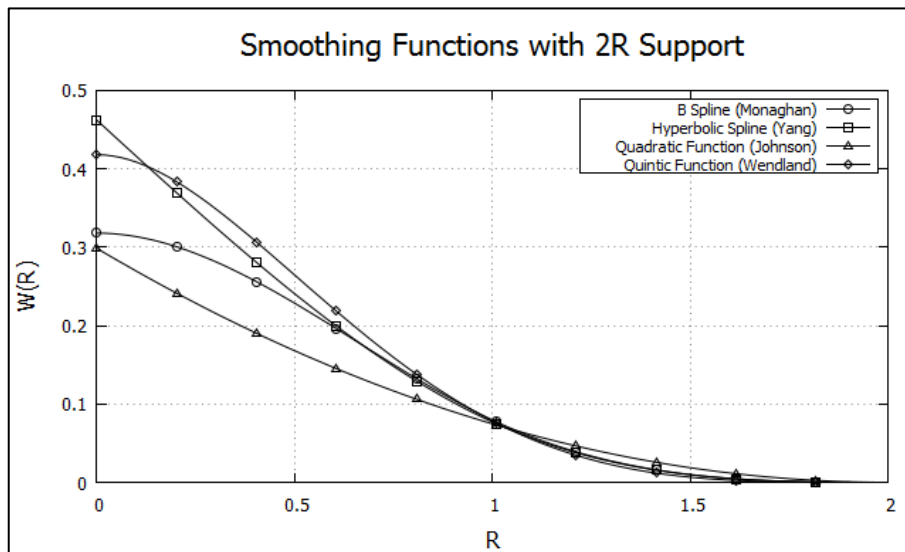
Figure 3-3 – Comparison of smoothing function with 2R support

A graph comparing the smoothing functions with $2R$ support is shown in Figure 3-3. The conservation equations in SPH form will involve the spatial derivative of the smoothing functions, for this reason, a close look at the smoothing function derivatives is more instructive. Figure 3-4 shows a comparison of the smoothing function derivatives. Both the quadratic and hyperbolic smoothing functions provide improved results for compressive stress fields. The reason for this can be inferred from the graphs; the value of the smoothing function derivative does not fall to zero as the particle pairs become increasingly close. This is beneficial for problems subject to large amounts of compression.
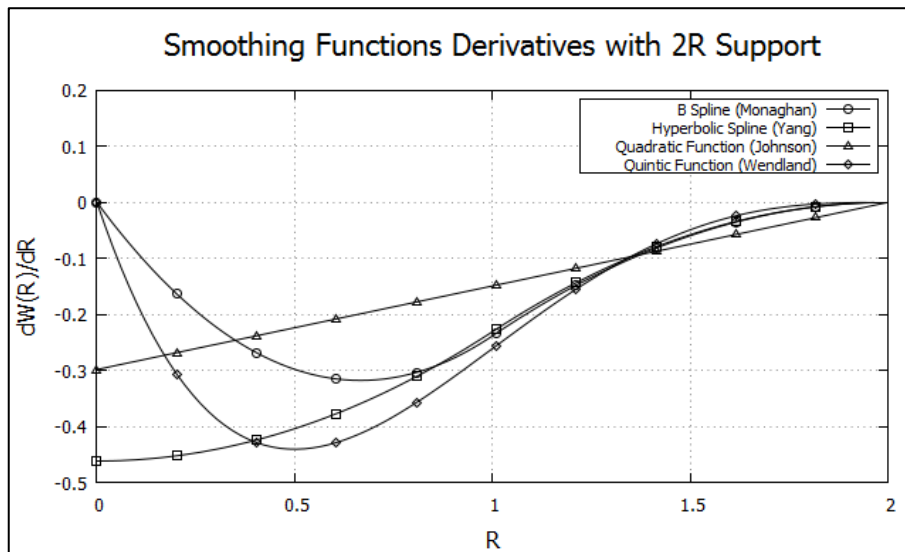
Figure 3-4 – Comparison of smoothing function derivatives with 2R support

Table 3-1 – Dimensional Specific Smoothing Constants

Smoothing Function	α_d - 2D	α_d - 3D	Support Size
Gaussian	$1/\pi h^2$	$1/\pi^{3/2}h^3$	∞
Cubic B-Spline (Monaghan)	$15/7\pi h^2$	$3/2\pi h^3$	2R
Quintic Function (Wendland)	$7/4\pi h^2$	$21/16\pi h^3$	2R
Quintic Spline (Morris)	$7/478\pi h^2$	$3/359\pi h^3$	3R
Quadratic Function (Johnson)	$2/\pi h^2$	$5/4\pi h^3$	2R
Hyperbolic Spline (Yang)	$1/3\pi h^2$	$15/62\pi h^3$	2R

3.3 SOLID MECHANICS APPROACH FOR FSW

In the fluid approach, the deviatoric stresses are found directly from:

$$\bar{\bar{S}} = 2\mu_{eff}\left(\dot{\bar{\bar{\varepsilon}}} - \frac{1}{3}tr(\dot{\bar{\bar{\varepsilon}}})\bar{\bar{I}}\right) \tag{3-28}$$

where μ_{eff} is an effective non-Newtonian viscosity and $\bar{\bar{I}}$ is the identity matrix (has a value of one along the diagonal of the tensor and zero elsewhere). This approach will neglect the effects of elastic strains. Such a simplification is admissible for certain types of studies on the FSW process (material flow in the weld zone, temperature distribution). Many researchers prefer to use the fluid approach [93-96, 98, 99]. On the other hand, limiting the simulation to only plastic strains leads to the inability to predict residual deformation, stress, and defect formation. In this project, the effects of elastic strains will be included using an appropriate solid mechanics formalism. In such an approach, the stress is found by integrating an appropriate objective stress rate, $\dot{\bar{\bar{S}}}$, equation. Here, the discrete form of the Jaumann rate equation has been adopted:

$$\frac{d\bar{\bar{S}}}{dt} = \dot{\bar{\bar{S}}} = 2G\left(\dot{\bar{\bar{\varepsilon}}} - \frac{1}{3}tr(\dot{\bar{\bar{\varepsilon}}})\bar{\bar{I}}\right) + \bar{\bar{S}}\bar{\bar{\Omega}} + \bar{\bar{\Omega}}\bar{\bar{S}} \tag{3-29}$$

where G is the shear modulus and $\bar{\bar{\Omega}}$ is the anti-symmetric spin tensor. Note that $\left(\dot{\bar{\bar{\varepsilon}}} - \frac{1}{3}tr(\dot{\bar{\bar{\varepsilon}}})\bar{\bar{I}}\right)$ is the deviatoric strain rate. This approach ensures that the determined stress is frame-indifferent (Galilean invariant). Other objective stress rates can be used such as Truesdell or Green-Nahgdi; more information can be found from [195-197]. The strain rate tensor in equation (3-29) is defined as:

$$\dot{\bar{\bar{\varepsilon}}} = \frac{1}{2}(\nabla\otimes\bar{v} + (\nabla\otimes\bar{v})^T) \tag{3-30}$$

and the spin tensor is:

$$\bar{\bar{\Omega}} = \frac{1}{2}(\nabla\otimes\bar{v} - (\nabla\otimes\bar{v})^T) \tag{3-31}$$

The deviatoric stress is composed of the total stress (commonly referred to as Cauchy stress) and the hydrostatic stress:

$$\bar{\bar{S}} = \bar{\bar{\sigma}} + p\bar{\bar{I}} \tag{3-32}$$

The hydrostatic (isotropic) pressure is:

$$p = -\frac{1}{3}tr(\bar{\bar{\sigma}}) \tag{3-33}$$

Section 3.15 will introduce the solution procedure used to integrate the set of equations in the simulation code. The main importance of using the solid approach is that stress is now a history variable (as opposed to a state variable in the fluid approach). This leads to the ability to keep track of the history of the deformation of the work pieces during the FSW process.

3.3.1   EQUATION OF STATE FOR WEAKLY COMPRESSIBLE MATERIAL

In the context of a compressible solid, the material point pressure is linked to the density using an appropriate equation of state (EOS). Many EOS' are possible; one such is the Gruneisen EOS. This formulation uses a different form in compression:

$$p = \frac{\rho_0\, c^2\, \mu_{comp}\left[1 + \left(1 - \frac{\gamma_0}{2}\right)\mu - \frac{a}{2}\mu_{comp}^2\right]}{\left[1 - (S_1 - 1)\mu_{comp} - S_2\left(\frac{\mu_{comp}^2}{1 + \mu_{comp}}\right) - S_3\left(\frac{\mu_{comp}^3}{(1 + \mu_{comp})^2}\right)\right]^2} + \left(\gamma_0 + a\,\mu_{comp}\right)U \tag{3-34}$$

and in tension:

$$p = \rho_0\, c^2\, \mu_{comp} + \left(\gamma_0 + a\,\mu_{comp}\right)U \tag{3-35}$$

where a, S_1, S_2, and S_3 are material specific constants. c is the isothermal speed of sound in the material, and γ_0 is the Gruneisen gamma defined as:

$$\gamma_0 = \frac{m}{\rho_o}\left(\frac{dp}{dU}\Big|_V\right) \tag{3-36}$$

where the derivative of pressure with respect to energy is at constant volume, V. By setting all the constants except c to zero and $\mu_{comp} = \rho/\rho_0 - 1$ (compression ratio), a linear model for the propagation of stress waves in the solid body is found:

$$p = c^2(\rho - \rho_o) \tag{3-37}$$

where ρ_0 is the reference state density and c is the isothermal sound speed in the material given by:

$$c = \sqrt{\frac{\partial p}{\partial \rho}\bigg|_T} \approx \sqrt{\frac{K + \frac{4}{3}G}{\rho}} = \sqrt{\frac{E(1-v)}{\rho(1+v)(1-2v)}} \tag{3-38}$$

The derivative of pressure with respect to density is taken considering constant temperature, T, K is the bulk modulus of the solid, v is Poisson's ratio, and E is the elastic modulus. Inspection of equation (3-37) shows that when the material is under compression (density increasing), the pressure will be positive, as expected. Neglecting the coefficients that multiply the higher order terms is a reasonable approximation for a weakly compressible formulation. In the FSW process, we have found that the compression ratio typically does not surpass 1%, for this reason, the linear assumption is valid. Note that a compression ratio of 1% would lead to a pressure of over 600 MPa, well in excess of the yield limit of most aluminum alloys. A comparison of the Gruneisen EOS for AA6061-T6 ($\gamma_0 = 1.97$, $a = 0.48$, $S_1 = 1.4$, $c = 4722 \, m/s$, [142] constant specific internal energy) and the linear EOS is provided in Figure 3-5. Note that for compression ratios of less than ~5%, the two curves are essentially identical.
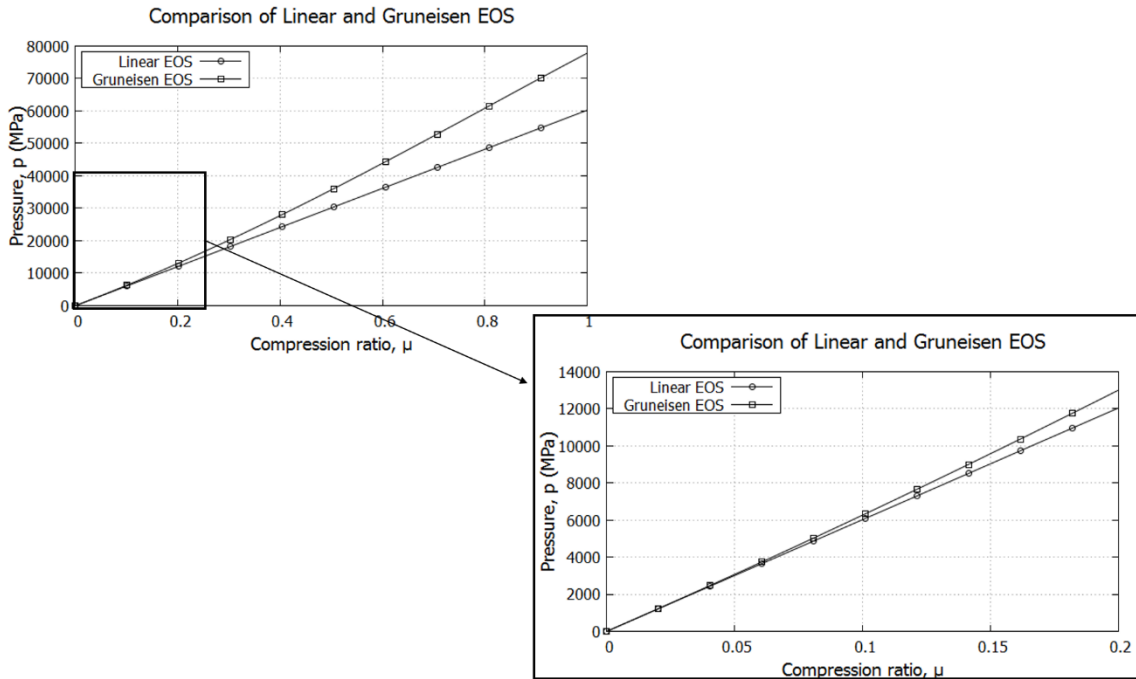
Figure 3-5 – Comparison of Linear and Gruneisen EOS

Since equation (3-29) involves the stress rate of the deviatoric stress, this equation must be integrated to obtain the stress at the current time step.

3.3.2 THERMAL EXPANSION

In the FSW process, thermal expansion plays an important role not only during the welding phases, but also plays a critical role once the weld has fully cooled, at which time the distortion and residual stresses will develop. For these reasons, thermal expansion is included in the developed simulation code. The total strain rate tensor is partitioned into an elastic rate ($\dot{\varepsilon}_e$) and a bulk isotropic thermal rate ($\dot{\varepsilon}_T$):

$$\dot{\bar{\bar{\varepsilon}}} = \dot{\bar{\bar{\varepsilon}}}_e + \dot{\bar{\bar{\varepsilon}}}_T \tag{3-39}$$

$$\dot{\bar{\bar{\varepsilon}}} = \frac{1}{2}(\nabla \otimes \bar{v} + (\nabla \otimes \bar{v})^T) + \alpha_{CTE}\frac{\partial T}{\partial t}\bar{\bar{I}} \tag{3-40}$$

where α_{CTE} is the coefficient of thermal expansion and $\partial T/\partial t$ is the time rate of change of temperature. Equation (3-30) shows that an increase in temperature leads to an increase in the total strain rate.

3.3.3 J2 PLASTICITY – VON MISES

The principle stresses play an important role in solid inelastic solid mechanics problems. They are the stresses that transform the stress state to the three principle planes, forming an orthogonal basis:

$$\bar{\bar{\sigma}} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \tag{3-41}$$

As can be seen from Equation (3-41), in 3D, there are three principle stresses; σ_1, σ_2, and σ_3. Determining the value of the stresses is an eigenvalue problem:

$$\left|\bar{\bar{\sigma}} - \lambda\bar{\bar{\delta}}\right| = -\lambda^3 + I_1\lambda^2 - I_2\lambda + I_3 = 0 \tag{3-42}$$

I_1, I_2, and I_3 are the first, second, and third stress invariants. Solving for the roots of the characteristic equation gives three eigenvalues. By convention the first principle stress is:

$$\sigma_1 = max(\lambda_1, \lambda_2, \lambda_3) \tag{3-43}$$

The third principle stress is:

$$\sigma_3 = min(\lambda_1, \lambda_2, \lambda_3) \qquad (3\text{-}44)$$

and the second principle stress will be the middle eigenvalue. A special case occurs when the stress state is hydrostatic, in this situation; the principle stresses will be equal to each other.

Analogously, the deviatoric stresses have invariants as well:

$$\left| \bar{\bar{S}} - \lambda \bar{\bar{I}} \right| = -\lambda^3 + J_1\lambda^2 - J_2\lambda + J_3 = 0 \qquad (3\text{-}45)$$

Solving this eigenvalue problem leads to the determination of the principle deviatoric stresses. The first and third deviatoric invariants have certain uses, however, the second invariant is often used as an effective stress measure when assuming a von Mises like yield surface. The second deviatoric invariant is:

$$J_2 = \frac{1}{2} \bar{\bar{S}} : \bar{\bar{S}} \qquad (3\text{-}46)$$

One of the best-known equivalent stress measures is that of von Mises. It is commonly used as a threshold for determining if a ductile material has undergone yielding. The von Mises stress, σ_v, is independent of the hydrostatic stress state in the material. As such, this stress metric is dependent only on the deviatoric stress state. Furthermore, plastic deformation is assumed to occur without material compression (incompressible). The von Mises stress is formulated form the second deviatoric stress invariant:

$$\sigma_v = \sqrt{3J_2} = \sqrt{\frac{3}{2} \bar{\bar{S}} : \bar{\bar{S}}} \qquad (3\text{-}47)$$

The assumption of von Mises plasticity is used for all the simulation work presented herein. Other plasticity approaches are possible, but not considered in this work.

3.3.4   RADIAL RETURN PLASTICITY

Plasticity is considered by using the radial return approach since J_2 plasticity with isotropic hardening is assumed. The goal of the stress update procedure is to find the plastic strain and deviatoric stress state at time step $k + 1$, given known values at time step k.
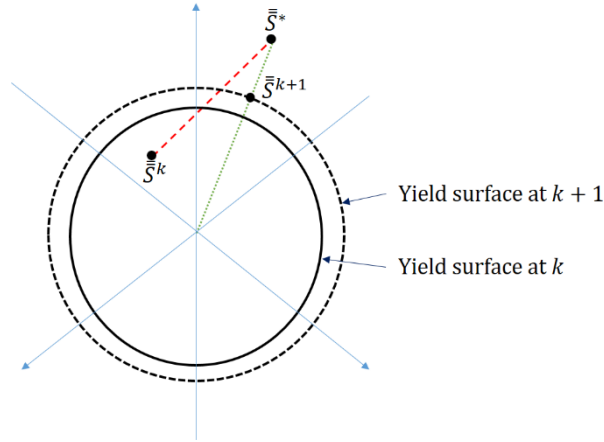
Figure 3-6 – Schematic of the update of the yield stress

The general idea is to assume that the material behaves elastically, and then check to see if the new stress state is outside of the yield surface. If it is, then the stresses are projected back onto the yield surface in the direction of the normal vector of the yield surface. In this sense, the radial return approach is a predictor-corrector method. One of the main underlying assumptions is that the plastic deformation phase is incompressible.

The process is shown schematically in Figure 3-6. For an elastic-plastic material with isotropic hardening, a trial stress (σ^{trial}) can be found from the deviatoric stress by initially assuming elastic behavior:

$$\sigma^{trial} = \sqrt{\frac{3}{2} \bar{\bar{S}}^* : \bar{\bar{S}}^*}$$

(3-48)

where $\bar{\bar{S}}^*$ is the deviatoric stress tensor from the elastic update. If the trial stress surpasses the yield stress (σ_y) of the material, the effective plastic strain increment is found from:

$$\Delta \varepsilon^p = \frac{\sigma^{trial} - \sigma_y{}^k}{3G + E_P}$$

(3-49)

G is the shear modulus, $E_P = \frac{E\,E_T}{E - E_T}$ is the hardening modulus, and E_T is the tangent modulus. The effective plastic strain is updated as:

$$\varepsilon^{p\,k+1} = \varepsilon^{p\,k} + \Delta \varepsilon^p$$

(3-50)

Note that for von Mises plasticity, the effective plastic strain is defined as:

$$\varepsilon^p = \sqrt{\frac{2}{3} \bar{\bar{\varepsilon}}^p : \bar{\bar{\varepsilon}}^p}$$

(3-51)

Once the effective plastic strain increment is found, the yield stress is updated (shown schematically in Figure 3-7) by:

$$\sigma_y{}^{k+1} = \sigma_y{}^k + E_P \Delta \varepsilon^p$$

(3-52)

Then the deviatoric stresses are scaled back to the yield surface using:

$$\bar{\bar{S}}^{k+1} = \frac{\sigma_y{}^{k+1}}{\sigma^{trial}} \bar{\bar{S}}^*$$

(3-53)

This forms an efficient and robust approach to include a non-iterative means to determine the plastic deformation of the material. The radial return algorithm can be extended to non-linear hardening for the flow stress model developed in Section 3.10. A Tylor series expansion with linearization about the current time allows determination of the plastic strain increment without the need for iteration. In this sense, the tangent modulus is found from:

$$E_T = \frac{\partial}{\partial \varepsilon^p} \sigma_y(\varepsilon^p, \dot{\varepsilon}, T^*)$$

(3-54)

This approach is only possible so long as the time step is sufficiently small to treat the non-linear hardening as locally linear (this is the case since explicit time integration is used). More information on the algorithm and its development can be found from [5, 7, 196, 198-202].
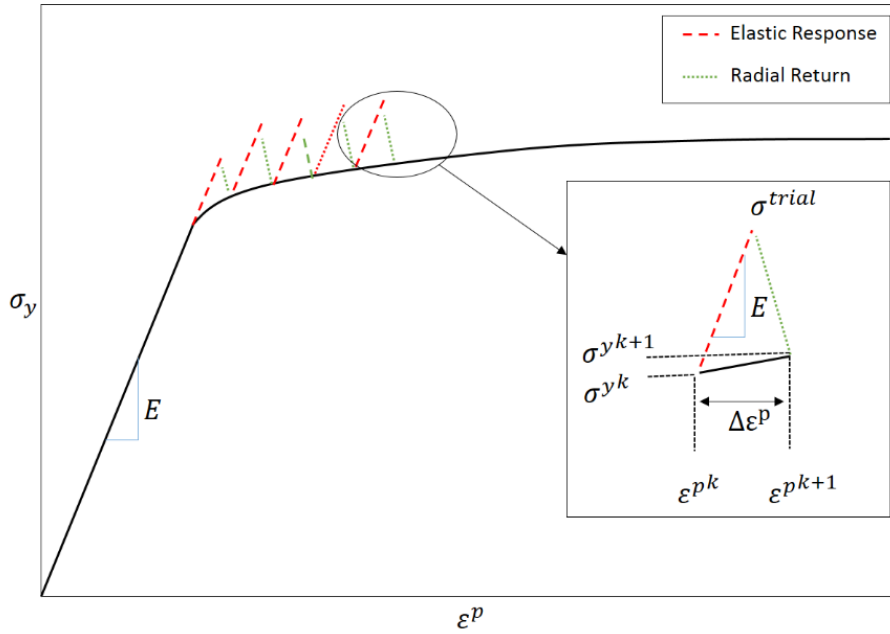
Figure 3-7 - Schematic of the update of the yield stress

3.4 SPH FORM OF THE CONTINUUM EQUATIONS

3.4.1 CONSERVATION OF MASS

By using directly equation (3-11), the discrete SPH equation for the conservation of mass can be written as:

$$\langle f(x_i{}^\alpha)\rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\alpha) W(r,h), \qquad where \; f(x_j{}^\alpha) = \rho(x_j{}^\alpha) = \rho_j \tag{3-55}$$

Making the appropriate simplifications, the result is:

$$\rho_i = \sum_{j=1}^{N_i} m_j W_{ij} \tag{3-56}$$

This is known as the summation density approach. It has the advantage that it directly conserves mass at the particle level. However, because of the nature of the particle approximation inconsistency, this leads to an underestimate of the density at particles on or near the boundary of the domain. A more common approach is to discretize the continuity equation (3-2). First start with a kernel estimate of the velocity divergence using Equations (3-14) and (3-12):

$$\langle \nabla \cdot f(x_i{}^\alpha) \rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\beta) \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\beta}, \qquad where \ f(x_j{}^\beta) = \bar{v}(x_j{}^\beta) = v_j{}^\beta \qquad (3\text{-}57)$$

Then substitute equation (3-57) into (3-2) to obtain:

$$\frac{d\rho_i}{dt} = -\rho_i \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} v_j{}^\beta \frac{\partial W_{ij}}{\partial x_i{}^\beta} \qquad (3\text{-}58)$$

Using equation (3-58) leads to edge effects due to an un-symmetrical particle-pair interaction. A better choice for the continuity equation can be found by using the following identity:

$$\langle \nabla \cdot \bar{f} \rangle \approx \langle \nabla \cdot \bar{f} \rangle - \bar{f} \langle \nabla \cdot 1 \rangle \qquad (3\text{-}59)$$

This then leads to a discretization that is symmetric at the particle-pair interaction level:

$$\frac{d\rho_i}{dt} = \rho_i \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} (v_i{}^\beta - v_j{}^\beta) \frac{\partial W_{ij}}{\partial x_i{}^\beta} \qquad (3\text{-}60)$$

Certainly, other discretizations are possible. Equation (3-60) is the form that is used in this research work as it is found to work well in the context of FSW simulation.

3.4.2   CONSERVATION OF MOMENTUM

Again, using Equations (3-13), (3-12), and (3-60), the discrete form of momentum conservation (Equation (3-5)) could be written with a naïve use of Equation (3-14) to get a kernel estimate for the stress divergence:

$$\langle \nabla \cdot f(x_i{}^\alpha) \rangle = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f(x_j{}^\beta) \frac{\partial W_{ij}}{\partial \bar{x}_i{}^\beta}, \qquad where \ f(x_j{}^\beta) = \bar{\bar{\sigma}}(x_j{}^\beta) = \sigma_j{}^{\alpha\beta} \qquad (3\text{-}61)$$

Equation (3-61) is then substituted into Equation (3-5) to get:

$$\frac{dv_i{}^\alpha}{dt} = \rho_i \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \sigma_j{}^{\alpha\beta} \frac{\partial W_{ij}}{\partial x_i{}^\beta} \qquad (3\text{-}62)$$

However, Equation (3-62) is a poor choice for the stress divergence for the same reason that was explained for Equation (3-58). A better choice is to develop the Equation of motion using the discrete Lagrangian, $\mathcal{L}$:

$$\mathcal{L} = \sum_{j\epsilon V} m_j \left[\frac{1}{2}\left(v_j^{\alpha}\right)^2 - u_j \right] \tag{3-63}$$

Next, the Euler-Lagrange equation is used:

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial v_i^{\alpha}}\right) - \frac{\partial\mathcal{L}}{\partial x_i^{\alpha}} = 0, \qquad \frac{\partial\mathcal{L}}{\partial v_i^{\alpha}} = m_i v_i^{\alpha}, \qquad \frac{\partial\mathcal{L}}{\partial x_i^{\alpha}} = -\sum_{j\epsilon V} m_j \frac{\partial u_j}{\partial \rho_j}\bigg|_s \frac{\partial \rho_j}{\partial x_j^{\alpha}} \tag{3-64}$$

where the change in internal energy with respect to density is at constant entropy, s. After providing relationships for entropy, kinetic energy, and density gradient (see Price [203] for full details), the Euler equation for linear momentum is:

$$\frac{dv_i^{\alpha}}{dt} = -\sum_{j=1}^{N_i} m_j \left(\frac{p_i}{\rho_i^2} + \frac{p_j}{\rho_j^2}\right)\frac{\partial W_{ij}}{\partial x_i^{\beta}} \tag{3-65}$$

Equation (3-65) considers only the effects of internal forces due to hydrostatic pressure (not isochoric). However, for the solid mechanics formulation, the effects of the volume preserving stresses (deviatoric stress) must be included as well. The isochoric deformation can be included as well following the same principles:

$$\frac{dv_i^{\alpha}}{dt} = \sum_{j=1}^{N_i} m_j \left(\frac{S_i^{\alpha\beta}}{\rho_i^2} + \frac{S_j^{\alpha\beta}}{\rho_j^2}\right)\frac{\partial W_{ij}}{\partial x_i^{\beta}} - \sum_{j=1}^{N_i} m_j \left(\frac{p_i}{\rho_i^2} + \frac{p_j}{\rho_j^2}\right)\frac{\partial W_{ij}}{\partial x_i^{\beta}} \tag{3-66}$$

Using Equation (3-32), the version of the equation of motion used in this work is:

$$\frac{dv_i^{\alpha}}{dt} = \sum_{j=1}^{N_i} m_j \left(\frac{\sigma_i^{\alpha\beta}}{\rho_i^2} + \frac{\sigma_j^{\alpha\beta}}{\rho_j^2}\right)\frac{\partial W_{ij}}{\partial x_i^{\beta}} \tag{3-67}$$

The discrete form of the energy equation is cast into SPH form starting from Equation (3-8) and omitting the $\nabla \cdot \bar{q}$ term (for the moment):

$$\frac{du_i}{dt} = \frac{1}{2}\sum_{j=1}^{N_i} m_j \left(\frac{p_i}{\rho_i{}^2} + \frac{p_j}{\rho_j{}^2}\right)\left(v_i{}^\beta - v_j{}^\beta\right)\frac{\partial W_{ij}}{\partial x_i{}^\beta} + \frac{1}{\rho_i} S_i{}^{\alpha\beta} \dot{\varepsilon}_i{}^{\alpha\beta}$$

(3-68)

Note that Equation (3-68) excludes the thermal energy portion of the total energy balance. Including this term will require significant development, and is the subject of Section 3.9. This will be broached in detail in section 3.9. For a transient solid mechanics problem, this Equation (3-68) will commonly be used to monitor the evolution of internal energy in comparison to kinetic energy. The ratio of kinetic to internal energy should remain small when using mass or time scaling. This will ensure that the results are not spoiled by excessive added inertial effects. However, the equation is insufficient to idealize the physics of the FSW process.

3.5 SPH FORM OF THE CONSTITUTIVE EQUATIONS

3.5.1 STRAIN RATE TENSOR

The strain rate, $\dot{\varepsilon}_i{}^{\alpha\beta}$, in a continuum solid is due to a deformation field caused by the rate of change of the material coordinates ($\bar{v}$) in the element. The strain is a rank 2 tensor with 6 non-identical terms (symmetric tensor), the strain can be written using the SPH approach as:

$$\dot{\varepsilon}_i{}^{\alpha\beta} = \frac{1}{2}\sum_{j=1}^{N_i}\left(\frac{m_j}{\rho_j} v_{ji}{}^\alpha \frac{\partial W_{ij}}{\partial x_i{}^\beta} + \frac{m_j}{\rho_j} v_{ji}{}^\beta \frac{\partial W_{ij}}{\partial x_i{}^\alpha}\right)$$

(3-69)

3.5.2 SPIN TENSOR

The spin tensor is needed in order to preserve rigid body motion of the particle system. The spin tensor is the difference between the divergence of the velocity field and the divergence of the transpose of the velocity field. Written in the SPH formulation, the spin tensor is:

$$\Omega_i{}^{\alpha\beta} = \frac{1}{2}\sum_{j=1}^{N_i}\left(\frac{m_j}{\rho_j} v_{ji}{}^\alpha \frac{\partial W_{ij}}{\partial x_i{}^\beta} - \frac{m_j}{\rho_j} v_{ji}{}^\beta \frac{\partial W_{ij}}{\partial x_i{}^\alpha}\right)$$

(3-70)

3.5.3 JAUMANN RATE EQUATION

The stress state can be updated in the material using a frame indifferent objective stress-rate equation. Many different stress rate equations that can be used; such as Truesdell, Green-Nahgdi or the Jaumann rate equation (others exist). The Jaumann rate has a relatively simple formulation, thus making it unassuming to implement in a CSM code. The rate equation in SPH form is:

$$\dot{S}^{\alpha\beta} = 2G\left(\dot{\varepsilon}^{\alpha\beta} - \frac{1}{3}\delta^{\alpha\beta}\dot{\varepsilon}^{\gamma\gamma}\right) + S^{\alpha\gamma}\Omega^{\beta\gamma} + \Omega^{\alpha\gamma}S^{\gamma\beta} \tag{3-71}$$

$\dot{S}^{\alpha\beta}$ is the time rate of change of the deviatoric stress, G is the shear modulus of the material, $\dot{\bar{\bar{\varepsilon}}}$ and $\bar{\bar{\Omega}}$ are the strain rate (from Eq. (3-69)) and spin tensor (from Eq. (3-70)). $\delta^{\alpha\beta}$ is the Kronecker delta and the $\frac{1}{3}\delta^{\alpha\beta}\dot{\varepsilon}^{\gamma\gamma}$ term is known as the mean stress, or the hydrostatic stress.

3.6 ARTIFICIAL VISCOSITY

The SPH momentum and energy equations are often written with an additional term called artificial viscosity, Π_{ij}. The concept of introducing an artificial viscosity into the hydrodynamics equations was originated by von Neumann and Richtmyer [204]. The general idea is to smooth a shock front across multiple calculation elements (removes the singularity associated with a shock). A similar approach has been adapted for SPH by Monaghan [205]. This approach was subsequently found to be an adequate way to alleviate certain stability issues. The acceleration vector associated with the artificial viscosity is given by:

$$\frac{dv_i^\alpha}{dt}\bigg|_{AV} = -\sum_{j=1}^{N_i} m_j(\Pi_{ij})\frac{\partial W_{ij}}{\partial x_i^\beta} \tag{3-72}$$

$$\Pi_{ij} = \begin{cases} \dfrac{-\alpha_\Pi c_{ij}\phi_{ij} + \beta_\Pi \phi_{ij}^2}{\rho_{ij}} & if \quad v_{ij}^\alpha x_{ij}^\alpha < 0 \\ 0 & otherwise \end{cases} \tag{3-73}$$

$$\phi_{ij} = \frac{h_{ij}v_{ij}^\alpha x_{ij}^\alpha}{\left|x_{ij}^\alpha\right|^2 + 0.1h_{ij}} \tag{3-74}$$

$$c_{ij} = \frac{1}{2}(c_i + c_j), \qquad \rho_{ij} = \frac{1}{2}(\rho_i + \rho_j), \qquad h_{ij} = \frac{1}{2}(h_i + h_j) \tag{3-75}$$

$$v_{ij}{}^{\alpha} = v_i{}^{\alpha} - v_j{}^{\alpha}, \qquad x_{ij}{}^{\alpha} = x_i{}^{\alpha} - x_j{}^{\alpha} \tag{3-76}$$

In this research work, no SPH artificial viscosity is used. The introduction of the artificial viscosity term can lead to excessive dissipation if great care is not exercised [183]. Nevertheless, it is included above for completeness.

3.7 VELOCITY AVERAGING USING THE XSPH APPROACH

The position of the particles can be updated by using the computed nodal velocities according to:

$$\frac{d\bar{x}}{dt} = \bar{v} \tag{3-77}$$

As usual, $\bar{x}$ is a position vector and $\bar{v}$ is the particle velocity vector. This equation can readily be integrated with a typical integration scheme (see section 3.15 for more details). For CSM type simulations, Monaghan [206] has shown that it is preferred to update the particle positions using the XSPH method. The average particle velocities are used for the position update. Equation (3-77) is altered to be:

$$\frac{d\bar{x}}{dt} = \bar{v} + \breve{v} \tag{3-78}$$

where $\breve{v}$ is an average particle velocity. The equation can be written in the discrete SPH formulation by using equation (3-11) and (3-59):

$$\breve{v}_i{}^{\alpha} = \zeta_{XSPH} \sum_{j=1}^{N_i} \frac{m_j}{\rho_{ij}} (v_j{}^{\alpha} - v_i{}^{\alpha}) W_{ij} \tag{3-79}$$

$$\frac{dx_i{}^{\alpha}}{dt} = v_j{}^{\alpha} + \breve{v}_i{}^{\alpha} \tag{3-80}$$

where $\rho_{ij} = (\rho_i + \rho_j)/2$. ζ_{XSPH} is a constant that is used to adjust the amount of velocity averaging that can range from 0.0 (no velocity averaging) to 0.5 (maximum recommended value). The XSPH approach helps to prevent particles from penetrating one another. The approach also helps to keep free surface elements from separating from the body. The interpolation kernel (W_{ij}) does not necessarily need to be the same as that used for the conservation equations, although, in practice it is advisable to do so. This

position update technique still conserves linear and angular momentum and is of $\mathcal{O}(h^2)$, which is consistent with the order of the conservation equations. The updated velocity should also be used in the conservation equations.

3.8 SOME COMMON ISSUES WITH THE SMOOTHED PARTICLE METHOD

SPH is a very powerful numerical method for solving partial differential equations. The method suffers from certain deficiencies, however. The main issues are related to the use of collocation to perform the spatial integration. The more prominent issues with SPH are a lack of completeness, consistency, tensile instability, as well as zero energy modes (not discussed here).

3.8.1 COMPLETENESS AND CONSISTENCY

SPH method was originally developed for astrophysics problems (infinite domain). Because of this, the discrete form of the conservation equations did not include the surface integrals (to resolve the boundary conditions). For finite domains, the SPH method suffers from incomplete interpolation, this concept is shown schematically in Figure 3-8.
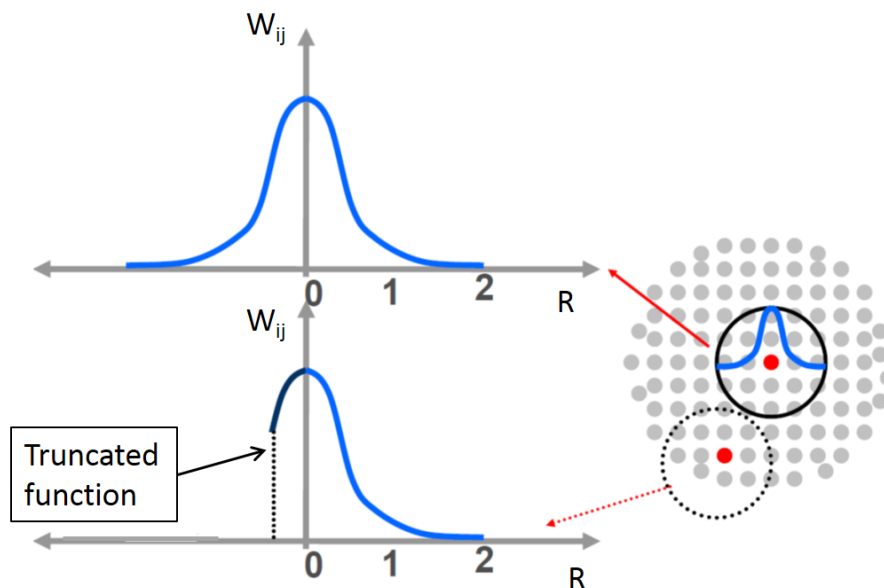
Figure 3-8 – Incomplete interpolation (adapted from Xu [207])

Here, a particle located on or near the boundary of the domain will not have a complete set of neighbors, leading to an underestimate of the value of the function at that point.

The consistency of a numerical method is measured by the ability of the method to reproduce a function of nth degree. A method that is able to reproduce at best a constant function is said to have zero-order consistency. In its standard form, SPH does not have zero-order consistency. The easiest way to see this is to try to reproduce the density field in an aluminum cube by using equation (3-56). Each particle has a mass of 3.6 kg, a 6x6x6 grid with equal spacing of 0.1 m is used.
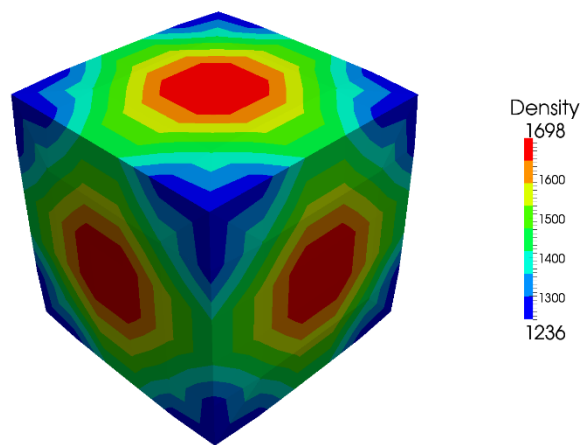
Figure 3-9 – Incomplete interpolation example

Although the calculated density should be 2700 kg/m^3, the result is far from that. The calculated density varies between 1236 and 1698 kg/m^3 as shown in Figure 3-9. In this example, the particle results are projected onto a smooth triangulated surface (see section 4.6 for more details on triangulation).

Various approaches to restore the zero-order particle consistency and to improve completeness have been proposed by other authors. The ghost particle approach of Libersky *et al.* [208] is popular. In this method, additional particles are provided on the domain boundaries to complete the neighborhood of the boundary particles. This method is interesting; however, it is difficult to implement for very large deformation problems.

Another approach that is able to reproduce constant functions was introduced by Libersky *et al.* [209] as well as Bonet and Lok [185]. With their approach, the value of a constant function can be calculated using a kernel re-normalization method:

$$f_i^{\alpha} = \frac{\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} f_j^{\alpha} W_{ij}}{\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} W_{ij}} \tag{3-81}$$

Using the improved interpolation approach of equation (3-81), the density field in the aluminum cube can be found by setting $f_i^{\alpha} = \rho_i$ and $f_j^{\alpha} = \rho_j$ (note that $\alpha = 1$ for a scalar). Figure 3-10 shows the improved results for the density calculation. Now the correct value of 2700 kg/m^3 is found throughout the entire domain.
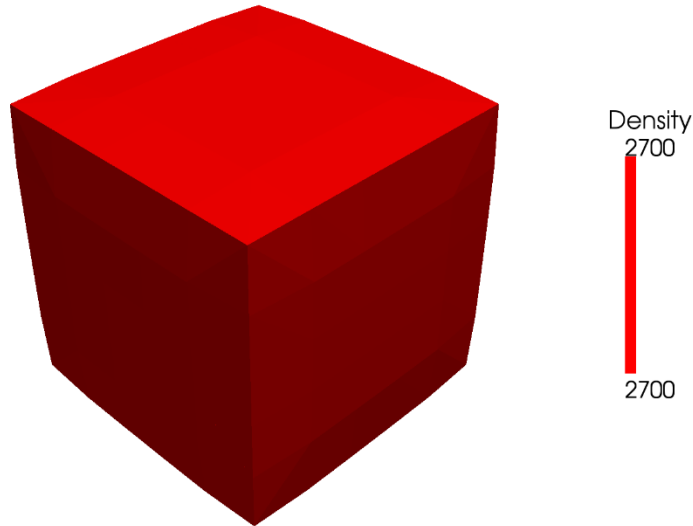
Figure 3-10 – Correct density field with re-normalization approach

The kernel re-normalization approach is very useful for obtaining the correct value of an interpolated value. However, this approach cannot be used to correct the kernel gradients that are used in equations (3-55) to (3-68). For this, the SPH method needs to be first-order consistent (reproduce linear functions). Bonet and Kulasegaram [210] have proposed a kernel gradient correction. Their method requires re-normalizing the kernel gradient to remove the boundary deficiency. The kernel is modified as follows:

$$\nabla \widehat{W}_{ij} = \bar{\bar{M}}_i^{-1} \nabla W_{ij} \tag{3-82}$$

where $\bar{\bar{M}}$ is a the kernel gradient modifying tensor defined as:

$$\bar{\bar{M}}_i = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \nabla W_{ij} \otimes (\bar{x}_i - \bar{x}_j) \tag{3-83}$$

When the correction is applied to an internal particle, $\bar{\bar{M}}$ will be equal to the identity matrix. As such, no correction will be introduced for these particles. For a particle near the domain boundary, $\bar{\bar{M}}$ will be different from the identity matrix. This will provide an anisotropic correction to the kernel. This method has application for small to moderate deformation; however, with a large deformation; the correction becomes ill defined and leads to divergence in the solution procedure. Attempts to use the kernel gradient correction for FSW have so far proved ineffective.

3.8.2   TENSILE INSTABILITY – ARTIFICIAL STRESS

The tensile instability has been a problem plaguing the SPH method. According to Monaghan [211], the problem is caused by a combination of a piecewise linear second derivative of the cubic B-spline smoothing function and the use of an Eulerian kernel function (using the current configuration for evaluation of field equations) for a Lagrangian method. The problem manifests as a pairing or "clumping" of particles as can be seen in Figure 3-11. Numerically, the problem is caused by the pairwise interactions. When the stress between two particle pairs is positive (tension), the force between the pairs is attractive. In a real atomistic pair, attractive and repulsive forces will provide for a stable configuration. In the standard SPH method, the attractive force between particle pairs is not balanced; this results in instability. Swegle *et al.* [212] provides an in-depth analysis of the tensile instability problem.
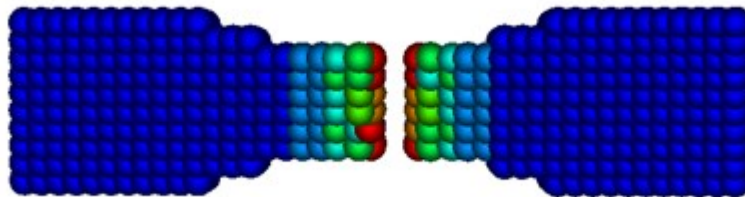
Figure 3-11 – Tensile instability

One approach to remedy the tensile instability was proposed by Vignjevic and Reveles [213]. They use a total Lagrangian approach to provide a stable solution of the conservation equations. The total Lagrangian formalism completely removes the instability because the field variables are now evaluated based on a reference position (start of simulation). The approach is very good for small to moderate levels of plastic deformation. However, for the extent of plastic deformation seen in FSW, this approach will not be beneficial. Since there is material mixing in the FSW process, field equations should be evaluated using the current configuration. Using a reference configuration would introduce large error and inaccuracy in the simulation.

In order to preserve the ability to treat large plastic deformation, Monaghan [211] and Gray *et al.* [214] have proposed a simple and effective solution to the tensile instability problem. They have added an artificial stress term in the momentum equation to remove the instability. The artificial stress acts as a repulsive force when the particle pair force is due to tension:

$$\left.\frac{dv_i{}^\alpha}{dt}\right|_{AS} = \sum_{j=1}^{N_i} m_j \left(R_{ij}{}^{\alpha\beta} f^n\right) \frac{\partial W_{ij}}{\partial x_i{}^\beta} \tag{3-84}$$

here f is a function that increases as the separation distance between two particles decreases. Monaghan proposes the function to be of the form:

$$f = \frac{W(r_{ij})}{W(\Delta p)} \tag{3-85}$$

$W(r_{ij})$ is an appropriate smoothing function that does not need to be the same as that used in the evaluation of the conservation equations, r_{ij} is the radial distance between two particles, and $W(\Delta p)$ is the value of the chosen smoothing function evaluated at the average particle spacing. The artificial stress term, $R_{ij}{}^{\alpha\beta}$, is defined as:

$$R_{ij}{}^{\alpha\beta} = \frac{R_i{}^{\alpha\beta}}{\rho_i{}^2} + \frac{R_j{}^{\alpha\beta}}{\rho_j{}^2} \tag{3-86}$$

$$R_i{}^{\alpha\beta} = \begin{cases} -\epsilon_{ASM} \sigma_i{}^{\alpha\beta}, & \sigma_i{}^{\alpha\beta} > 0 \\ 0, & \sigma_i{}^{\alpha\beta} \leq 0 \end{cases} \tag{3-87}$$

$$R_j{}^{\alpha\beta} = \begin{cases} -\epsilon_{ASM}\sigma_j{}^{\alpha\beta}, & \sigma_j{}^{\alpha\beta} > 0 \\ 0, & \sigma_j{}^{\alpha\beta} \le 0 \end{cases} \tag{3-88}$$

ϵ_{ASM} is a constant that depends on the simulation and is, typically, between 0.1 and 0.5. One of the drawbacks to this approach is that the stress state for each particle pair must be monitored. A conditional statement is required to assign values to $R_{ij}{}^{\alpha\beta}$, which increases the computational time. Nevertheless, the method is a viable option for the large plastic strains present during the FSW process. Using equations (3-84) to (3-88), the same tension problem can now be solved without the instability. Figure 3-12 shows the same tension test as in Figure 3-11, this time the instability has been removed and a uniform stress field is calculated.
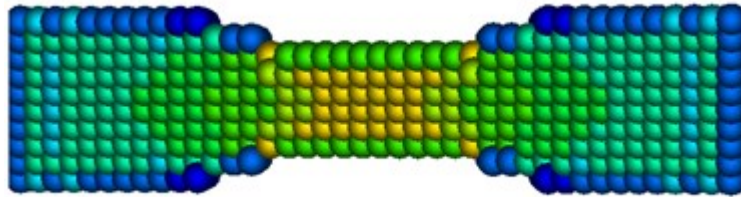
Figure 3-12 – Tensile instability removed

Since the artificial stress method preserves the SPH ability to treat large plastic deformation, this approach is well suited to simulation of FSW. Generally, the stress field is primarily compressive in FSW and no special measures are required. In some situations, most notably in the bobbin tool FSW process, there is a strong tension field as the tool is entering the work pieces. Another common case is during cooling, the weld zone residual stresses are tensile. In such cases, the artificial stress approach can be used to stabilize the solution.

3.9   HEAT TRANSFER

In the FSW process, the evolution of the temperature field within the work pieces (and, to a certain extent, in the support structure and tool) is of primary importance. This falls at the heart of the reason why FSW is possible for common engineering materials; as the energy state in the material increases, the material softens. This allows the material to be easily deformed, allowing the two work pieces to mix and form a strong mechanical joint. FSW would not be possible if these common materials did not soften due to an

increase in temperature. For this reason, thermal energy must now be included in the total energy balance. Recalling equation (3-8):

$$\rho \frac{DU}{Dt} = \bar{\bar{\sigma}}(\nabla \cdot \bar{v}) - \nabla \cdot \bar{q} + \dot{q} \tag{3-89}$$

Evidently, expressions for the divergence of the heat flux as well as the source term must be developed. First off, the heat flux is related to the temperature field, T, through:

$$\bar{q} = -k\nabla T \tag{3-90}$$

where k is the thermal conductivity of the material. By assuming that the link between energy and temperature is simple, one can write:

$$U = C_p T \tag{3-91}$$

where C_p is the heat capacity of the material. The $\bar{\bar{\sigma}}(\nabla \cdot \bar{v})$ term is not retained since the $\nabla \cdot \bar{v}$ term is the velocity divergence, which is non-zero only in the case of compressible materials. In the case of a material undergoing irreversible plastic deformation, the plastic phase is considered incompressible. The $\bar{\bar{\sigma}}(\nabla \cdot \bar{v})$ term would account for elastic compressions. However, in a weakly compressible hydrodynamic approach, there can be local non-physical variations in the density of the material, which can lead to a non-physical change of energy. For the aforementioned reason, the $\bar{\bar{\sigma}}(\nabla \cdot \bar{v})$ is omitted. With this in mind, the energy equation can be stated as:

$$\rho C_p \frac{\partial T}{\partial t} = \nabla \cdot (k\nabla T) + \dot{q} \tag{3-92}$$

With this approach, the various heat sources such as plastic and friction work, and dissipation such as convection and radiation, will be included in the $\dot{q}$ term. Although radiation is insignificant for the aluminum work pieces (emissivity less than 0.1), some of the simulations presented in this work involve the use of a flat black paint (emissivity of ~0.95) on the top surface of the work pieces, which makes the radiation term significant. Using the product rule property:

$$\nabla \cdot (\phi \bar{F}) = \nabla \phi \cdot \bar{F} + \phi(\nabla \cdot \bar{F}) \tag{3-93}$$

where ϕ is a scalar and $\bar{F}$ is a vector and the discrete SPH Laplace operator [215] is:

$$\langle \nabla^2 f_i \rangle = -2 \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} (f_j - f_i) \tilde{x}_{ij} \frac{\partial W_{ij}}{\partial x_i^\beta}, \quad where \ \tilde{x}_{ij} = \frac{x_i - x_j}{|x_i - x_j|^2} = \frac{x_{ij}}{|x_{ij}|^2} \tag{3-94}$$

here f is a scalar field. Take $\phi = k$, $\bar{F} = \nabla T$, $f = T$, and further expand (3-93) to:

$$\nabla \cdot (k \nabla T) = \frac{1}{2} (k \nabla^2 T + T \nabla^2 k - T \nabla^2 k + k \nabla^2 T) \tag{3-95}$$

After substitution of (3-94) into (3-93), simplification and rearrangement, the discrete SPH equation for heat transfer is:

$$\frac{dT_i}{dt} = \frac{2}{\rho_i C_{p_i}} \left[\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} K_{ij} (T_i - T_j) \frac{x_{ij}}{|x_{ij}|^2} \frac{\partial W_{ij}}{\partial x_i^\beta} + \dot{q}_i \right] \tag{3-96}$$

Some research groups [216, 217] have used the average thermal conductivity:

$$K_{ij} = \frac{1}{2} (k_i + k_j) \tag{3-97}$$

However, Cleary and Monaghan [218] have shown that using the harmonic mean:

$$K_{ij} = 2 \left(\frac{k_i k_j}{k_i + k_j} \right) \tag{3-98}$$

improves the stability of the algorithm for large variations in thermal conductivities and even in the presence of discontinuities. So then, equation (3-96) is cast as:

$$\frac{dT_i}{dt} = \frac{1}{\rho_i C_{p_i}} \left[\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \frac{4 k_i k_j}{k_i + k_j} \frac{(T_i - T_j)}{|x_{ij}|^2} x_{ij} \frac{\partial W_{ij}}{\partial x_i^\beta} + \dot{q}_i \right] \tag{3-99}$$

This form is used to model heat transfer in this research work. To this point, a means to calculate the diffusion of heat from one point to another has been presented; as of yet, nothing has been stated regarding the boundary conditions or the source terms. The total heat source term per unit volume will be composed of plastic work, $\dot{q}_{PW}$, friction work, $\dot{q}_{FW}$, surface convection, $\dot{q}_{conv}$, and surface radiation, $\dot{q}_{rad}$:

$$\dot{q} = \dot{q}_{PW} + \dot{q}_{FW} + \dot{q}_{conv} + \dot{q}_{rad}$$

(3-100)

Although frictional heating, convection, and radiation are surface integrals, these terms can be approximated as volume integrals without any loss of precision for the FSW simulations. The error introduced by converting the surface integrals to volume integrals is negligible in comparison to the other sources of error and uncertainty. In some cases, discretizing the surface integrals with the SPH formulation can be beneficial, see the Appendices for details.

3.9.1   PLASTIC WORK

First, the contribution of plastic work will be considered, starting with the definition of work, W, along a curvilinear path, Γ:

$$W = \int_{\Gamma} \bar{F} \cdot d\bar{x}$$

(3-101)

In the SPH code, the time integration scheme guarantees that the line integral will be reasonably approximated due to the very small time-steps (on the order of microseconds). This allows (3-101) to be written as:

$$W = \bar{F} \cdot \Delta\bar{x}$$

(3-102)

$\Delta\bar{x}$ is the incremental change in position of an SPH element from the previous time step to the current one and $\Delta\bar{x} = \bar{x}^{t+dt} - \bar{x}^{t}$. Next, recognize that the stress, $\bar{\bar{\sigma}}$, in a solid body is related to the force and that the incremental displacement is related to the strain, $\bar{\bar{\varepsilon}}$, so that:

$$W = \bar{\bar{\sigma}} : d\bar{\bar{\varepsilon}}$$

(3-103)

This can be generalized as being the area under the stress strain curve for the solid body. The total strain ($\bar{\bar{\varepsilon}}$) can be partitioned into elastic ($\bar{\bar{\varepsilon}}^{e}$) and plastic ($\bar{\bar{\varepsilon}}^{p}$):

$$\bar{\bar{\varepsilon}} = \bar{\bar{\varepsilon}}^{e} + \bar{\bar{\varepsilon}}^{p}$$

(3-104)

Although the true area under the curve would consider the area under the elastic and plastic portions of the curve, for metals, the area under the elastic portion is negligible. Taking rates on both sides of (3-103) leads to the equation for the plastic power:

$$\dot{W}_{PW} = \bar{\bar{\sigma}} : \frac{d\bar{\bar{\varepsilon}}^p}{dt} \tag{3-105}$$

Note that the $d\bar{\bar{\sigma}}/dt : \bar{\varepsilon}^p$ term is omitted form Equation (3-105) since it is many orders of magnitude smaller than the $d\bar{\varepsilon}^p/dt : \bar{\sigma}$ term. Finally, the heat source term due to plastic work, is:

$$\dot{q}_{PW} = \dot{W}_{PW} = \chi_{PW} \left(\bar{\bar{\sigma}} : \frac{d\bar{\bar{\varepsilon}}^p}{dt} \right) \left[\frac{J}{m^3 \; s} \right] \tag{3-106}$$

The factor χ_{PW} is the Taylor-Quinney factor, which is typically set between 0.8 and 1.0 depending on the situation. Since the formulation is based on the plastic work, the χ_{PW} factor is most aptly set to 1.0.

3.9.2 FRICTION WORK

Next, the relationship for friction work can be developed starting form equation (3-101), noting that the friction force that does work is the tangential force, $\bar{F}_T$, and taking rates on each side, the friction power is:

$$\dot{W}_{FW} = \bar{F}_T \cdot \frac{d\bar{x}_T}{dt} \tag{3-107}$$

here, the $d\bar{x}_T/dt$ term represents the incremental movement of an SPH particle due to friction in the tangential direction per time step. As in Section 3.9.1, the same argument holds for the omission of the $d\bar{F}_T/dt \cdot \bar{x}_T$ term. This rate of change of the position of an SPH element is the same as the relative tangential velocity, $\bar{v}_{\text{Rel}_T}$, of the ith SPH element with respect to the jth element on the sliding surface:

$$\bar{v}_{\text{Rel}_T} = v_{T_{ij}}{}^{\alpha} = \frac{\Delta \bar{x}_T}{\Delta t} = \bar{v}_{T_i} - \bar{v}_{T_j} \tag{3-108}$$

Using equations (3-107) and (3-108), the friction power is succinctly described by:

$$\dot{W}_{FW} = \bar{F}_T \cdot \bar{v}_{\text{Rel}_T} \tag{3-109}$$

This power is developed at the interface between the two contacting bodies. By considering each of the bodies as semi-infinite solids (see [219] for full details), the ratio of power that goes into the ith body is:

$$\lambda_i = \frac{\sqrt{k_i C_{p_i} \rho_i}}{\sqrt{k_i C_{p_i} \rho_i} + \sqrt{k_j C_{p_j} \rho_j}} \tag{3-110}$$

Perfect thermal contact is assumed to exist between the two bodies. This is a reasonable assumption since the workpiece becomes very plastic and tends to fill in the asperities in the tool surface. The λ_i parameter is required since the work pieces are discretized with meshfree elements. Since the calculation point is not physically at the free surface of the body, λ_i is needed to distribute the power in to the meshfree calculation points. In a sense, λ_i is akin to the penalty factor in Eq. (3-155). In the case of imperfect thermal contact, a thermal resistance can be incorporated by adding a factor, β_{FW}, that can degrade the friction power as a function of the contact resistance. The final form of the heat source due to friction is:

$$\dot{q}_{FW} = \frac{\beta_{FW} \lambda_i}{V} \dot{W}_{FW} = \frac{\beta_{FW} \lambda_i}{V} \left(\bar{F}_T \cdot \bar{v}_{Rel_T} \right) \left[\frac{J}{m^3 \, s} \right] \tag{3-111}$$

where $V = m/\rho$, is the volume of the SPH element. In this work, all simulations are performed with $\beta_{FW} = 1.0$.

3.9.3 FINDING THE FREE SURFACE ELEMENTS

Before treating SPH thermal boundary conditions, an approach is needed to determine the free surface of an SPH domain. This is very important for setting up a general tactic to describe surface heat flux and convection. In order to determine the particle surface normal, the fact that particles on the surface of the SPH domain have an incomplete set of neighbors can be used. A vector $(x_{COM_i}{}^{\alpha})$ between the ith particle and the center of the particle cluster by:

$$x_{COM_i}{}^{\alpha} = \frac{1}{M} \sum_{j=1}^{N_i} m_j x_{ij}{}^{\alpha} \tag{3-112}$$

This can be thought of as a normal vector pointing out of the solid at each surface particle. The total mass of the particle cluster is M, the mass of the jth particle is m_j, and the distance vector between the ith and the jth particle is $x_{ij}{}^{\alpha}$. One can immediately recognize that the particles that are within the body of the solid will essentially have a null value for $x_{COM_i}{}^{\alpha}$. This normal vector for the ith SPH element will be important for the thermal boundary conditions, it is found by normalizing the center of mass:

$$\hat{n}_{SPH} = \frac{\bar{x}_{COM}}{|\bar{x}_{COM}|} \tag{3-113}$$

Next, the particles that reside on the surface of the solid domain must be determined. This can be accomplished by comparing the length of the normal vector to the smoothing length of the particle and to a specific number of neighbors (ξ).

$$Surface_{node} = \begin{cases} 1, & x_{COM_i} \geq \frac{1}{4}h_i \text{ and } N_i \leq \xi \\ 0, & Otherwise \end{cases} \tag{3-114}$$

In this work $\xi = 46$ is used. In this manner, the particles that are on the surface of the domain will be tagged with a value of one and the internal particles will be tagged with a value of zero. One can see how well the algorithm works for an arbitrary shaped domain in Figure 3-13. The left side of the image shows the normal vectors and the right side shows only the surface particles. Other research groups have used similar methods. Marrone *et al.* [220] use an algorithm that is ideal for fluid simulations. They use a two-step method, first particles that are close to the free surface are found by using the properties from the renormalization matrix, $\bar{\bar{M}}_i$ (as described in section 3.8.1). Then, in the second step, they evaluate the geometric properties of the SPH elements found in the first step. Their process requires scanning a conical region beyond the surface elements. Their method supposes that the renormalization approach has been used. This method would be significantly more complicated to implement in parallel.

Normal Vectors
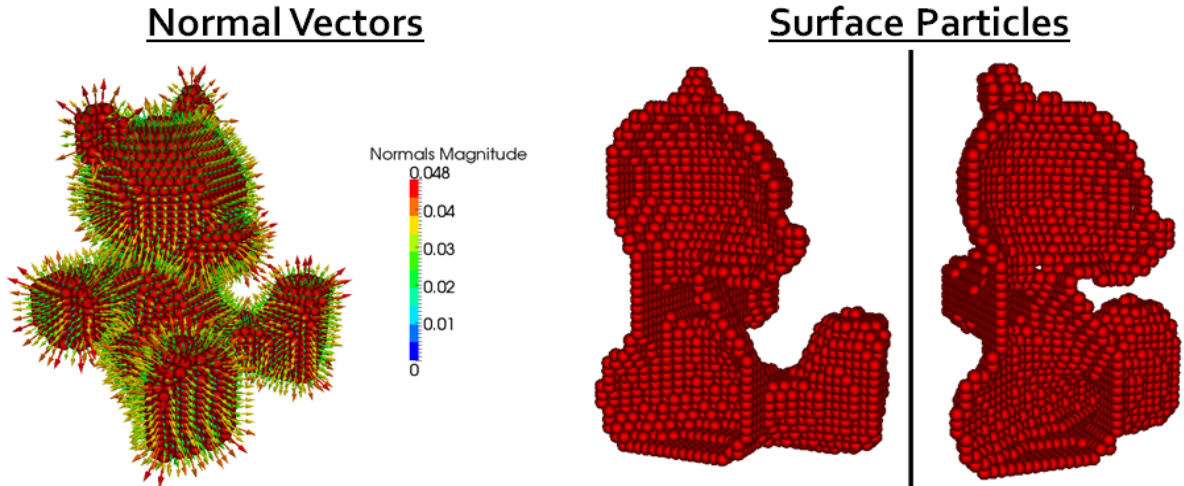
Surface Particles

Figure 3-13 – Normal Vectors and Surface Particles – Left: Normal vectors; right: Free surface particles

Other authors such as Randles and Libersky [190] have used a "discrete color" approach. The color of the particles is evaluated by exploiting the completeness of the interpolation of a SPH element. If the interpolation is incomplete, the element is assigned a color that associates it as a boundary particle.

The proposed method in this work is much less complicated, as the algorithm simply needs to find the center of mass of each neighbor group. This is accomplished with a very simple equation that is evaluated very quickly. This approach is very efficient and can easily be performed in parallel on the GPU.

3.9.4   VISUALIZING HEAT FLUX VECTORS

Determining flow of heat within the solid can be important for visualizing the heat flow. The heat flux vectors can also be used to transfer heat to an external system. For example, SPH can be thermally coupled to FEM using this approach. The heat flux vector from the FEM and from the SPH can be determined. These "external" heat fluxes can then be included in the energy balance. A simple approach to approximate the heat flux is given by:

$$\bar{q}_i{}^\alpha = k_i \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \left(T_i - T_j\right) \frac{\partial W_{ij}}{\partial x_i{}^\beta} \tag{3-115}$$

This approach will lead to a heat flux that is incorrect on the border of the SPH domain because of the truncation of the influence domain of the surface elements. An effective method to correct these vectors is to use the filter previously described in equation (3-81). Alternatively, the kernel gradient correction approach shown in equation (3-82) could be used so that equation (3-115) would become:

$$\bar{q}_i{}^\alpha = k_i \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \left(T_i - T_j \right) \widehat{\nabla W}_{ij} \tag{3-116}$$

Equation (3-116) does not include heat transfer through the surface boundary, indeed only thermal conduction in the body is taken into consideration. Equation (3-116) is not used in the code to perform a thermal energy balance, it is only meant as a post-processing tool to approximately visualize the heat flow.
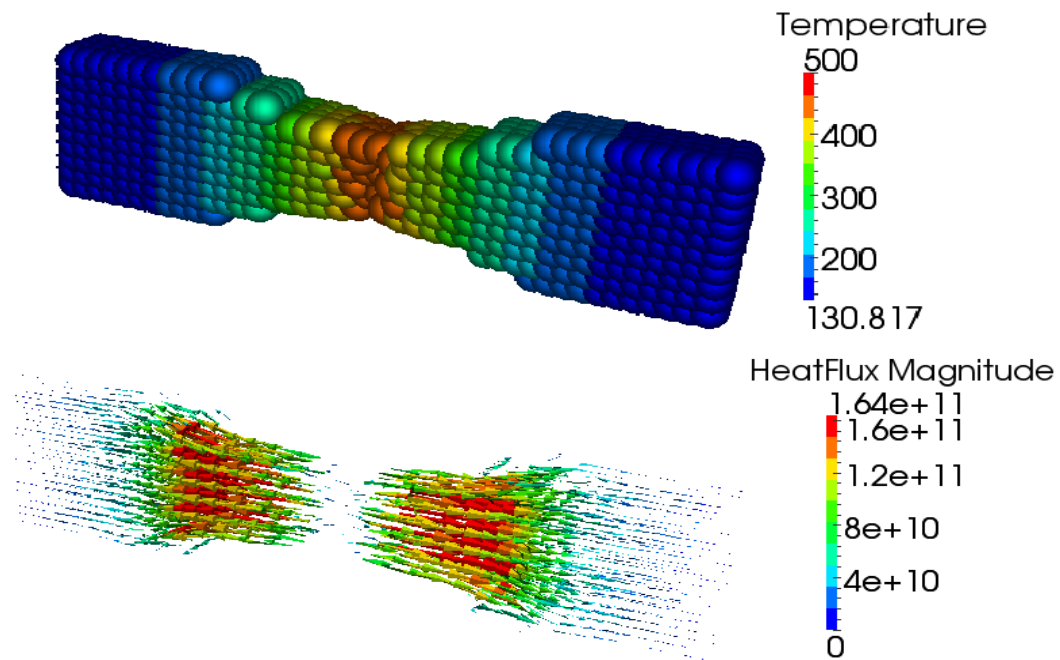
Figure 3-14 – Heat flux vectors

An example of heat flux vectors is shown in Figure 3-14 for a plastically deforming tensile specimen. Note that the temperature is increasing in the center of the specimen where there is plastic deformation. In addition, the heat flux vectors are showing a null heat flow through all the surfaces; this is because

the particles on the free surface do not have elements to transfer heat to outside of the body; as such, the surface material points are naturally treated adiabatically.

3.9.5   SURFACE CONVECTION

Convection plays an important role in dissipating heat from the work pieces, support structure, and the FSW tool surfaces. The heat flux due to convection is:

$$\bar{q}_{conv} = h_{conv}(T_\infty - T_s)\hat{n}_{SPH} \tag{3-117}$$

where h_{conv} is the coefficient of convection, T_∞ is the ambient temperature, and T_s is the surface temperature. A schematic of the boundary condition is shown in Figure 3-15. The temperature increase per unit time at the boundary will be:

$$\frac{dT}{dt} = \frac{\bar{q}_{conv} \cdot \hat{n}_{SPH}}{mC_p} A_s = \frac{h_{conv}A_s(T_\infty - T_s)}{mC_p} \tag{3-118}$$

Convection (Robin BC)

$$-k\frac{\partial T}{\partial x}\bigg|_{x=0} = h(T_\infty - T_s)$$

$$h, T_\infty$$

$$T_s$$

$$T(x, t)$$

$$x$$
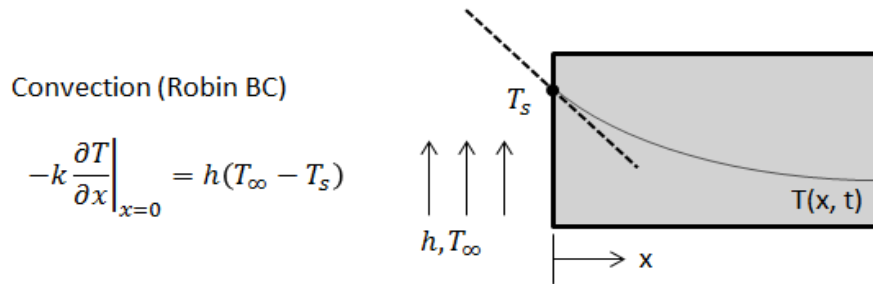
Figure 3-15 – Surface convection boundary condition
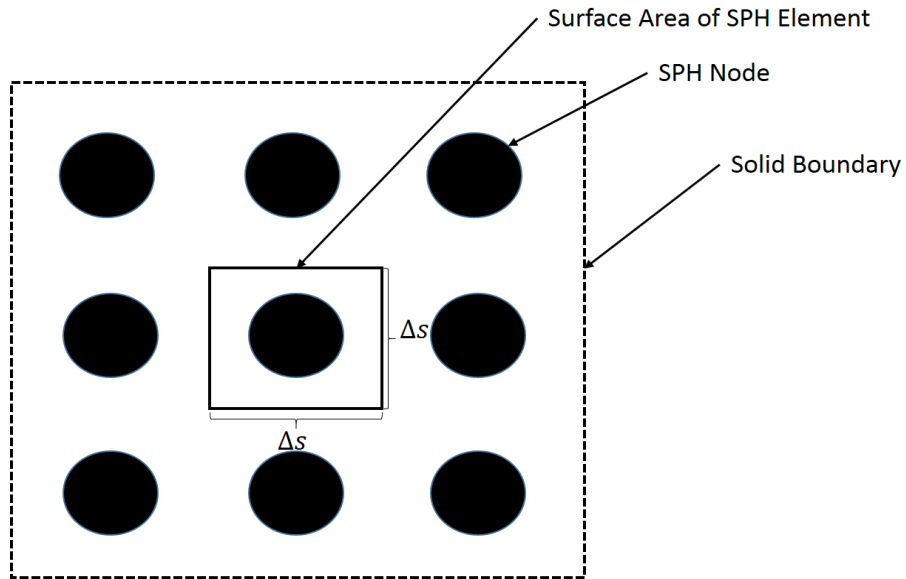
Figure 3-16 – Equivalent surface area of an SPH element

The surface area, A_s, is related to the equivalent surface area of an SPH element. Think of a square with edges that are equidistant from the center of the element with sides equal to the particle spacing as shown in Figure 3-16. The particle spacing, Δs, can be found from:

$$\Delta s = \sqrt[3]{V} = \sqrt[3]{\frac{m}{\rho}}$$

(3-119)

Now, the temperature rate is:

$$\frac{dT}{dt} = \frac{h_{conv}\Delta s^2(T_\infty - T_s)}{mC_p}$$

(3-120)

This can be cast into a convection heat source term:

$$\dot{q}_{conv} = \frac{\rho C_p h_{conv}\Delta s^2(T_\infty - T_s)}{mC_p} = \frac{\rho h_{conv}\Delta s^2(T_\infty - T_s)}{m} \quad \left[\frac{J}{m^3\,s}\right]$$

(3-121)

The convection source term is only applied to the elements on the free surface (found according to section 3.9.3) of the work piece domain.

3.9.6   SURFACE RADIATION

The emissivity, ϵ_r, of aluminum is very low (~0.1 [219]); typically radiation would not need to be considered for the aluminum work pieces for this reason. However, in this research work, the surfaces of the work pieces are painted black to improve the quality of the image that can be obtained with an infrared camera (Thermal camera). Since the plates are painted, radiation must be considered in the energy balance. Following a similar development as for surface convection and noting that:

$$\bar{q}_{rad} = \epsilon_r \sigma_{SB} \left(T_{surr}{}^4 - T_s{}^4 \right) \hat{n}_{SPH} \tag{3-122}$$

where σ_{SB} is the Stefan-Boltzmann constant, T_{surr} is the temperature of the surroundings, and T_s is the surface temperature. The heat source term for radiation is then:

$$\dot{q}_{rad} = \frac{\rho C_p \epsilon_r \sigma_{SB} A_s \left(T_{surr}{}^4 - T_s{}^4 \right)}{m C_p} = \frac{\rho \epsilon_r \sigma_{SB} \Delta s^2 \left(T_{surr}{}^4 - T_s{}^4 \right)}{m} \quad \left[\frac{J}{m^3 \, s} \right] \tag{3-123}$$

The radiation source term is only applied to the elements on the free surface (found according to section 3.9.3) of the work piece domain.

3.10 FKS FLOW STRESS MODEL

The material behavior of aluminum is a topic that has been extensively studied by a great number of research groups. The extent of the available flow stress models is astounding. In this section, a new flow stress model is developed and presented that is ideally suited for the large range of plastic deformation common in the FSW process. The model is composed of three main parts: strain hardening, strain rate stiffening, and thermal softening. The new model will be called the Fraser-Kiss-St-Georges flow stress model (FKS). Although the model has been tailored to the behaviour of AA6061-T6 aluminum alloy, it can easily be adapted to other homogeneous isotropic engineering materials.

3.10.1   EXPERIMENTAL WORK – COMPRESSION TESTING

A set of compression tests was recently conducted using a Gleeble dynamic material research system (shown in Figure 3-17) at high and low strain rates and different temperatures. The aluminum (AA6061-

T6) test samples were machined (long axis transverse to the rolling direction) from the bar stock that is used for the FSW experiments.

Figure 3-17 - Gleeble dynamic material research system

The cylinders had initial diameters, d_o of 10 mm and lengths, l_o, of 15 mm. This size was chosen since the material was being machined form 12.7 mm plate stock. Constant strain rate and temperature runs were used since this compares closely to typical flow stress curves for material testing. A sample temperature history is shown for 150°C, 350°C, and 500°C for $\dot{\varepsilon} = 1.0$ in Figure 3-18.

Note that the temperature is controlled well by the Gleeble machine within the test region. The test region is selected for each test to ensure that the strain rate and temperature are constant. The Gleeble system uses Joule heating to heat uniformly the cylinder during testing. This type of heating ensures a uniform heat source; however, there can be minute temperature variation in the longitudinal and radial directions. Thermocouples were welded to the surface of the cylinder at the ends and center to ensure a uniform temperature distribution. Typically, for a tension testing program, once the specimen starts to neck, the joule heating will cause the necked region to increase in temperature drastically. However, in the case of a compression test, the temperature remains more stable throughout the test. This can be shown by comparing the temperature at the center and end of the specimen for the 150°C case (shown in Figure 3-19, T1 and T2 are at the end and center of the specimen respectively).

Temperature Comparison - Strain Rate 1.0
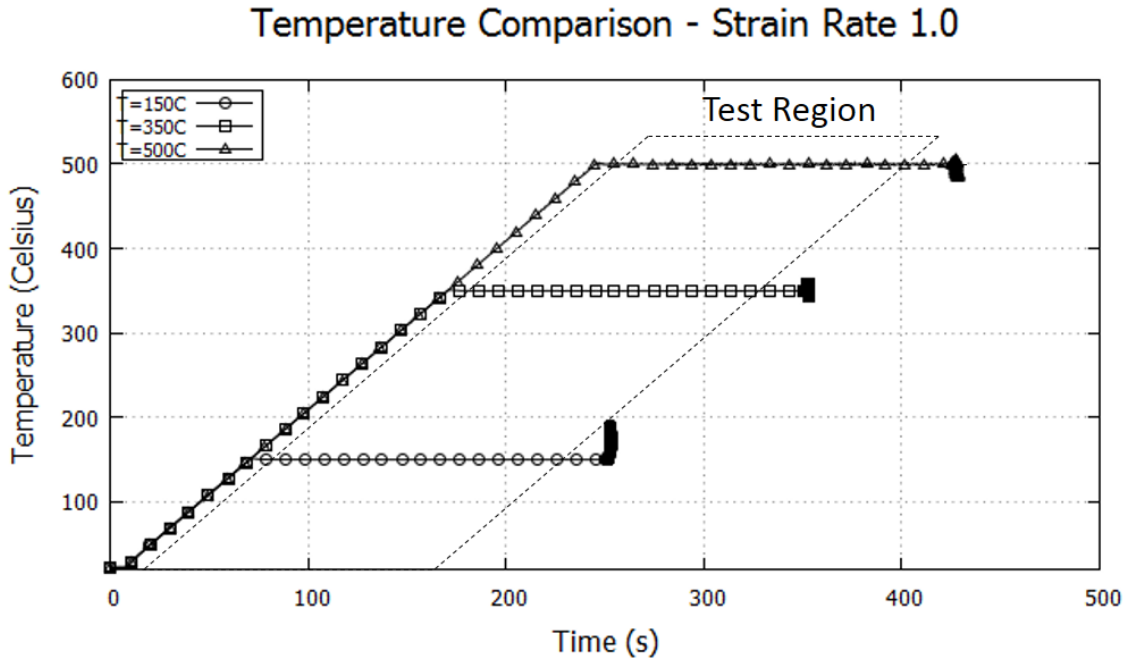
Figure 3-18 – Temperature comparison at center of specimen for $\dot{\varepsilon} = 1.0$

Thermocouple comparison - Strain Rate 0.001 T=150C
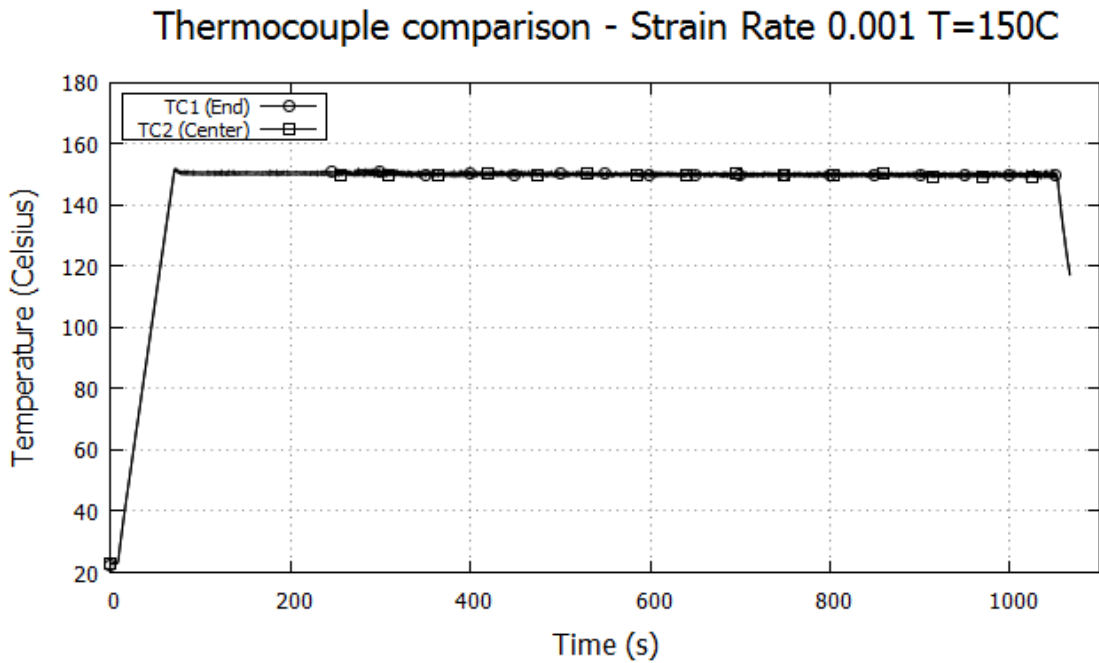
Figure 3-19 – Thermocouple comparison during compression testing ($\dot{\varepsilon} = 0.001$, T=150°C)

In all cases, true stress as a function of true strain was used. Engineering stress, σ, and strain, ε, are converted to true stress, σ_{True}, and strain, ε_{True} using the change in gauge length, Δl:

$$\varepsilon_{True} = \ln\left(\frac{l_o + \Delta l}{l_o}\right) = \ln(1 + \varepsilon) \tag{3-124}$$

$$\sigma_{True} = \sigma\, e^{\varepsilon_{True}} = \sigma(1 + \varepsilon) \tag{3-125}$$

The flow stress curves in this section are presented as true stress as a function of true plastic strain. The plastic strain is determined by shifting the flow stress curves to remove the elastic portion, ε^e, of the flow curves:

$$\varepsilon^p = \varepsilon_{True} - \varepsilon^e = \varepsilon_{True} - \frac{\sigma_y}{E} \tag{3-126}$$

The results from the compression tests did not fit well into the currently available flow stress models (such as Johnson-Cook, Hansel-Spittel, Norton-Hoff, etc.). The main reason was that the compression results showed no strain rate sensitivity of the aluminum at room temperature. However, at elevated temperatures, the material did exhibit strain rate stiffening. Dorbane *et al.* [140] found similar results in their recent work. For this reason, a new flow stress model has been developed that takes into account strain hardening, thermal softening and strain rate effects.
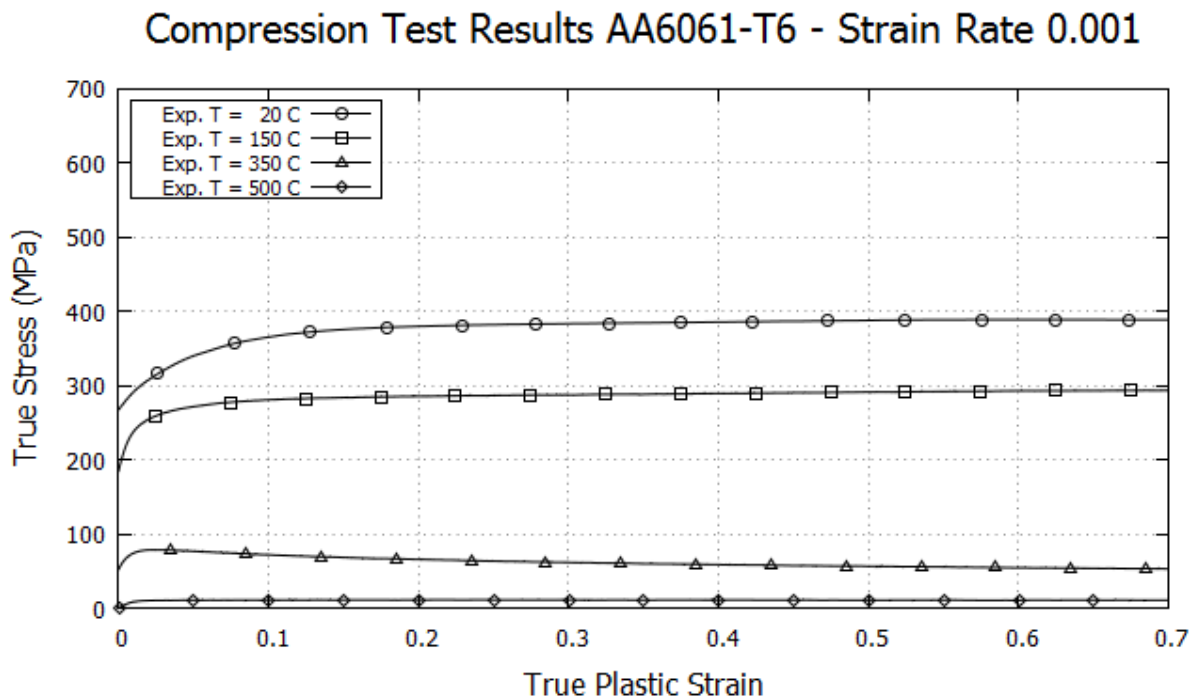
Figure 3-20 – Compression test results: $\dot{\varepsilon} = 0.001$

Kirk Fraser - April 2017
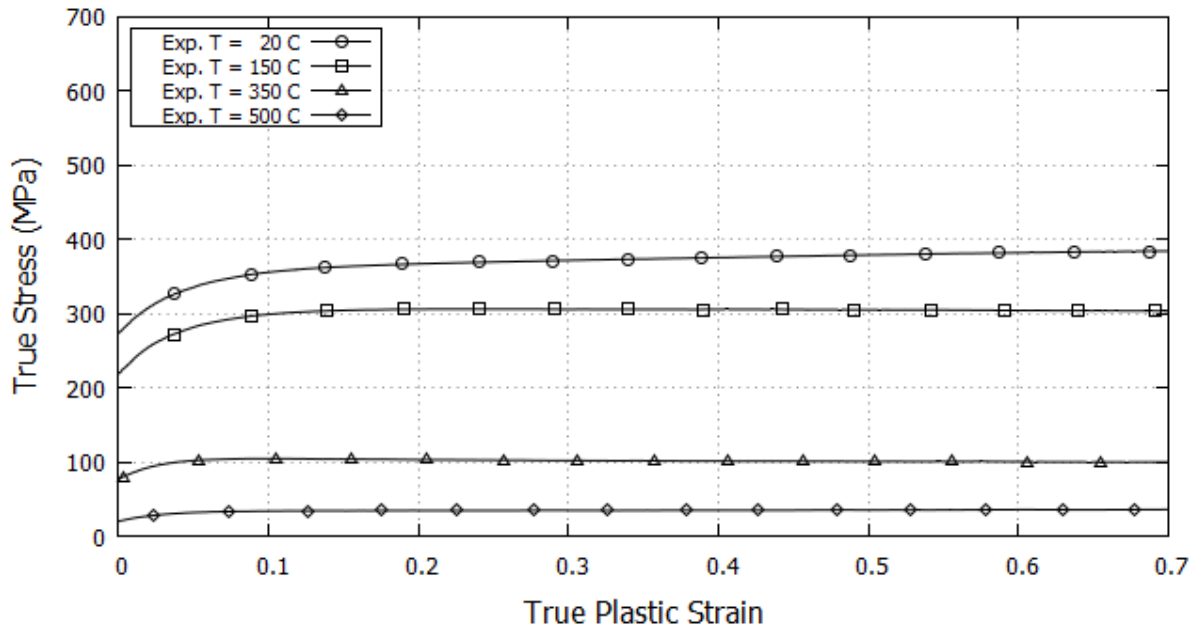
Figure 3-21 - Compression test results: $\dot{\varepsilon} = 1.0$
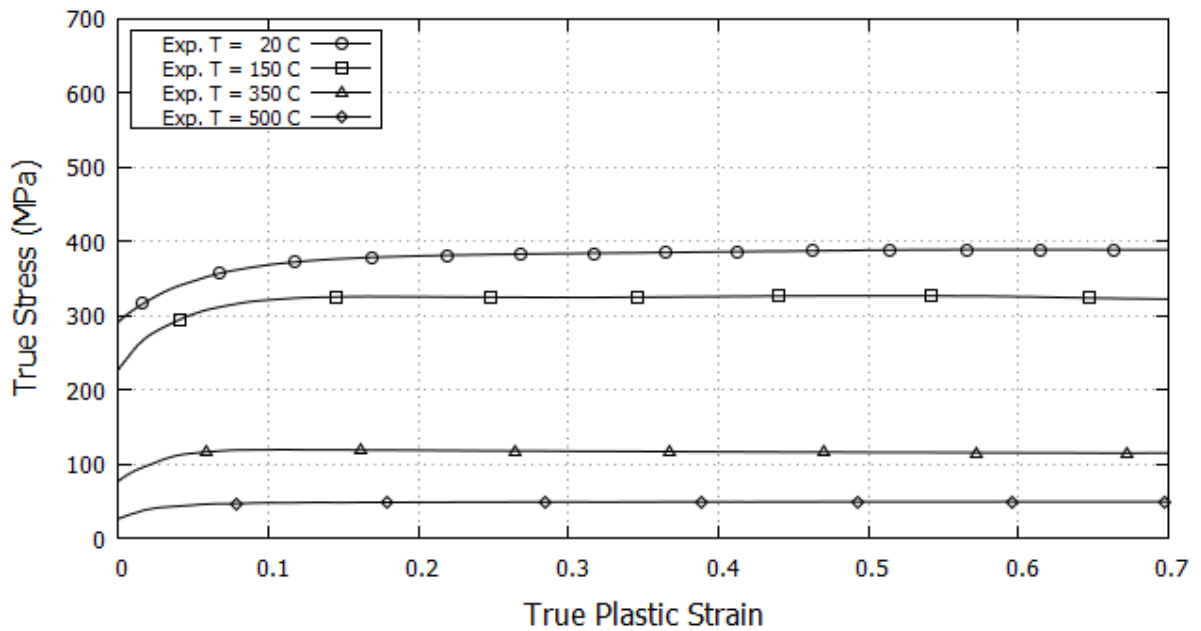
Figure 3-22 - Compression test results: $\dot{\varepsilon} = 10.0$

The final shape of the compressed test specimens for $\dot{\varepsilon} = 0.001$ is shown in Figure 3-23. Notice the elliptical bulging shape of the samples at 20°C and 350°C. This shows that there is a certain amount of orthotropy in the AA6061-T6 plate stock (strong axis in rolling direction). Taking into account such orthotropic behaviour in the constitutive equations would drastically increase the solution time due to the material mixing during the FSW process. An orthotropic model could be developed in the future, but is out of the scope of this research work.
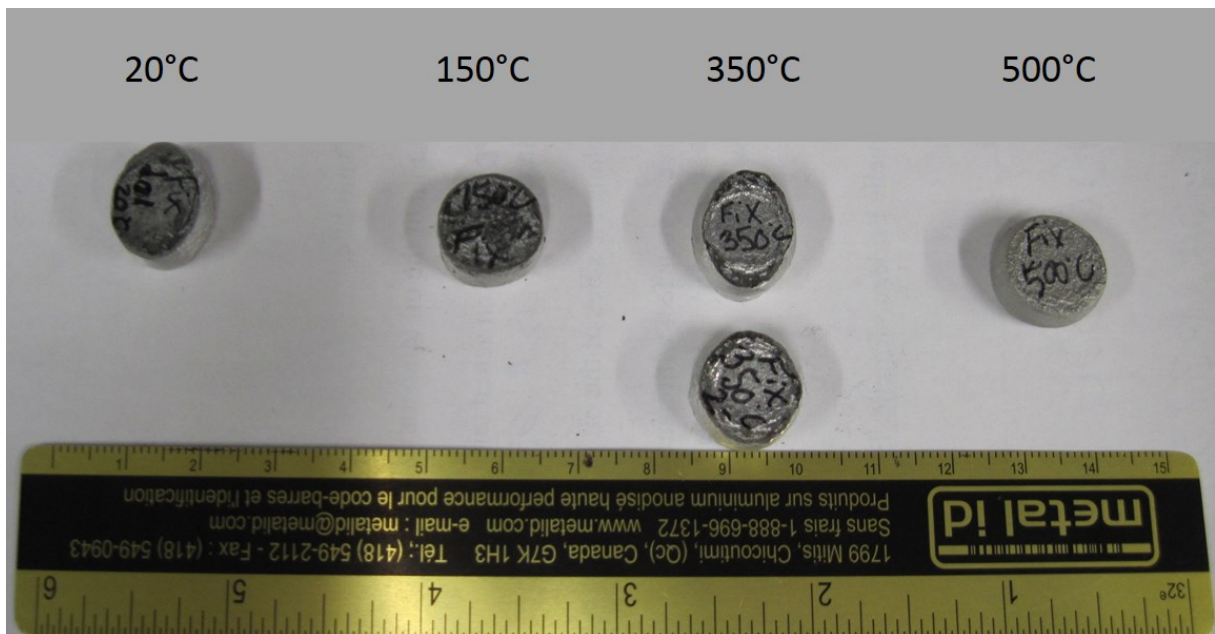
Figure 3-23 – Final shape of the compressed cylinders ($\dot{\varepsilon} = 0.001$)

3.10.2   STRAIN HARDENING EFFECT

The results from the compression tests showed that the flow stress typically attained a plateau. For this reason, an arctangent formulation is proposed to account for the strain hardening, $H(\hat{\varepsilon}^p)$, portion of the new flow law:

$$H(\varepsilon^p) = a_1 + a_2 tan^{-1}(a_3 \varepsilon^p)$$

(3-127)

3.10.3   STRAIN RATE EFFECT

The most important realization of the compression testing campaign was that the AA6061-T6 material exhibited no strain rate sensitivity at room temperature; however, at higher temperatures, the strain rate

did play a role in the flow stress. A function is proposed, $\Lambda(\dot{\varepsilon}, T^*)$, that is a combination of the homologous temperature, and the strain rate:

$$\Lambda(\dot{\varepsilon}, T^*) = 1 + \left[b_1 T^{*b_2}\right]\left[b_3 ln\left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_0}\right)\right] \tag{3-128}$$

The homologous temperature, T^* is

$$T^* = \frac{T - T_{melt}}{T_{melt} - T_{room}} \tag{3-129}$$

The reference strain rate, $\dot{\varepsilon}_0$, is the rate at which no strain rate effects should be attributed in the flow stress model.

3.10.4 THERMAL SOFTENING EFFECT

The thermal softening, $\Theta(T^*)$, aspect will be incorporated using a sigmoid type function called the general logistic function. This function is chosen, as it is completely customizable to obtain any type of thermal softening behavior:

$$\Theta(T^*) = 1 - \frac{1}{(1 + e^{-c_1 T^*})^{\frac{1}{c_2}}} \tag{3-130}$$

Note that the true general logistic function involves more constants; however, they are typically taken as unity and have been converted appropriately in eqn. (3-130).

3.10.5 THE FLOW STRESS MODEL

The resulting model is a multiplicative combination of the three terms:

$$\sigma_y(\varepsilon^p, \dot{\varepsilon}, T^*) = H(\varepsilon^p)\Lambda(\dot{\varepsilon}, T^*)\Theta(T^*) \tag{3-131}$$
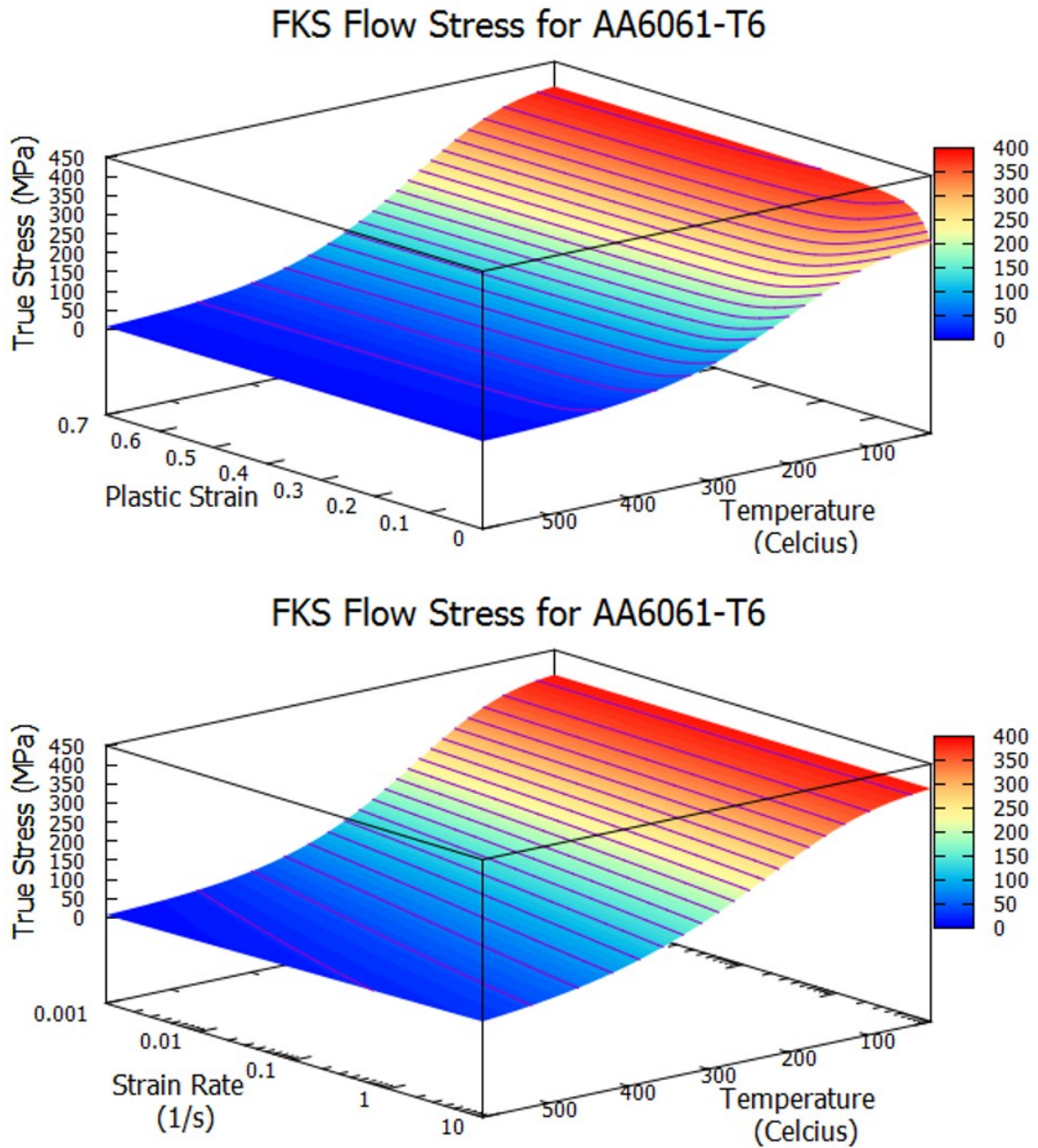
Figure 3-24 – FKS flow stress surface, $\sigma_y(\varepsilon^p, T)$ upper, $\sigma_y(\dot{\varepsilon}, T)$

A comparison of the proposed flow stress model against the experimental data is shown in Figure 3-26. Figure 3-24 shows the yield as a function of plastic strain and temperature (above) and yield as a function of strain rate and temperature (below). There is an excellent correlation between the proposed model and the experimental data. The constants used in the model are provided in Table 3-2. With these values,

the R^2 value is 0.97, showing that the model is excellently suited for the flow stress of the AA6061-T6 alloy. The value of a_1 can be attributed to the initial yield stress of the material at room temperature. a_2 is a strain hardening factor that can be seen to be the maximum increase in flow stress due to strain hardening at room temperature and at the reference strain rate. a_3 controls the time for the flow stress to plateau. b_1 and b_2 control the effect of temperature on the strain rate sensitivity. Since the value of T^* is zero at room temperature, no strain rate effects are incorporated. c_1 and c_2 control the shape of the sigmoid logistic function.
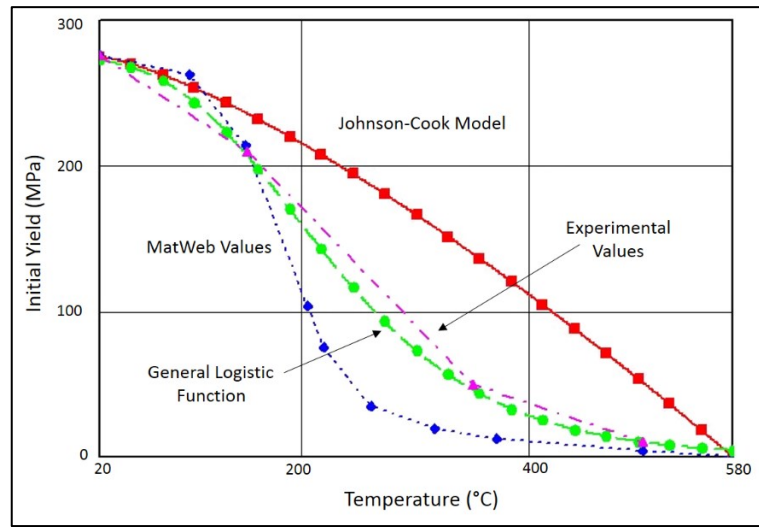
Figure 3-25 – Comparison of the yield values

In order to show the effect of the thermal softening in the proposed model, a comparison to the Johnson-Cook model and experimental values are shown in Figure 3-25. Note that the values presented by MatWeb show a faster decay than was found experimentally.

Table 3-2 – Constants used for proposed flow stress model

Constant	Value	Units	Constant	Value	Units	Constant	Value	Units
a_1	276.0	MPa	b_1	1.7	-	c_1	6.0	-
a_2	74.0	MPa	b_2	2.4	-	c_2	0.16	-
a_3	50.0	m/m	b_3	0.1	-			

The yield values will obviously be dependent on the manner in which the plate stock is produced (rolled, drawn, or extruded), the retailer, as well as the specific thermal treatment. For this reason, it would not be expected that the obtained values would be the same as MatWeb or even to another plate stock from a different manufacturer. By using the general logistic function, excellent correlation to the decay of the yield strength with increasing temperature is possible. The new flow stress model can be adapted to use the radial return (section 3.3.4) algorithm by defining the tangent to the curve (tangent modulus, E_T) as $\frac{\partial}{\partial \varepsilon^p} \sigma_y(\varepsilon^p, \dot{\varepsilon}, T^*)$:

$$E_T = \frac{\partial}{\partial \varepsilon^p} \sigma_y(\varepsilon^p, \dot{\varepsilon}, T^*) = \frac{a_2 a_3 \Lambda(\dot{\varepsilon}, T^*) \Theta(T^*)}{a_3{}^2 \varepsilon^{p2} + 1} \tag{3-132}$$

The relation between the hardening, E_P, and tangent modulus:

$$E_P = \frac{E\,E_T}{E - E_T} \tag{3-133}$$

This provides a reasonable explicit update of the plastic strain.
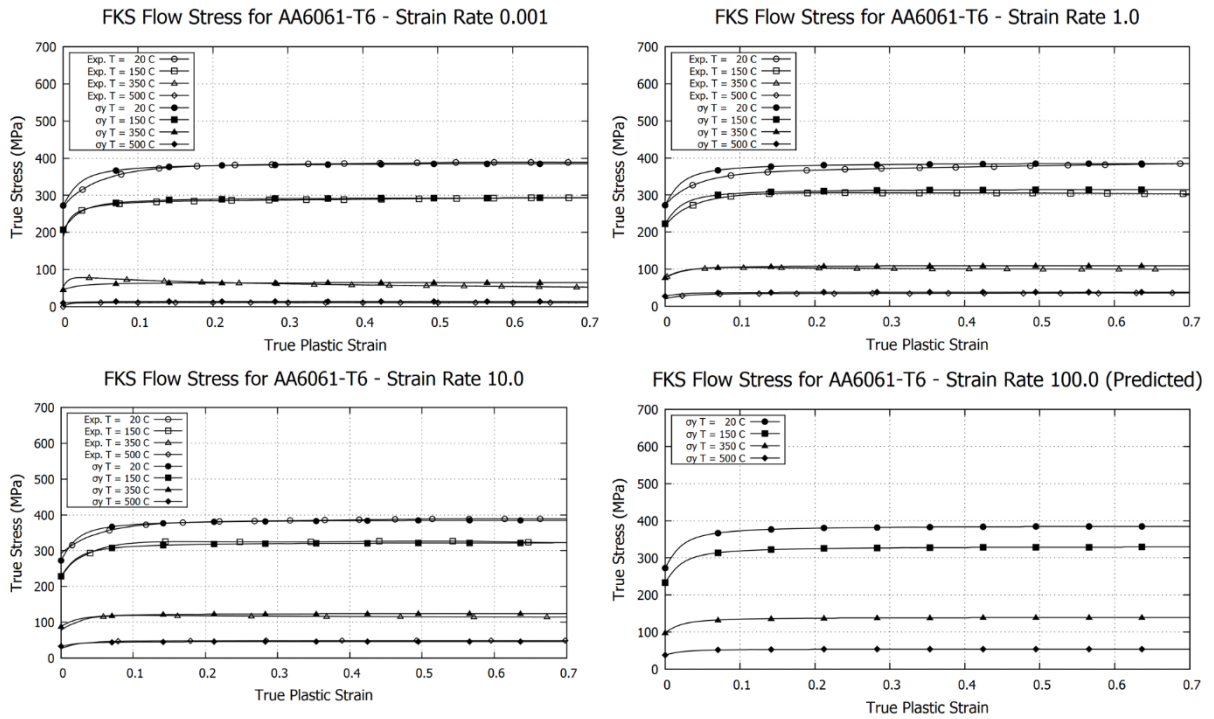
Figure 3-26 – Compression test results for strain rate 0.001, 1.0, 10.0, and 100.0

3.10.6 ITERATIVE PLASTICITY

The radial return algorithm is, in most cases, a sufficient approach to find the unknown increment in plastic strain while assuming J2 Plasticity with local linear strain hardening. However, there are some situations where such an algorithm can provide reduced precision:

- Tangent modulus is large at the current time step

- The time step is large enough to result in significant overshoot of the correct plastic strain

- Finite strains

- Viscoplastic material model (rate dependent)

In these cases, an iterative approach is beneficial whereby the Newton-Raphson (or another appropriate non-linear solution procedure) is used. In this work, an iterative procedure is not deemed necessary, and is not expected to provide significant improvement of the solution precision. We have found that the initial non-linear hardening portion of the flow stress curve is well approximated using the approach explained in Section 3.3.4.

3.11 SPH-FEM HYBRID THERMOMECHANICAL CONTACT

Node to surface contact algorithms are a very popular approach to couple SPH elements with finite elements. The SPH elements are treated as the slave and the finite elements as the master. For FSW simulation, the tool will be treated as a rigid body. The surface of the tool (and support structure if needed) will be meshed with rigid zero thickness triangular plate elements. The penalty formulation will be used to impose the contact constraint to prevent the two bodies from penetrating one within the other. A contact example is shown in Figure 3-27. Here, a flexible body with an initial velocity is meshed with SPH elements (red part). The rigid impact surface (grey part) is meshed with finite elements.
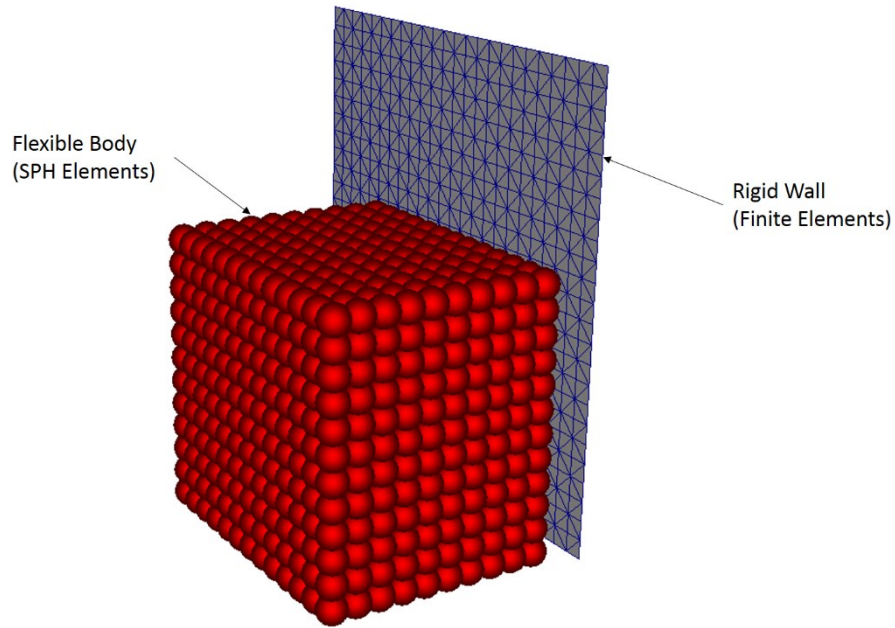
Figure 3-27 – Contact example

The general layout for the SPH and finite elements is shown in Figure 3-28. The i^h SPH element has a center at $\bar{x}_i$ and has smoothing length (radius of influence) of h_i. The jth triangular element has vertices A, B and C; their positions are then $\bar{x}_{A_j}$, $\bar{x}_{B_j}$ and $\bar{x}_{C_j}$ respectively. The triangular element has a surface normal vector that is $\hat{n}_{FEM_j}$. The normal vector is found from:

$$\hat{n}_{FEM_j} = \frac{\bar{U}_j x \bar{V}_j}{\|\bar{U}_j x \bar{V}_j\|}$$

(3-134)
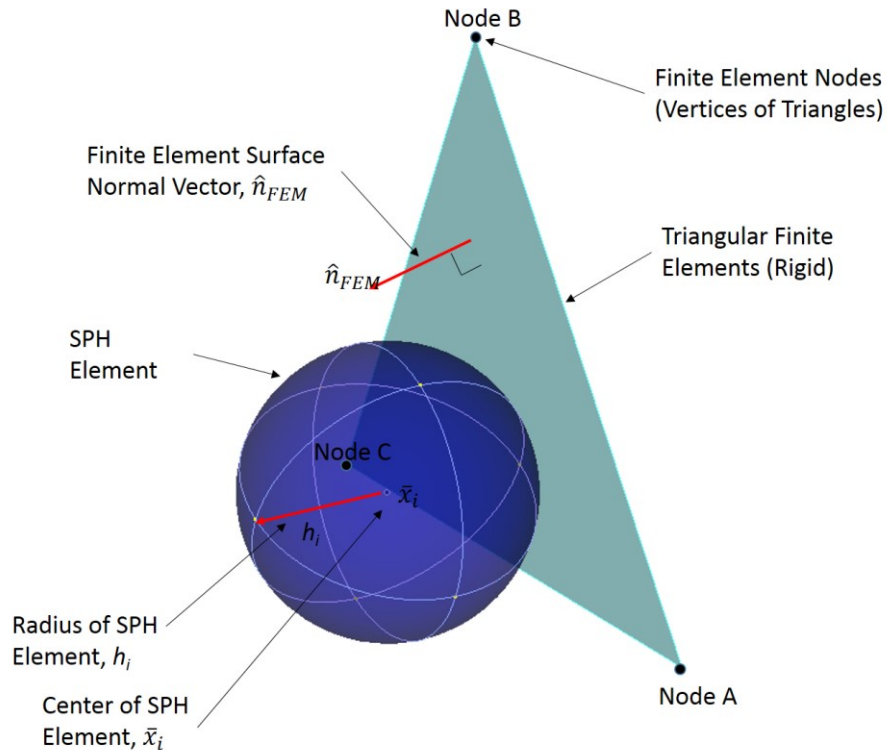
Figure 3-28 – Node to surface contact

Where $\bar{U}_j = \bar{x}_{B_j} - \bar{x}_{A_j}$ and $\bar{V}_j = \bar{x}_{C_j} - \bar{x}_{A_j}$. The $\| \quad \|$ signifies the magnitude of the vector and the $\widehat{}$ represents a normalized unit vector.

3.11.1 CONTACT PAIR BUCKET SORT

Knowing the proximity of an SPH element to a finite element is vital to the contact algorithm. In a brute force approach, contact detection can be evaluated by comparing the location of each SPH element to each finite element; this has order $\mathcal{O}(N^2)$, where N is the number of SPH and finite elements. This approach is prohibitively expensive and would lead to excessively long calculation times. A simple way to improve the search time is to place the SPH and finite elements into buckets (also known as bins).

The bucket sort is a very popular and efficient searching method that is of order $\mathcal{O}(N\ nNeib)$, where $nNeib$ (~56 in 3D) is the SPH average number of neighbors. The general approach is rather simple and has been tackled by many groups. Since the SPH elements require a neighbor search that is based on

the bucket sort, the algorithm can be modified slightly to include the finite elements. This will form a list of SPH elements that are within close range to the finite elements.
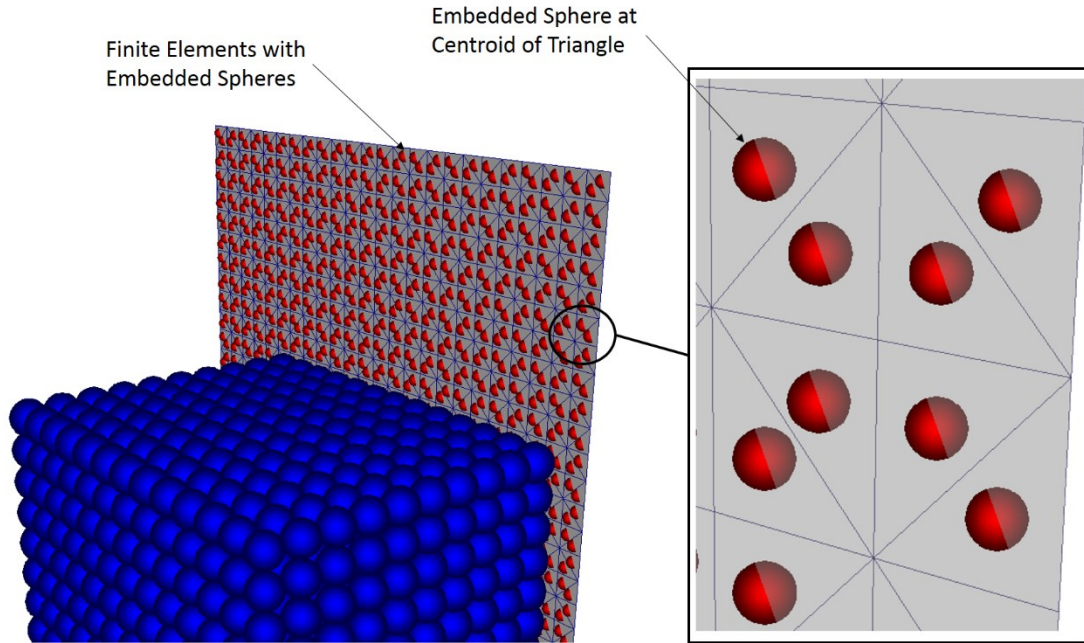
Figure 3-29 – Embedded spheres in triangular finite element mesh

To start, spheres are embedded into the triangular finite elements (because the neighbor search is most efficient when using spheres). The sphere is inserted at the centroid of the triangle, an example is shown in Figure 3-29 (note that the spheres are not shown with their true radius to improve visibility). The location, $\bar{x}_{sphere_j}$, is found by:

$$\bar{x}_{sphere_j} = \frac{1}{3}\left[\bar{x}_{Aj} + \bar{x}_{Bj} + \bar{x}_{Cj}\right]$$

(3-135)

Next, the size of the embedded sphere must be found. In order for the contact detection to work correctly, the sphere must completely enclose the finite element within it is embedded. This is done by ensuring that the radius of the sphere, r_{sphere_j}, is large enough to include the node that is the furthest away from

$\bar{x}_{sphere_j}$. This is accomplished by evaluating the magnitude of three vectors, $\bar{r}_{1j} = \left|\bar{x}_{sphere_j} - \bar{x}_{Aj}\right|$, $\bar{r}_{2j} = \left|\bar{x}_{sphere_j} - \bar{x}_{Bj}\right|$, and $\bar{r}_{3j} = \left|\bar{x}_{sphere_j} - \bar{x}_{Bj}\right|$, where | | signifies the absolute value. The radius of the embedded sphere is then:

Kirk Fraser - April 2017

$$r_{sphere_j} = max\left[\left[\left\|\bar{r}_{1_j}\right\|, \left\|\bar{r}_{2_j}\right\|, \left\|\bar{r}_{3_j}\right\|\right]\right] \qquad\qquad (3\text{-}136)$$
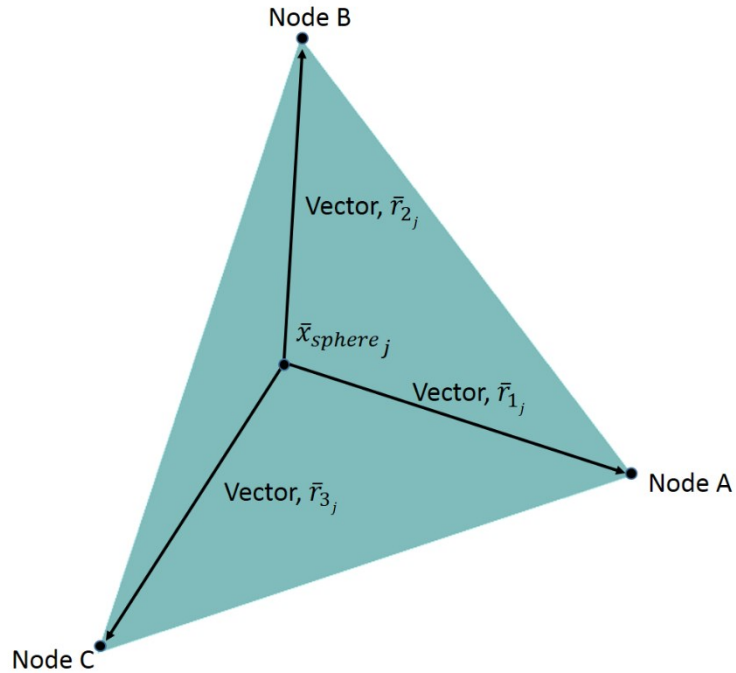
Figure 3-30 – Determining the radius of the embedded sphere

Once the size and location of the embedded spheres is found, the neighbor search (as described in section 4.5) is modified to include the embedded spheres. A list of "Contact Neighbors", $Neib_{contact}$, is found by scanning the original neighbors list, Neib. A cut-off radius is used that is $r_{cut} = 0.5(h_i + h_j)$. This allows for a maximum of ~12 contact neighbors. This makes the contact detection phase very efficient since only 12 possible contact pairs must be checked for contact.

3.11.2   NODE TO SURFACE CONTACT DETECTION

The contact detection approach that we use is based on a ray tracing algorithm found in [221]. Each potential contact pair in $Neib_{contact}$ is checked to test for an intersection point that lies within the triangular finite element. Some changes to the standard ray-triangle intersection algorithm have been made to take into account the movement of both the SPH and finite elements as well as the radius of the SPH elements. Start by calculating the relative velocity, $\bar{v}_{Rel}$, between the SPH, $\bar{v}_i$, and finite element, $\bar{v}_j$:

$$\bar{v}_{Rel} = \bar{v}_i - \bar{v}_j \tag{3-137}$$

Before going any further in the contact test, the penetration rate, $\dot{\delta}$, must be calculated:

$$\dot{\delta} = -(\hat{n}_{FEM_j} \cdot \bar{v}_{Rel}) \tag{3-138}$$
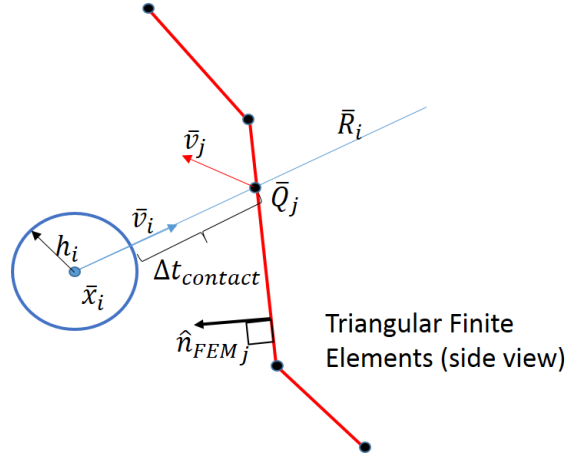
Figure 3-31 – Contact detection

Next, perform a check to ensure that the SPH and finite elements are approaching each other (impending contact). This is done by checking to ensure that $\dot{\delta}$ is greater than zero. If the penetration rate is less than or equal to zero, the pair are moving away and are not considered to be contact candidates.

There exists a ray that extends from the center of the SPH element in the direction of the velocity. The ray, $\bar{R}_i$, is defined by:

$$\bar{R}_i = \bar{x}_i + \Delta t_{contact} \bar{v}_{Rel} \tag{3-139}$$

The parametric equation for a plane that passes through the three points of the triangular finite element is given by:

$$\hat{n}_{FEM_j} \cdot \bar{x} = p_{plane} \tag{3-140}$$

where any $\bar{x}$ must lie on the plane. p_{plane} is the plane coefficient and is given by:

$$p_{plane} = \hat{n}_{FEM_j} \cdot \bar{x}_{A_j} \tag{3-141}$$

Note that any point on the triangle ($\bar{x}_{A_j}$, $\bar{x}_{B_j}$ or $\bar{x}_{C_j}$) could be used, but it is convenient to use node A (see Figure 3-30). The time increment for the contact to occur, $\Delta t_{contact}$, this can be found from:

$$\Delta t_{contact} = \frac{h_i + p_{plane} - \left(\hat{n}_{FEM_j} \cdot \bar{x}_i\right)}{-\dot{\delta}} \tag{3-142}$$

The next step is to check whether contact will occur within the current time step. If $\Delta t_{contact} \leq \Delta t$ then the SPH element will at least come into contact with the plane that passes through the vertices of the finite element. The maximum time step size is found from:

$$\Delta t = min(\Delta t_{CFL}, \Delta t_{ext}) \tag{3-143}$$

The mechanical time step is found based on the Courant-Friedrichs-Lewy (CFL) criteria:

$$\Delta t_{CFL} = CFL\left(\frac{h}{c + \|\bar{v}\|}\right) \tag{3-144}$$

This condition ensures that an elastic stress wave propagating at the isothermal bulk speed of sound will be captured as it moves through an element of support size h. The time step size due to external forces (from contact) is:

$$\Delta t_{ext} = \epsilon_{ext}\left(m\frac{2\|\bar{v}\|}{\|\bar{F}_N\|}\right) \tag{3-145}$$

This condition ensures that the time step will be small enough to prevent an element from penetrating excessively within a body with which it is in contact.

Next, the point of contact, $\bar{Q}_j$, on the plane is found from:

$$\bar{Q}_j = \bar{x}_i + (\bar{v}_i \Delta t_{contact}) - (h_i \cdot \hat{n}_{FEM_j}) \tag{3-146}$$

This approach ensures that the closest point on the SPH element is taken to be the contact point (similar to a closest point projection algorithm). Now that the contact point has been found, the next step is to determine if this point falls within the bounds of the triangular finite element (see Figure 3-32 for details).
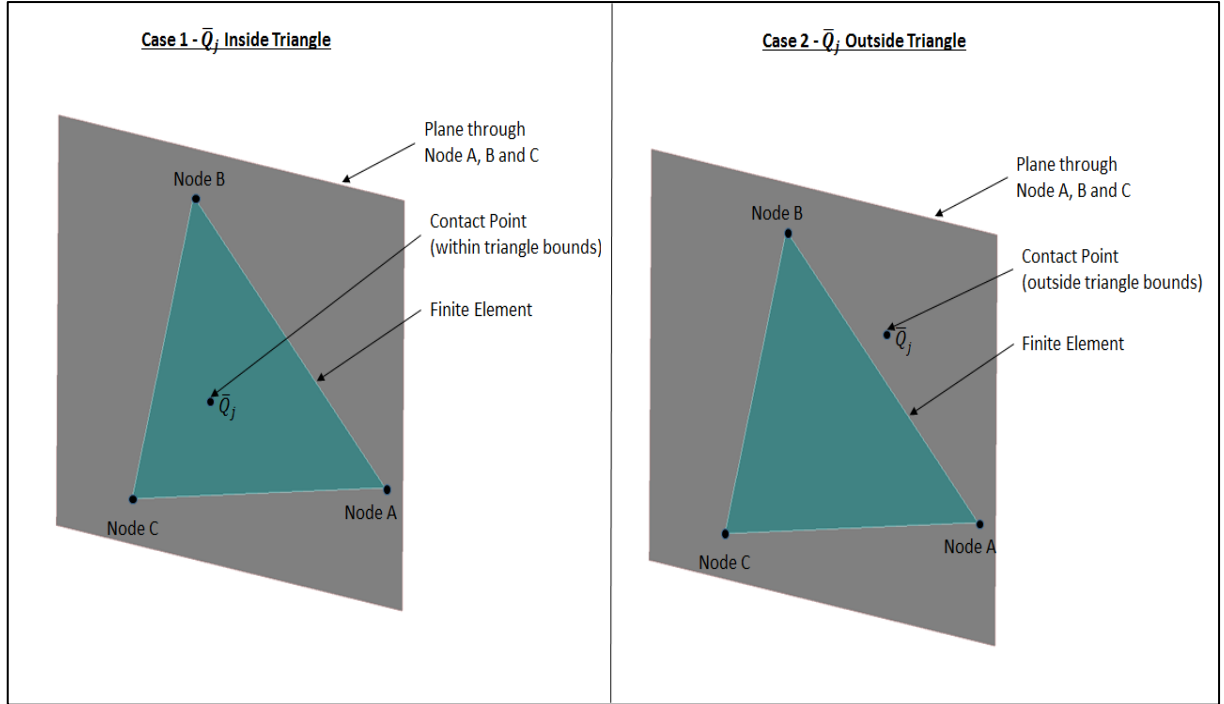
Figure 3-32 – Contact point inside or outside of finite element

For contact to be actually present between the SPH and finite element, $\bar{Q}_j$ must lie within the bounds of the triangle. This can be tested using the "inside-out" test. The general idea is to find a vector that is normal to a vector from one of the nodes to the contact point and to a side of the triangle. Then we test to see if this normal vector is in the same direction as $\hat{n}_{FEM_j}$. This is repeated for all the sides of the triangle. If all the calculated normals are in the same direction as $\hat{n}_{FEM_j}$ then the point is considered to be within the triangle. If any ONE of the tests fails, the point is not within the triangle. The test for inclusion can be stated algebraically as:

$$\left[\left(\bar{x}_{B_j} - \bar{x}_{A_j}\right) \times \left(\bar{Q}_j - \bar{x}_{A_j}\right)\right] \cdot \hat{n}_{FEM_j} \geq 0 \; and \tag{3-147}$$

$$\left[\left(\bar{x}_{C_j} - \bar{x}_{B_j}\right) \times \left(\bar{Q}_j - \bar{x}_{B_j}\right)\right] \cdot \hat{n}_{FEM_j} \geq 0 \; and$$

$$\left[\left(\bar{x}_{A_j} - \bar{x}_{C_j}\right) \times \left(\bar{Q}_j - \bar{x}_{C_j}\right)\right] \cdot \hat{n}_{FEM_j} \geq 0$$

The greater or equals account for the case that $\bar{Q}_j$ is on an edge of the triangle. Once a valid contact pair has been found, the relative velocity must be recalculated using the velocity of the contact point, $\dot{\bar{Q}}_j$, on the finite element mesh:

$$\bar{v}_{Rel} = \bar{v}_i - \dot{\bar{Q}}_j \tag{3-148}$$

This is an important step if the FE mesh is much coarser than the SPH element spacing. Certainly, if the FE mesh and SPH density is similar, this step will not lead to a drastic improvement. Nevertheless, this step should always be performed to ensure a robust and precise contact solution. The next step is to proceed to calculate the penetration depth, δ:

$$\delta = (dt - \Delta t_{contact})\dot{\delta} \tag{3-149}$$

The penetration depth is used in the contact formulation, which will be described in the next section.

3.11.3   MECHANICAL CONTACT FORMULATION

Once a valid contact point has been found between a SPH and a finite element, the next step is to determine a force that will apply to the SPH element to remove the penetration. Although the bodies do not actually penetrate one within the other, the concept has a close relationship to the actual compression in the bodies. This can be shown by considering two elastic bodies, body i and body j that have come into contact (see Figure 3-33).
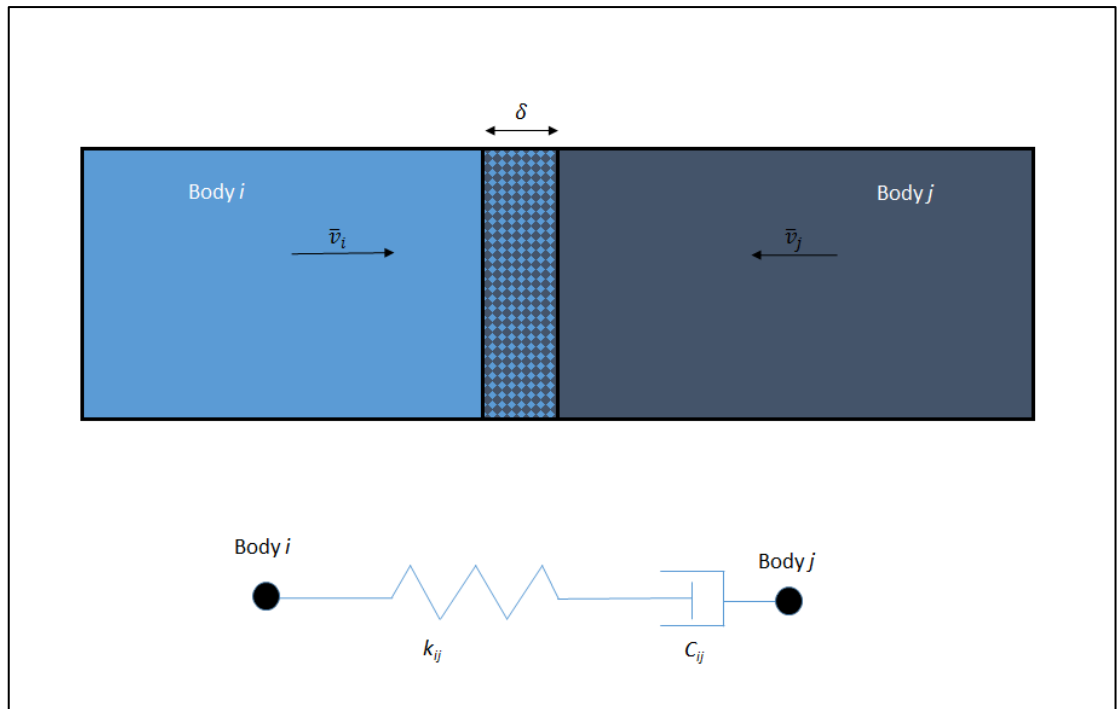
Figure 3-33 – Spring and damper contact model

A representative spring and damper model is shown below the penetrating bodies in Figure 3-33. The reaction force is taken to be proportional to the penetration depth and the stiffness at the contact interface, k_{ij} which will be related to the stiffness of the two contacting bodies. Contact damping is included to damp out oscillations in the contact behaviour. The damping is proportional to the relative velocity of the contacting bodies.

The amount that the bodies penetrate must be the same as the sum of the compression in each body. Furthermore, when an elastic body is compressed, it will want to return to its original shape. The force felt as the body tries to regain its original shape will be equal and opposite to the force required to compress the body.

The relationship between force, displacement, and stiffness according to Hooke's Law (in 1D) is:

$$F = k\delta \tag{3-150}$$

Under axial compression, the deflection, δ, of a prismatic bar is:

$$\delta = \frac{F\,L}{A\,E} \tag{3-151}$$

where F is the force acting on the bar, L is the length of the bar, A is the cross section area, and E is the modulus of elasticity. Combining the two equations and re-arranging for k:

$$k = \frac{E\,A}{L} \tag{3-152}$$

In the case of an SPH element, imagine that the element represents a cube like volume of material with sides equal to the mesh point spacing (inter particle spacing), Δs. The volume of the element would be Δs^3 and the stiffness of an individual i^{th} SPH element would then be:

$$k_i = \frac{E_i\,\Delta s^2}{\Delta s} = E_i\,\Delta s \tag{3-153}$$

As previously mentioned, the contact interface stiffness, is composed of the stiffness' of both bodies i and j. Continuing with the spring analogy, we can write the contact stiffness, k_{ij}, according to the rule for two springs in series:

$$k_{ij} = \frac{k_i k_j}{k_i + k_j} = \frac{E_i\,E_j\,\Delta s^2}{\left(E_i + E_j\right)\Delta s} = \left(\frac{E_i\,E_j}{E_i + E_j}\right)\Delta s \tag{3-154}$$

To provide some fine-tuning, a penalty factor is added, $PFAC$, so that the resulting contact interface stiffness in the SPH simulation code is:

$$k_{ij} = PFAC\left(\frac{E_i\,E_j\Delta s}{E_i + E_j}\right) \tag{3-155}$$

The penalty factor is typically set to 1.0 (in this work $PFAC = 1.0$ is used exclusively), but can be set to lower values in the case of high velocity impact. Now, contact damping will be included to prevent unwanted numerical noise. Consider the contact interface to be a simple spring-mass system. The natural frequency, $\omega_{contact}$, of such a contact system is then:

$$\omega_{contact} = \sqrt{\frac{k_{ij}}{M_{contact}}} = \sqrt{k_{ij}\left(\frac{m_i + m_j}{m_i m_j}\right)} \tag{3-156}$$

where $M_{contact}$ is the equivalent mass at the contact interface and m_i is the mass of the ith contacting element in the pair. Then, the critical damping, $\zeta_{critical}$, is:

$$\zeta_{critical} = 2m_{avg}\omega_{contact} = (m_i + m_j)\sqrt{k_{ij}\left(\frac{m_i + m_j}{m_i m_j}\right)}$$

(3-157)

In the SPH code, the contact damping, $\zeta_{contact}$ is applied as a fraction of the critical damping:

$$\zeta_{contact} = DFAC\, \zeta_{critical}, \qquad 0 \leq DFAC \leq 1$$

(3-158)

In this work, all the simulation are performed with $DFAC = 0.2$. Finally, the force vector that is added to the ith SPH element in the direction of the surface normal to prevent penetration of the contacting pair is:

$$\bar{F}_{N_i} = \left(k_{ij}\delta - \zeta_{contact}\dot{\delta}\right)\hat{n}_{FEM_j}$$

(3-159)

3.11.4   THERMAL CONTACT APPROACH

In FSW, quantifying the propagation of heat throughout the work pieces is of key importance. Of equal importance is to know how heat flows from the work pieces to the welding tool (and vice versa). The tool interacts mechanically with the work piece through the previously described contact algorithm. The aforementioned mechanical contact algorithm does not allow the transfer of heat across the contact interface.

In this section, the implementation of an efficient thermo-mechanical contact algorithm for smoothed particle hydrodynamics simulations will be described. The idea is to treat the mechanical contact as previously described with the node to surface penalty approach and treat the thermal contact by allowing heat to flow from one part to the other using the SPH heat equation. In order to do this, different neighbor's lists must be set up for the mechanical part and the thermal part. As a point of efficiency, a separate list is also used for the potential contact pairs (for mechanical contact). If the thermal part and the mechanical part shared the same neighbors list, a no-slip type of situation would occur between the tool and the work pieces. This would negate the development of an advanced friction model (see Section 3.12). Figure 3-34 shows a simulation model of a rigid block (red) that slides across a flexible aluminum block (green). The surface of the base of the rigid block is meshed with finite elements.
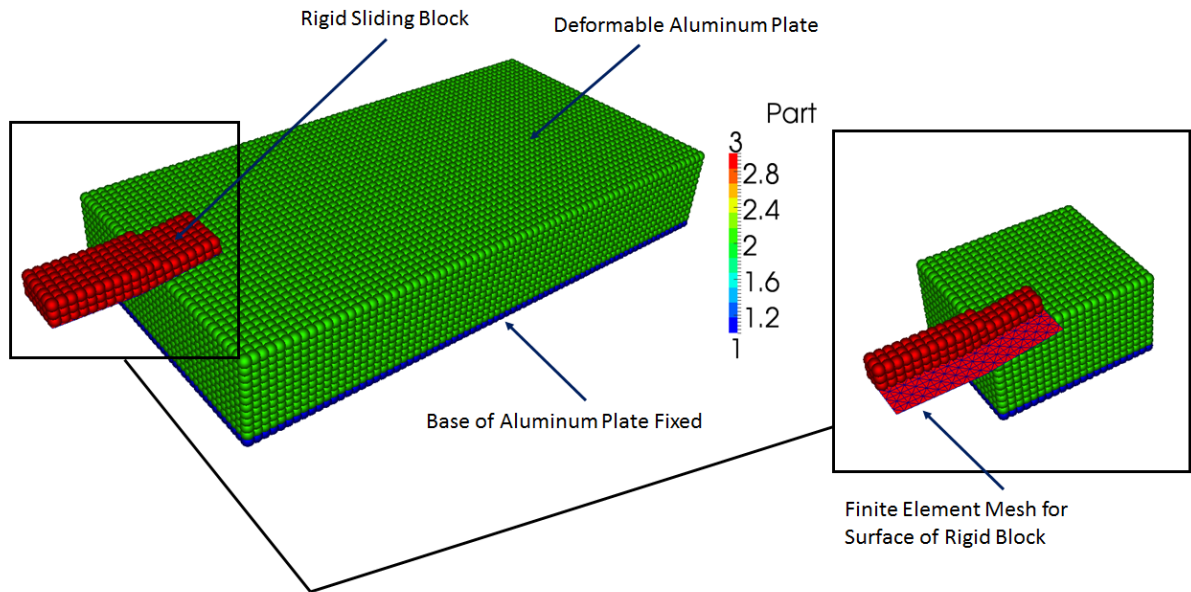
Figure 3-34 – Hybrid Thermo-Mechanical Contact Example

To simulate the discontinuity due to the contact of two bodies we need to make sure that the SPH elements on the different bodies are not included in the same neighbor list. The following lists are set up in the code:

1. *Neib*   - This is the full list containing all the element in the model

2. *NeibMech*   - This is a list that contains the neighbors only within the deformable parts

3. *NeibTherm*   - This is a list that contains the neighbors for the thermal problem

4. *NeibContact*   - This is a list that contains the neighbors for the potential contact pairs

The different lists can be visualized in ParaView for the test case shown in Figure 3-34. The full neighbor list is shown in Figure 3-35. This list contains the SPH elements in the whole model along with the embedded spheres in the finite element mesh.
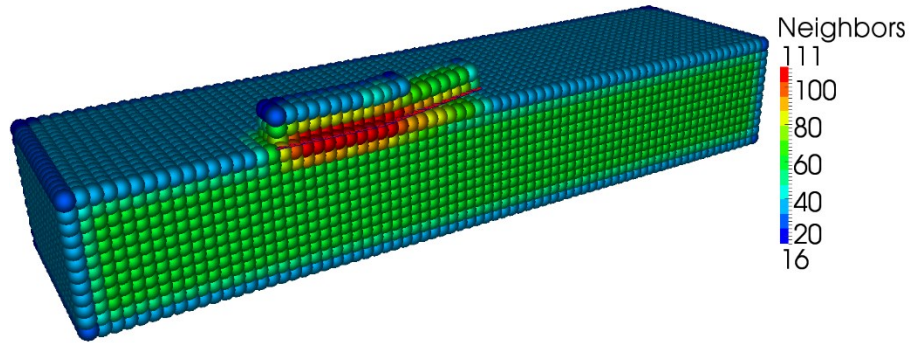
Figure 3-35 – Full neighbors list (Neib)

The $NeibMech$, $NeibTherm$ and $NeibContact$ lists are formed by scanning the full neighbors list. Although building the full neighbor list ($Neib$) is $\mathcal{O}(NlogN)$, building the other lists is $\mathcal{O}(N)$ and is easily computed in parallel on the GPU.

The neighbors for the deformable body are shown in Figure 3-36. Notice that only the SPH elements for Part 1 and 2 (see Figure 3-34 for Parts) have neighbors. On the other hand, the $NeibTherm$ list contains the elements in Parts 1, 2 and 3 as shown in Figure 3-37. The potential contact pairs are stored in $NeibContact$ as shown in Figure 3-38. Here we can see that only the SPH element on the surface of Part 2 in direct contact with Part 3 have neighbors. This allows the contact algorithm to be very efficient. Notice that the maximum number of contact neighbors is 14. This is important so that the neighbor search for the contact pairs (and the full neighbor list, $Neib$, as well) does not have to be updated every single time step. This significantly improves the computational time of the simulation.
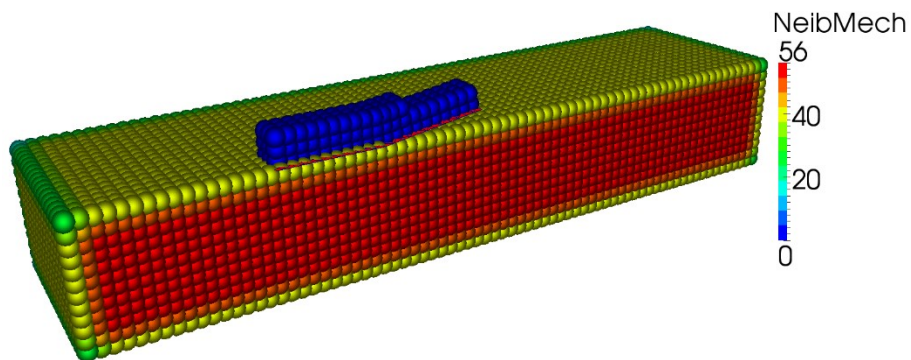
Figure 3-36 – Deformable body neighbors ($NeibMech$)
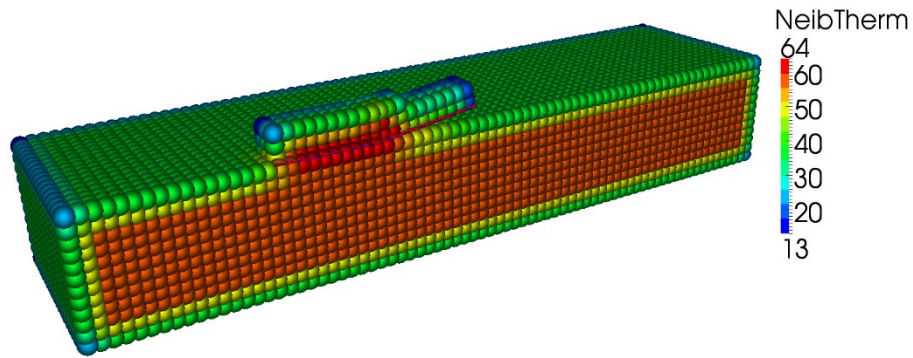
Figure 3-37 – Thermal problem neighbors ($NeibTherm$**)**
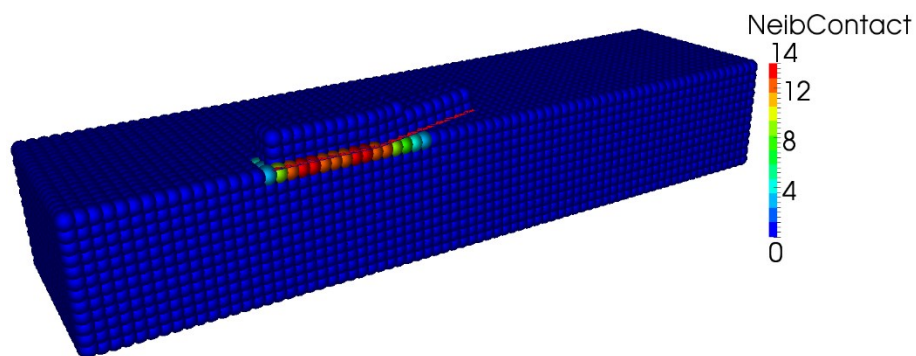
Figure 3-38 – Potential Contact Pairs Neighbor ($NeibContact$**)**

3.12 CUMULATIVE DAMAGE FRICTION MODEL

The friction behavior between the tool and the work pieces is significant in FSW. The correct treatment of this aspect is numerically difficult to calculate due to a lack of understanding of the underlying physics involved at the contact interface. The contact behavior is certainly a complex phenomenon in the FSW process; the interface between the tool and the work pieces is not just a case of two flat parallel surfaces. As the tool rotates and advances, change occurs at the interface as can be seen in Figure 3-39. Of particular importance are the presences of surface agglomerates (particles) and asperities (unseen) on the surface of the weld. These particles are likely due to the breakdown of contact asperities, followed by agglomeration of individual surface particles and hardening.
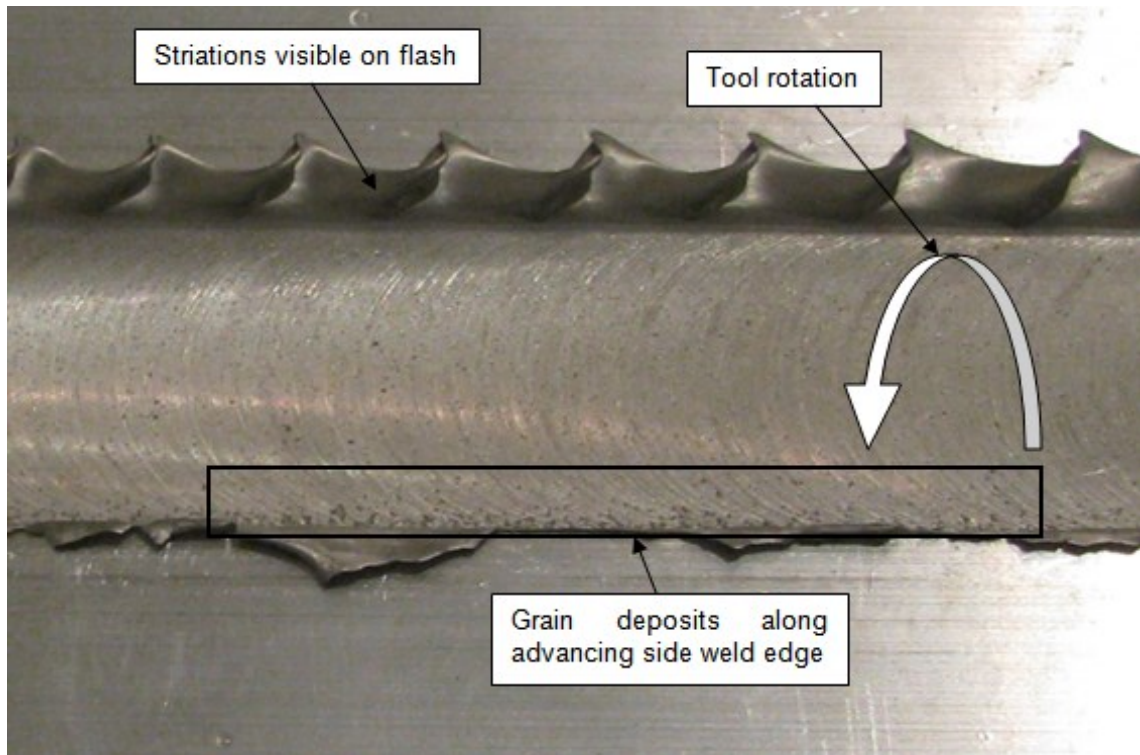
Figure 3-39 – Weld surface of a FSW joint

There is certain plausibility behind the various friction models proposed by other authors for FSW. However, there is a definitive need for improved friction models. To this end, a new fiction behaviour model is proposed that will tie a cumulative damage approach to the friction model. The goal is to represent phenomenologically the stick-slip friction behavior, asperity evolution, and surface particle agglomeration in the FSW process.

3.12.1   THE STICK-SLIP FRICTION MODEL

The standard stick-slip friction model that has been used in numerical simulation of FSW by other authors [222-226] is:

$$\tau_\varphi = \delta\tau_y + (1 - \delta)\mu p_c \tag{3-160}$$

The friction law relates the contact shear stress, τ_φ, to a linear combination of the shear strength, τ_y, of the material (sticking) and a Coulomb type approach (slipping, involving the coefficient of friction, μ, and the contact pressure, p_c). The amount of sticking or slipping is controlled by the slip ratio δ. In the limit

of $\delta = 0$, the friction model is the same as the Coulomb model so that $\tau_\varphi = \mu p_c$, obtaining a true slip model. On the other hand, when $\delta = 1$, the model becomes purely a stick model and thus $\tau_\varphi = \tau_y$.

3.12.2   THE CONTACT INTERFACE

The transfer of motion (or force) at the contact interface is strongly dependent on the nature of the surface topology at the interface. Initially, the interface will be populated with sharp asperities (peaks and valleys); then as the tool rotates on the work piece material, the asperities will be ground down. Eventually the surface will have fewer asperities and will be more uniformly flat. We suggest that initially, the sharp asperities lead to a sticking behavior (interlocking of the peaks and valleys). Then, as the asperities are damaged, more and more slipping will be present. Figure 3-40 shows a representation of the asperity evolution in the contact surface.
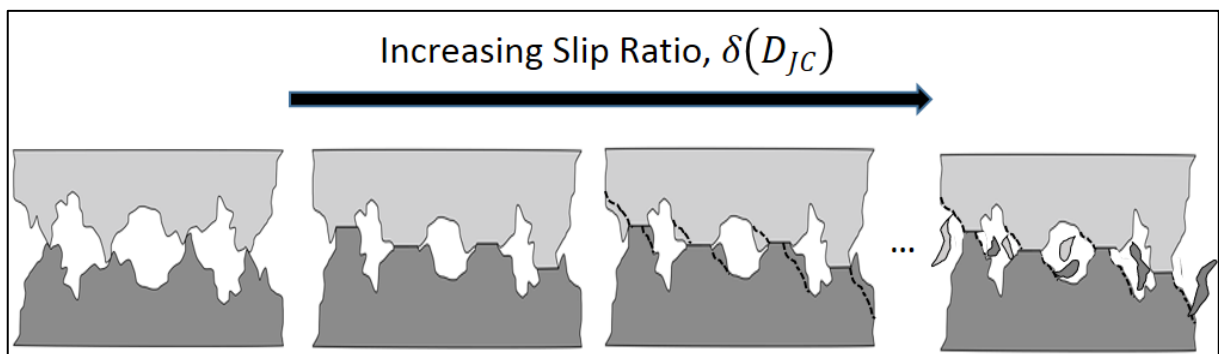
Figure 3-40 – Asperity evolution with relation to slip ratio

When the slip ratio attains a value of one, the element is considered to have sheared away from the surface and is deposited elsewhere in the contact interface. We offer that when it is deposited, it is done so in a chaotic manner (a singularity in this case). Furthermore, we propose that this newly deposited element will be rough due to surface particle agglomeration and strain hardening, providing an approximate shape similar to an asperity. The damage evolution then continues and the slip ratio evolves accordingly.

3.12.3 IDEALIZATION OF THE CONTACT INTERFACE WITH A CUMULATIVE DAMAGE MODEL

A number of models exist that can take into account damage accumulation. A straightforward model is that of Johnson and Cook [145] (other details can be found in [227]). The damage parameter, D_{CDF}, is a history variable and is defined as:

$$D_{CDF} = \sum \frac{\Delta \hat{\varepsilon}^p}{\varepsilon^f} \tag{3-161}$$

where $\Delta \hat{\varepsilon}^p$ is the current time step effective plastic strain increment and ε^f is the current step failure strain:

$$\varepsilon^f = max\left[\left(D_1 + D_2 e^{(D_3 \sigma^*)} \right)(1 + D_5 T^*), \varepsilon^f{}_{min} \right] \tag{3-162}$$

The damage parameters, D_1 to D_5 are taken from Schwer [142] and have the values shown in

Table 3-3 – Cumulative damage Constants used for proposed friction model

Constant	Value	Constant	Value
D_1	-0.77	D_4	0.0
D_2	1.45	D_5	1.6
D_3	-0.47	$\varepsilon^f{}_{min}$	0.2

The minimum failure strain, $\varepsilon^f{}_{min}$, must be included due to "divide by zero" problems under certain stress states. σ^* is the Lode parameter defined as: $\sigma^* = p/\sigma_{eff}$ (ratio of material point pressure to effective stress). More information on the determination of the damage constants can be found in [228]. The damage parameter is then related to the slip ratio:

$$\delta(D_{CDF}) = \begin{cases} D_{CDF}, & D_{CDF} < 1 \\ D_{CDF} \leftarrow 0, & D_{CDF} \geq 1 \end{cases} \tag{3-163}$$

This behavior is best explained by a diagram: Figure 3-41 shows the evolution of the slip ratio for an SPH element at the contact interface as a function of time. The slip ratio increases until the damage attains a value of one. At that point, the element is considered to have released and been deposited

elsewhere forming a new asperity. At that time, both the damage parameter and the slip ratio are reset to zero.
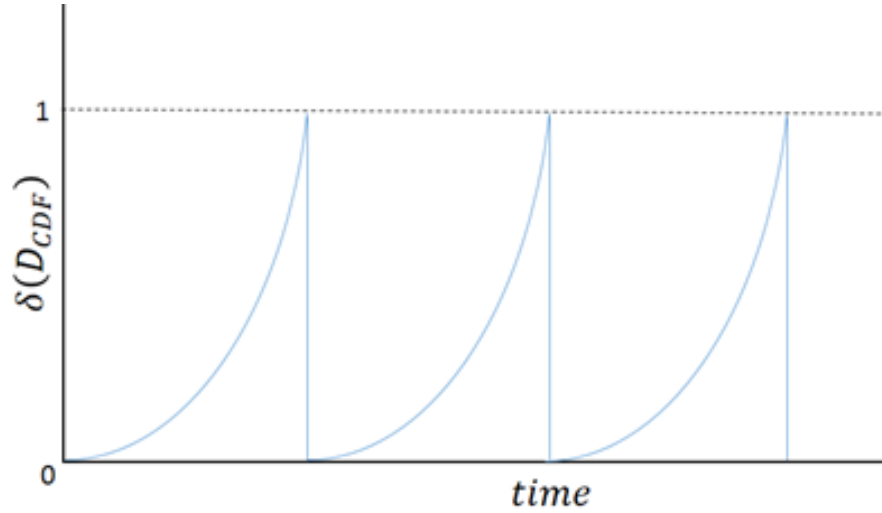
Figure 3-41 – Evolution of the slip ratio as a function of surface damage

3.12.4 CUMULATIVE DAMAGE FRICTION FORCE EQUATION

The proposed approach associates the slip ratio to the cumulative damage friction parameter, D_{CDF}. The resulting cumulative damage stick-slip friction model takes the form:

$$\tau_\varphi = \delta_{CDF}(D_{CDF})\tau_y + (1 - \delta_{CDF}(D_{CDF}))\mu p_c \tag{3-164}$$

The underlying idea is that asperities at the surface in the contact interface will be damaged in a complex way that can only be accounted for by considering an accumulation of damage. In the context of an explicit dynamics code, the contact shear stress is more aptly converted into a tangential contact force (as described in section 3.11):

$$\bar{F}_T = \delta_{CDF}(D_{CDF})\frac{\sigma_y(\hat{\varepsilon}^p, \dot{\varepsilon}, T^*)}{\sqrt{3}}A_c + (1 - \delta_{CDF}(D_{CDF}))\mu\|\bar{F}_N\|\hat{n}_{Tj} \tag{3-165}$$

Where $\sigma_y(\hat{\varepsilon}^p, \dot{\varepsilon}, T^*)$ is as described in section 3.10, A_c is the contact surface area of the SPH element, which is assumed to be:

$$A_c = \Delta s^2 \tag{3-166}$$

and $\|\bar{F}_N\|$ is the magnitude of the contact force along the unit normal direction at the contact interface. $\hat{n}_{T_j}$ is a unit vector at the contact interface in the tangential direction. More specifically, this is the friction direction:

$$\hat{n}_{T_j} = \frac{\bar{v}_{rel_T}}{\|\bar{v}_{rel_T}\|} \tag{3-167}$$

The relative tangential velocity at the contact interface is ($\dot{\bar{Q}}_j$ as described in (3-146) and (3-148)):

$$\bar{v}_{rel_T} = \dot{\bar{Q}}_j - \left(\dot{\bar{Q}}_j \cdot \hat{n}_{FEM_j} \right) \hat{n}_{FEM_j} \tag{3-168}$$

This is little doubt that other forms of the cumulative damage model are possible (and in some cases more representative of the contact interface). The main goal of this development is to show that a relationship can be formed between the changing surface conditions and the contact interface behaviour. The CDF friction model will be used in the test cases in the dissertation. In section 5.1.3 a comparison of different friction modeling approaches (stick, slip, CDF) will be investigated. At that point, the value and merit of the proposed model will become evident.

3.13 PIN THREAD AND SHOULDER SCROLL MODEL

One of the advantages of a meshfree method such as SPH is that reasonable precision results can be obtained with significantly fewer calculation points than would be possible with a grid-based method. On the other hand, typically in order to account for a threaded pin or a scrolled shoulder, a high mesh density (for a meshed based method) is required near the tool to resolve the effects of the threaded tool on the material flow Figure 3-42 [229].
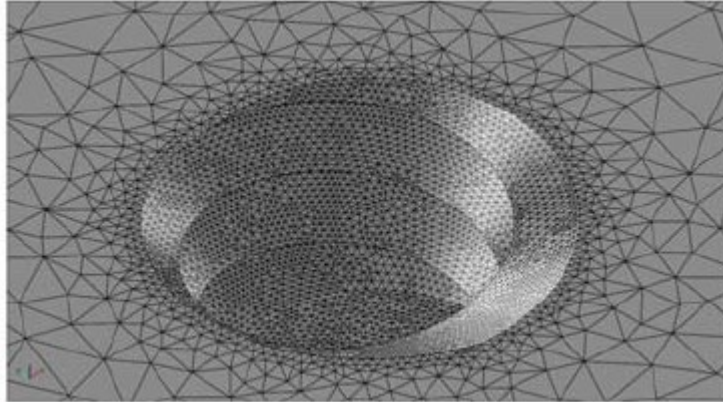
Figure 3-42 – Typical mesh density in a threaded tool CFD simulation model [229]

To circumvent this difficulty, a kinematic constraint approach can be used to approximate the transmission of forces to the workpiece material by the threads. To this end, a smooth (not threaded) finite element mesh can be used for the tool geometry as shown in Figure 3-43.
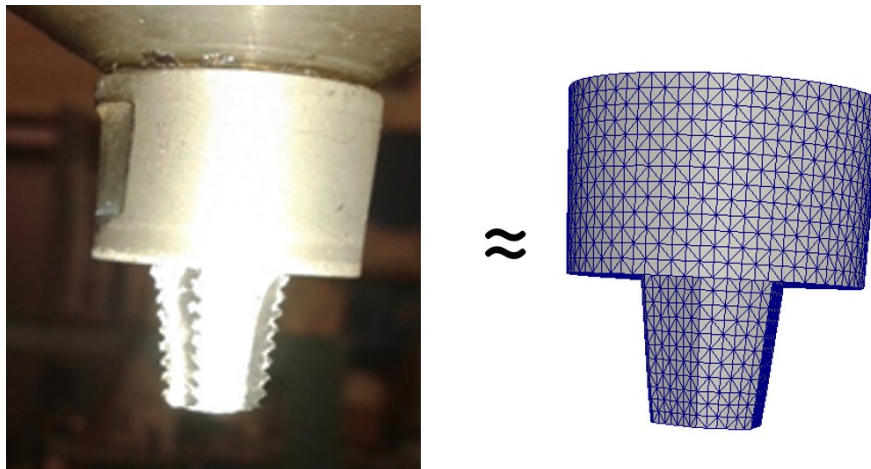
Figure 3-43 – Equivalent tool model without threads

Approximate forces can be added to satisfy the condition that the material in contact with the threads should move at a rate proportional to the rpm of the tool and the thread pitch, p_{thread}. The force required to incur the desired material movement can be derived from the kinematic equation:

$$\Delta \bar{x} = \bar{v}\,\Delta t + {}^{1}\!/_{2}\,\bar{a}\,\Delta t^{2} \tag{3-169}$$

Certainly, a material point that remains in contact with the threaded region should be moved though $\Delta \bar{x}$ for a certain amount of tool rotation per time step (due to the pitch of the threads). For the considered

simulations, the value of $\bar{v}$ can be assumed to be zero since the distance that the particle will have to move is small and the time step size is on the order of 10^{-7} s. From the definition of thread pitch, one can infer that the material will be moved through a full pitch distance per tool rotation. In this sense, the material movement must be:

$$\Delta \bar{x} = p_{thread} \; rps \; \Delta t \; \left(\hat{n}_{FEM_j} \times \hat{n}_{T_j} \right)$$

(3-170)

where rps is the number of tool revolutions per second, p_{thread} is the thread pitch, and $\hat{n}_{FEM_j} \times \hat{n}_{T_j}$ represents the vector pointing in the direction of the material motion due to the threads (as shown in Figure 3-44).
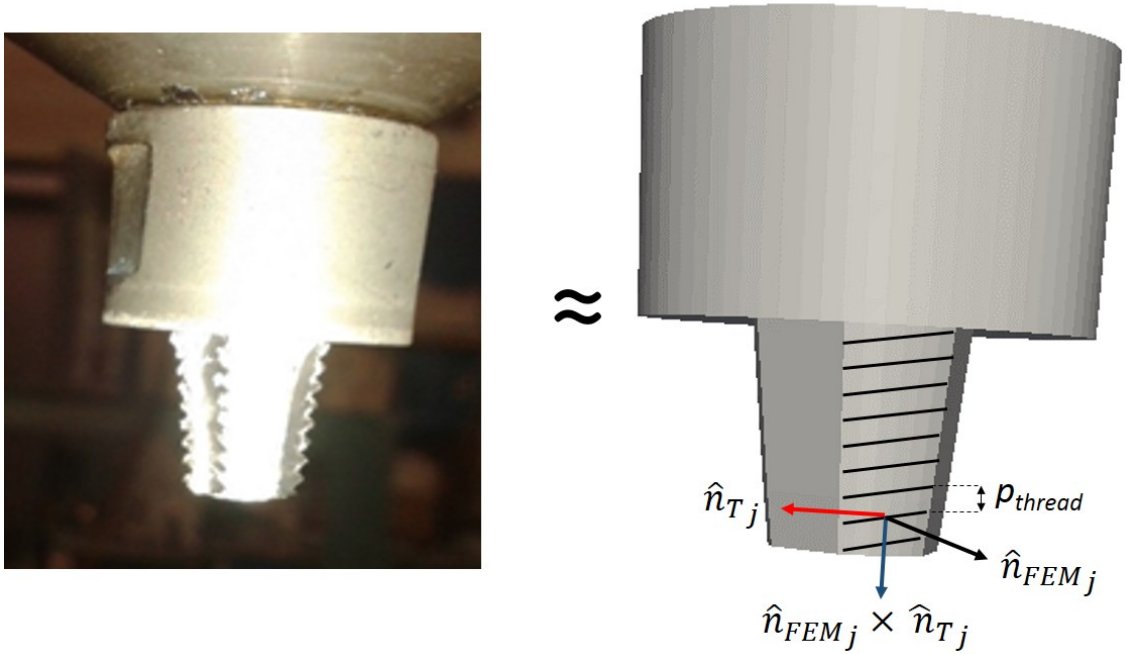
Figure 3-44 – Equivalent thread force model

Combining equations (3-169) and (3-170) and noting that $\bar{F} = m \, \bar{a}$, the force required to move the ith SPH material point in contact with the threaded section of the tool per time step is:

$$\bar{F}_{thread_i} = \frac{2m_i}{\Delta t} \left(p_{thread} \; rps \; \left(\hat{n}_{FEM_j} \times \hat{n}_{T_j} \right) \right)$$

(3-171)

This modelling approach provides a robust and efficient means to include the effect of pin threads and shoulder scrolls. This development allows coarse mesh models to obtain excellent results in far less time than would be possible using a fine mesh with the threads explicitly modeled.

3.14 TOOL WEAR PREDICTION

The life span of a FSW tool is relatively short; a typical tool could last anywhere from 100 to just over 500 meters of weld. Given the current cost of a single tool (a specialized tool can be over 1000$), there is a great interest in being able to predict the wear of the tool. Such a model should allow the analyst to perform parametric studies to evaluate the effect of varying process parameters on the wear life of the tool. Wang *et al.* [230] showed the amount of material worn from the tool for different tool material and geometries. Other authors have worked on quantifying the resulting joint quality and microstructure as a function of tool wear [231-233]. The wear behaviour can be explained by Archand's wear law based on Zmitrowitz [234], the wear depth rate, $\dot{d}_{wear}$, is given by:

$$\dot{d}_{wear} = \frac{k_{wear}}{H_v(T)A_c} \|\bar{F}_N\| \|\bar{v}_T\| \tag{3-172}$$

$$H_v(T) = \frac{1}{3} \sigma_{yo_{Tool}} \left(1 - \frac{T - T_{melt}}{T_{melt} - T_{room}}\right) \tag{3-173}$$

where H_v is Vickers hardness of the tool, k_{wear} is a material specific wear constant, and A_c is the contact area. $\sigma_{yo_{Tool}}$ is the room temperature yield strength of the tool (a typical tool would have $\sigma_{y0_{Tool}} >$ 1500 MPa). T is the current temperature of the tool, T_{melt} is the melt point of the tool ($T_{melt} > 1420 \,°C$). Equation (3-172) shows that the wear rate will increase with increasing contact pressure, $\|\bar{F}_N\|$, and relative tangential velocity, $\|\bar{v}_T\|$, at the contact interface. The wear will increase when a softer tool material is used. Note that the contact pressure (in this work) is an indirect function of the hardness of the aluminum since the stiffness (contact penalty from equations (3-155) and (3-159)) of the material will be affected by the flow stress law used. Figure 3-45 shows an example of predicted tool wear from the model described in section 5.3. This is a convex tool with a tapered cylindrical pin. The results show wear patterns that are intuitive for this type of tool geometry.
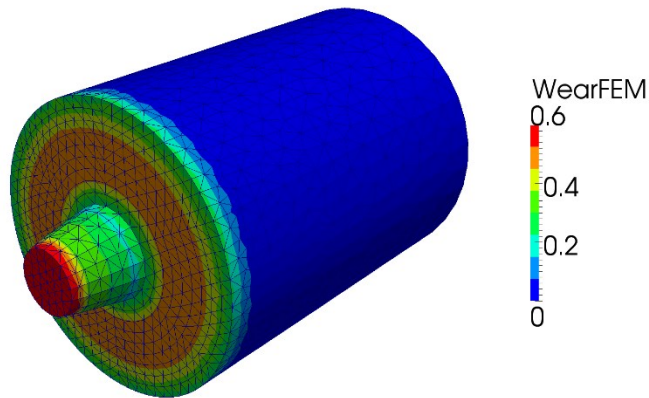
Figure 3-45 – Tool wear prediction example

The developed tool wear model is simple, but at the same time provides a powerful approach to predict the effects of different process parameters on the wear of the tool. In this work, no experimental data is available and as such, the wear model should only be used to evaluate relative wear patterns. To be able to predict the actual life of the tool would require the determination of the value of the wear coefficient (k_{wear} in equation (3-172).

3.15 SOLUTION PROCEDURE

The solution of the set of coupled continuum mechanics equations is undertaken with an explicit (forward difference) time stepping scheme. Due to the large plastic deformation (highly non-linear), an explicit procedure is more efficient and robust in comparison to an implicit solution strategy. In the explicit method, the set of equations are evaluated algebraically; there is no need to assemble and solve a system of equations. This makes the relative per time step calculation cost orders of magnitude faster than is possible with an implicit scheme (assuming an iterative sparse solving approach). Since the per-step cost is higher for the implicit time stepping approach, one should use a much larger time step size. This is, however, unfeasible for the FSW process due to the rate at which the contact between the tool and the work piece is changing. It turns out that the time step requirements for FSW is controlled mainly by the contact condition, making the use of an implicit method incredibly computationally expensive.

Certainly, in less dynamic phases of FSW, such as the cooling phase, an implicit form would be beneficial (formulation is developed in section 9 – Appendices).

The solution procedure seeks to determine a set of unknown variables: $\bar{x}$, $\bar{v}$, ρ, $\bar{\bar{\sigma}}$, T, and d_{wear} at $t = t + \Delta t$ based on known values at $t = t$. The explicit scheme used is coined the "modified Euler" [209, 235, 236]. The procedure is as follows:

1. Initialize the mesh and field variables.

2. Move the boundary particles (or surface mesh) with the imposed velocity or displacement.

3. Perform the neighbor search algorithm.

4. Calculate the average particle velocities with the XSPH method and adjust the material point velocity.

5. Evaluate the right hand side of the density equation, $\frac{d\rho}{dt}$.

6. Calculate the material point pressure from the equation of state.

7. Calculate the right hand side of the deviatoric stress update, $\frac{d\bar{\bar{S}}}{dt}$.

8. Calculate the effects from normal contact, $\bar{F}_N$, friction (tangential), $\bar{F}_T$, and assemble them together:

$$\bar{F}_{ext} = \bar{F}_N + \bar{F}_T \tag{3-174}$$

9. Update the material point densities and deviatoric stresses with:

$$\rho^{t+\Delta t} = \rho^t + \Delta t \frac{d\rho}{dt} \tag{3-175}$$

$$\bar{\bar{S}}^{t+\Delta t} = \bar{\bar{S}}^t + \Delta t \frac{d\bar{\bar{S}}}{dt} \tag{3-176}$$

10. Call the appropriate material law and determine the plastic strain, deviatoric stress and the total stress.

    a. Evaluate if the material point stress surpasses yield, if not, retain the stress state and continue to Step 11, otherwise continue to Step 10.b,

    b.   Use the radial return algorithm to find the plastic strain increment and stress scale factor,

    c.   Scale the deviatoric stresses back to the yield surface ,

    d.   Use the equation of state to calculate the pressure,

    e.   Calculate the total stress.

11. Calculate the right hand side of the momentum equation, $\frac{d\bar{v}}{dt}$.

12. Calculate the acceleration vector from the artificial stress term to remove the tensile instability, $\frac{d\bar{v}}{dt_{AS}}$.

13. Calculate artificial viscosity, $\frac{d\bar{v}}{dt_{AS}}$, if required.

14. Calculate the total acceleration vector:

$$\frac{d\bar{v}}{dt}_{Total} = \frac{d\bar{v}}{dt} + \frac{d\bar{v}}{dt}_{AS} + \frac{d\bar{v}}{dt}_{AV} + \frac{1}{m}\bar{F}_{ext} \tag{3-178}$$

15. Calculate the right hand side of the energy equation and/or the heat diffusion equation, $\frac{de}{dt}$ and/or $\frac{dT}{dt}$.

16. Update the velocity, position and energy and temperature:

$$\bar{v}^{t+\Delta t} = \bar{v}^t + \Delta t \frac{d\bar{v}}{dt}_{Total} \tag{3-179}$$

$$e^{t+\Delta t} = e^t + \Delta t \frac{de}{dt} \tag{3-180}$$

$$T^{t+\Delta t} = T^t + \Delta t \frac{dT}{dt} \tag{3-181}$$

$$\bar{x}^{t+\Delta t} = \bar{x}^t + \Delta t \frac{d\bar{x}}{dt} + \frac{1}{2}\Delta t^2 \frac{d\bar{v}}{dt}_{Total} \tag{3-182}$$

17. Evaluate the wear rate, $\frac{d}{dt}d_{wear}$, and update the surface wear depth:

$$d_{wear}{}^{t+\Delta t} = d_{wear}{}^t + \Delta t \frac{d}{dt}d_{wear} \tag{3-183}$$

18. Return to step 2, continue until termination time is reached.

The maximum time step size is found Equation (3-143) through (3-145), and is repeated here for completeness:

$$\Delta t = min(\Delta t_{CFL}, \Delta t_{ext})$$

The mechanical time step is found based on the Courant-Friedrichs-Lewy (CFL) criteria:

$$\Delta t_{CFL} = CFL \left(\frac{h}{c + \|\bar{v}\|} \right)$$

This condition ensures that an elastic stress wave propagating at the bulk speed of sound will be captured as it moves through an element of support size h. The time step size due to external forces (from contact) is:

$$\Delta t_{ext} = \epsilon_{ext} \left(m \frac{2\|\bar{v}\|}{\|\bar{F}_N\|} \right)$$

This condition ensures that the time step will be small enough to prevent an element from penetrating excessively within a body with which it is in contact. The value taken for the CFL constant is problem specific. Various authors [188, 189, 236, 237] have shown that it should be no greater than 0.3 when the leapfrog or predictor-corrector integration scheme is used. For the modified Euler approach adopted in this work, Randles and Libersky [190] have shown that this integration scheme is stable for values up to 1.5. A value of 0.7 appears to be sufficient for most FSW simulation with extensive non-linear transient behaviour. A value of $\epsilon_{ext} = 1.0$ is generally sufficient to ensure the contact behaviour is correctly captured.

The explicit thermal time step is controlled by:

$$\Delta t_{therm} = \chi \frac{0.3 \rho C_p}{k} h^2 \qquad (3\text{-}184)$$

where χ is a stability factor. When it is set to one, the time step size is at the limit of stability. In practice, $\Delta t_{therm} \gg \Delta t$, which could allow performing the heat transfer calculations less frequently than the mechanical calculations. However, for the FSW process, we have found that the heat transfer

calculations must be calculated at the same frequency as for the mechanical part. This is due to the

interdependence on friction and plastic heat generation on the mechanical solver and vice-versa.

4 PARALLEL PROGRAMMING

Many types of engineering simulations require a large amount of computational time due to the complexity of the numerical model and/or the sheer size of the computational domain. In the case of FSW, capturing all the aspects of the process requires a multi-physics approach that is very computationally burdensome. A typical FSW simulation can take many days or even weeks running on a single processing unit (sequential approach). For this reason, it is critical to be able to find an efficient means to run the simulation code in parallel. The idea is to split the domain into sub-regions and assigns them to individual processing units.

A number of different parallelization strategies can be used to improve the performance of a simulation code. A popular method for small to medium sized models is to use a shared memory parallel (SMP, also referred to as symmetric multiprocessing) approach wherein each processor has its own sets of tasks, but the processors share memory. In this sense, all the simulation data is stored in a common memory location. OpenMP [238] is a very common directives-based programming language that can be used for SMP codes running on central processing units (CPU). Once the models become larger, the memory requirements will surpass the SMP model, in such cases distributed memory parallel (DMP or massively parallel processing, MPP) is a preferred option. Such a tactic is often employed with large numbers of CPUs, whereby the model and the data in memory is split up and assigned to individual compute "nodes". This approach is called distributed-memory parallel and requires the individual compute "nodes" to be linked by a network. A message passing interface (MPI) [239] is required to provide the communication.

Another parallelization strategy that has become very popular is to use the GPU. Today's GPUs have hundreds, and in most cases thousands of "cores". NVIDIA has developed an application programming interface (API) language that makes parallel programming on the GPU straightforward. The language is called "compute unified device architecture" (CUDA). In this chapter, the underlying principles of parallel

programming with CUDA will be detailed. The following sections will explain the programming strategy used in this research project:

Section 4.1 introduces the layout of the GPU and why it is ideal for the developed SPH code, SPHriction-3D.

Section 4.2 will discuss some of the main features of CUDA Fortran and why this language was chosen for the SPH code.

Section 4.3 presents a number of the programming concepts employed to improve the performance of the SPH code.

Section 4.4 shows the performance of the thermal solver in SPHriciton-3D for a simple heat transfer test case.

GPU programming is a vast and rich field of study; the approaches presented in this chapter are meant to be efficient, robust, and simple. We have found that the implementation of highly specialized algorithms meant to improve performance by a few percent is not worth the extra effort. Such undertakings will typically overcomplicate the code. The interested reader is directed towards the work of other authors leading the way in innovative GPU programming [240-255].

4.1 GPU ARCHITECTURE

Figure 4-1 shows a schematic of the architecture of a typical GPU. Each multi-processor is composed of a large number of "thread processors". The GPU has its own memory called global memory that is accessed by all the multi-processors. Data must be explicitly transferred between the CPU (dynamic random access memory, DRAM) and the GPU. The bandwidth for global memory access is ~336 GB/s on a GTX 980 Ti (state of art at time of press). However, transferring data from the CPU to the GPU is much slower (~6 GB/s). For this reason, as much as possible of the code should be programmed on the GPU to limit the amount of data transfer between the CPU and GPU.

The SPH method is an excellent candidate for parallelization with the GPU. The large number of streaming multi-processors on a GPU is perfect for the computationally heavy nature of SPH. SPH codes

written to take advantage of the GPU can typically achieve speed-up factors of 20x to 100x over an equivalent serial CPU (see Dalrymple *et al.* [252] for example). In some cases, speed-up factors of over 150x are possible; although, these are typically problems that are set up to fully exploit the architecture of a specific GPU.

The ultimate goal of SPHriction-3D is to optimize the FSW process parameters through simulation. Currently, a single simulation run can take anywhere from a number of days to many weeks. This is just not feasible for performing virtual prototyping. The optimization process requires a number of iterations. If each iteration requires a number of days of calculation, then the full optimization process will take on the order of months. In order to optimize the process parameters, the GPU can be used to minimize the calculation time as much as possible.
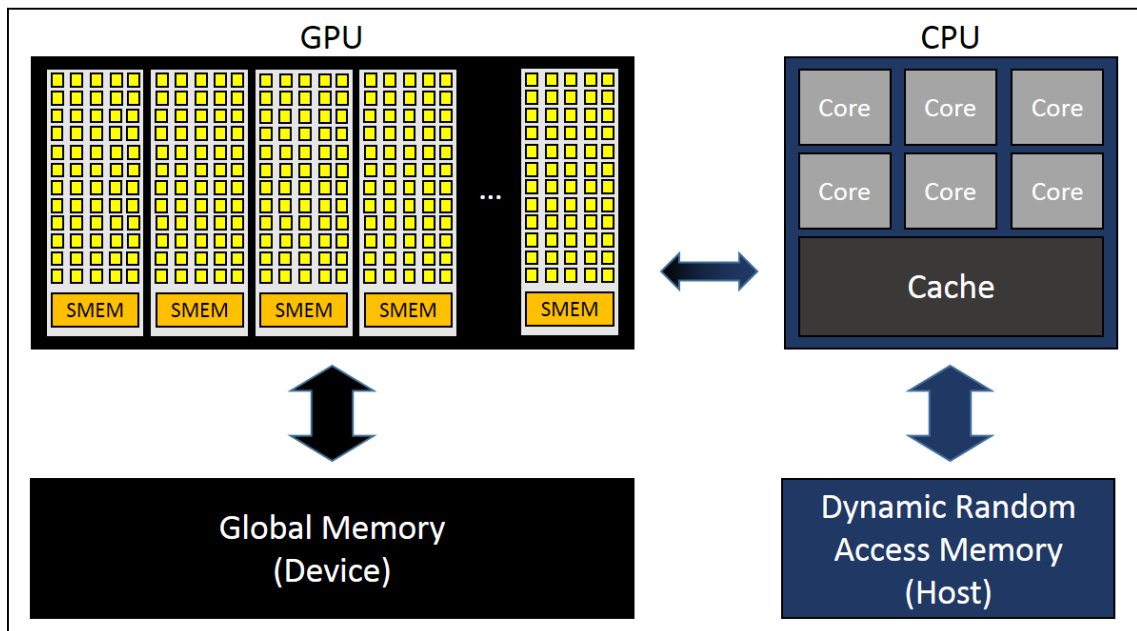
Figure 4-1 – GPU architecture

Presently, there are two main open source codes available for fluid simulation with SPH on the GPU: DualSPHysics [256] and GPUSPH [247] as well as a commercial code, Impetus AFEA [257]. However, at this point, there does not exist an open source or commercial implementation of SPH for solid mechanics on the GPU. Many different fluid flow problems have been solved with the SPH by various authors [242, 246, 251, 253, 255, 258-267].

4.2   CUDA FORTRAN

Fortran is a programming language that is deeply seeded within the scientific numerical simulation community. It has been heavily used for engineering simulations and has a very good numerical methods library. The choice between using C/C++, Fortran, Python, OpenCL, OpenGL… depends on the programmers experience with a certain language and what the intended audience is. We have chosen to use CUDA Fortran for SPHriction-3D because of the simple programming model. CUDA Fortran is Fortran with a set of extensions that allows the programmer to take advantage of the parallel nature of NVIDIA's GPUs. CUDA is inherently a fine-grained data parallelisation. As such, there is very little overhead in creating a parallel instance (in CUDA terminology, this is known as a thread). In this section, some of the basic principles of CUDA Fortran will be discussed. The majority of the discussion will be from Ruetsch and Fatica [250].

A comparison of Fortran and CUDA Fortran is shown in Figure 4-2. Here we have the same program functionality (time integration of a scalar quantity using Euler explicit scheme), the main difference is that the Do loop in the CPU version is replaced with parallel threads. Also, calling the subroutine is slightly different in CUDA Fortran (*Attributes(Global) Subroutine* used as opposed to *Subroutine* in the CPU version). Calling the subroutine is slightly different on the GPU, we must tell the compiler how many threads and thread blocks to launch, this is known as a CUDA kernel call and is accomplished with the triple chevrons syntax, *<<<grid,tBlock>>>*. This simple comparison shows how strikingly similar CUDA Fortran is to Fortran90. In comparison to OpenCL, CUDA Fortran is very straightforward.

Fortran Subroutine

```
Module Module_Integrate

Use Module_Precision
Use Module_Neighbors


Contains

    Subroutine Euler_singleCPU(Start_node,End_node,F,dFdt,dt)

        IMPLICIT NONE

        Integer:: nTotal, Start_node, End_node
        Integer:: i
        Real(fp_kind):: dt
        Real(fp_kind), Intent(IN):: dFdt(nTotal)
        Real(fp_kind), Intent(INOUT):: F(ntotal)

        Do i = Start_node, End_node
            F(i) = F(i) + dt*dFdt(i)
        End Do

    End Subroutine Euler_singleCPU

End Module Module_Integrate
```

CUDA Fortran Subroutine

```
Module Module_Integrate

Use Module_Precision
Use Module_Neighbors
Use cudafor

Contains

    Attributes(Global) Subroutine Euler_single
(Start_node,End_node,F,dFdt,dt)

        IMPLICIT NONE

        Integer, Value:: nTotal, Start_node, End_node
        Integer:: i
        Real(fp_kind), Value:: dt
        Real(fp_kind), Intent(IN):: dFdt(nTotal)
        Real(fp_kind), Intent(INOUT):: F(ntotal)

        i = (blockIdx%x-1)*blockDim%x + threadIdx%x

        If (i >= Start_node .and. i <= End_node) Then
            F(i) = F(i) + dt*dFdt(i)
        End If

    End Subroutine Euler_single

End Module Module_Integrate
```

Do loop replaced by parallel threads

Fortran Main Program

```
PROGRAM SPHriction_HT

Use Module_Integrate

    Implicit None
    ... Vars ...

    Other code

    Call Euler_singleCPU(2501,nTotal,T,dTdt,dt)

    ... Other Code ...

END Program SPHriction_HT
```

CUDA Fortran Main Program

```
PROGRAM SPHriction_HT

    Use Module_Integrate

    Implicit None
    ... Vars ...

    ... Some Code ...
    Call Euler_single<<<grid,tBlock>>> 2501,nTotal,T_d,dTdt_d,dt

    ... Some Code ...

End Program SPHriction-3D
```

Subroutine call replaced with a CUDA kernel call

Figure 4-2 – Code Comparison between Fortran and CUDA Fortran

A typical GPU is composed of many thread processors that reside on a multiprocessor. The threads are grouped into thread bocks that are launched in a grid. The parallelization comes about by launching a large number of threads (see Figure 4-3 and Figure 4-4). Essentially, a thread should operate on a single element. So for example in SPH, a thread would be assigned to perform some calculations for a single SPH element. The number of concurrent threads that can be launched depends on the specific architecture of the GPU that will perform the calculations. For example, a GTX 660M, (NVIDIA Kepler generation), has a total of 1024 threads per block available. The device has 384 CUDA cores arranged on two multi-processors. It can have a total of 2048 threads per multi-processor (so a total of 4096 concurrent threads across the whole device).
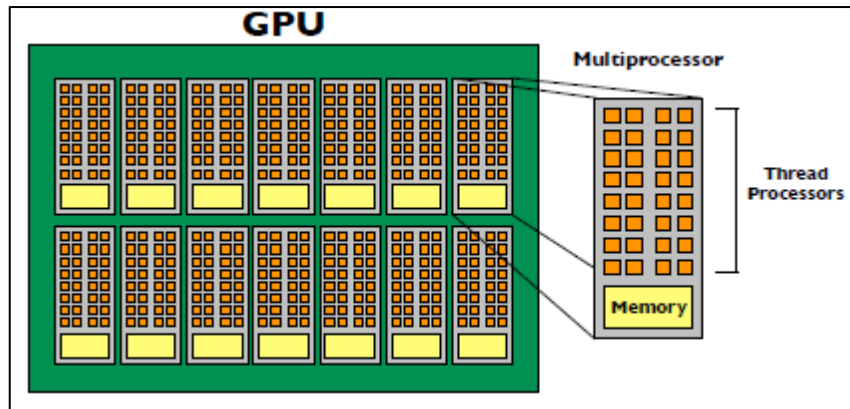
Figure 4-3 – GPU Architecture (Ruetsch and Fatica [250])
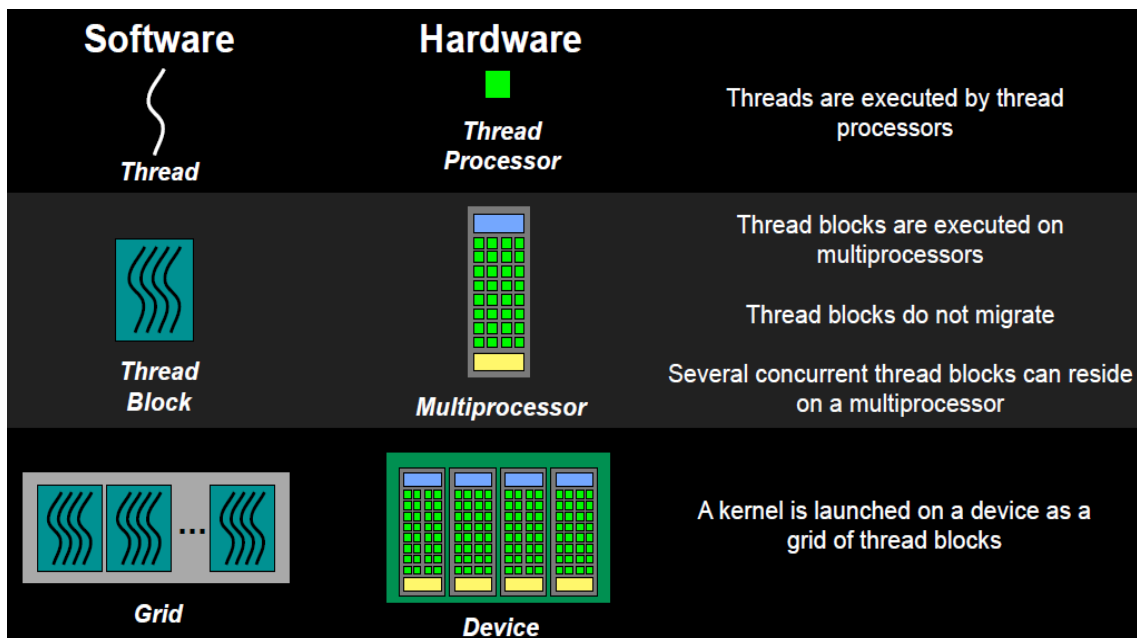
Figure 4-4 – Software and Hardware Schematic (NVIDIA, [268])

At this point in time, the GPU has its own memory banks. This means that variables must be transferred from the CPU (Host) to the GPU (Device) and back again. This is often a major bottle neck for scientific simulations on the GPU. Every attempt must be made to reduce the data transfers from Host to Device. Data transfers from Device to Host are needed to send results from the GPU to the CPU. This can be done in an intelligent way to limit the time wasted on transfers. Typically, the bandwidth of transfers from Host to Device is one to two orders of magnitude slower than memory bandwidth within the GPU (for

Kirk Fraser - April 2017

example, H to D bandwidth on the GTX 980Ti is around 6 Gb/s whereas on the GPU can achieve up to 366 Gb/s).

4.2.1 MEMORY

Efficient use of memory is the key to good performance in CUDA. Each GPU has a limited amount of memory that can be used, so care must be taken. In addition, different types of memory are available depending on the type of task. A schematic of the memory model for the CPU and the GPU is shown in Figure 4-5. The Host and Device have separate memory locations. Memory must be transferred across the PCIe Bus from the Host to Device and vice versa.
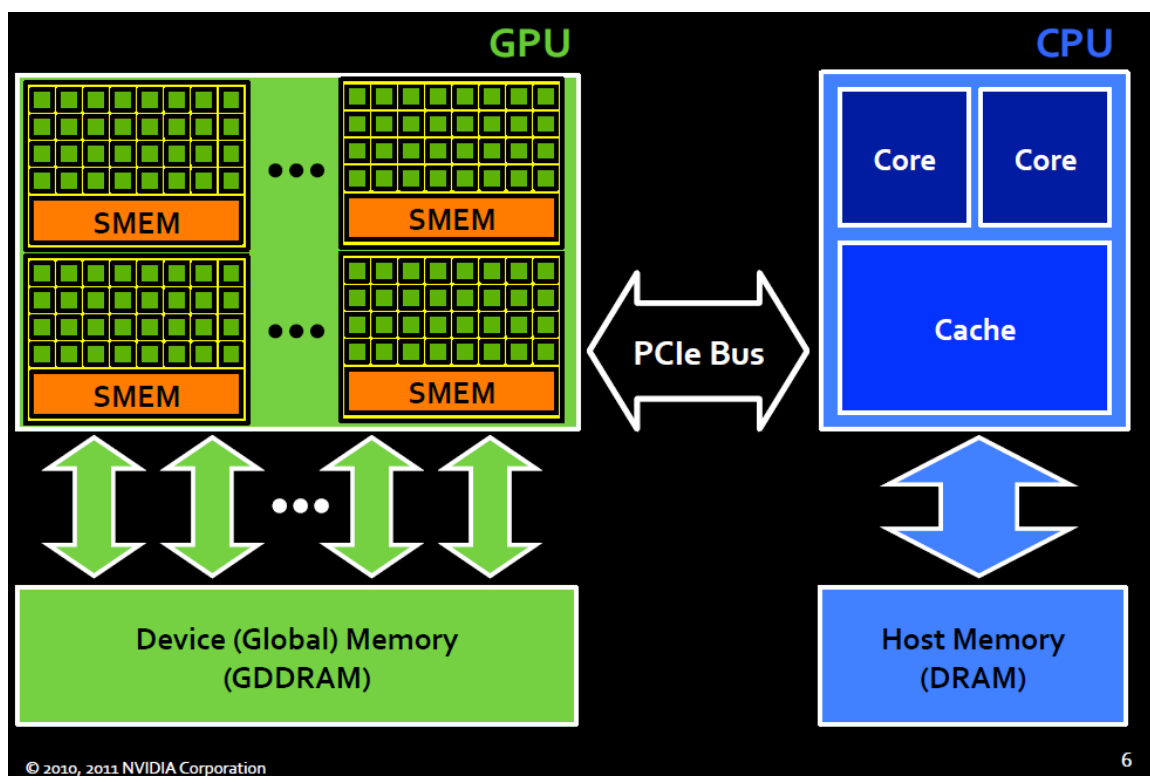
Figure 4-5 – Host and Device Memory Schematic (Balfour [212])

As seen in Figure 4-5, the GPU has a certain amount of device memory called global memory. This is typically three to twelve gigabytes on modern GPUs. The GPU also has a small amount of on-chip memory (such as constant, texture, and shared). Using constant memory is straightforward and is often used for variables that do not change during the simulation (hence the *constant* specifier). Texture and shared memory is difficult and very difficult to use respectively. Using these types of memory requires a

very different parallel strategy. Using texture memory was tested during the research project. No improvement in calculation time in the code resulted; the code simply became more complex. For this reason, only global memory is used. This approach makes the code efficient and robust without over complicating the programming strategy.

For a code that passes a significant amount of data from the device to the host and vice versa, the use of pinned memory can be beneficial. Other improvements can be obtained by interleaving the data transfers with kernel executions using *cudamemcpysync()*. Since the code developed in this research work is entirely on the GPU, these techniques complicate the code without providing any significant improvements. Neither pinned memory or asynchronous memory transfers have been implemented for these reasons. More information on the CUDA memory model can be found from [241, 245, 249, 250, 254].

4.2.2   CUDA KERNELS

At the heart of CUDA is the kernel; they are the subroutines that contain the parallel threads to be executed on the Device. There is a short example of a comparison between calling a Host subprogram and a CUDA kernel in Figure 4-2. More examples can be found from PGI's CUDA Fortran reference manual [245].

Global Kernel

A Global kernel is a subroutine that is called by the Host and is executed on the Device. Global kernels can only be called from Host code. They are defined using the syntax:

```
Attributes(Global) Subroutine SubName(Var1,Var2,...)
```

Other modules can be used within Global kernel, as can variables be declared as they typically would be in a standard Fortran90 subroutine. The *Value* variable qualifier:

```
Real, Value:: VarVal
```

is useful for passing Host variable directly to a kernel without explicitly passing the variable to the Device memory.

Device Subprogram

A Device subprogram is a subroutine that is called by the Device and is executed on the Device. Device subprograms can be called from within a Global kernel or from another Device kernel. They are defined using the syntax:

Attributes(Device) Subroutine SubName(Var1,Var2,...)

One slight limitation with Device subprograms is that they should be in the same Module as the Global kernel that calls the Device Subprogram. PGI has recently come up with a way to re-distribute kernels from one Module to another (allowing Device subprograms to be out of the calling kernels Module). The problem stems from linking the different modules together; at this point, there is no Device linker.

Invoking a Kernel

Calling a CUDA kernel in CUDA Fortran is very similar to the method used in CUDA C/C++. A set of triple chevrons are used specify the launch configuration:

Call SubName<<<grid,tBlock>>>(Var1,Var2,...)

Grid is the number of blocks to launch and $tBlock$ is the number of threads per block. In SPHriction-3D, kernel launches are performed by declaring grid and $tBlock$ to be $Dim3$ variables (structure of x,y,z). A typical declaration is:

tBlock = Dim3(256,1,1)
grid = Dim3(ceiling(Real(nTotal)/tBlock%x),1,1)

This will launch 256 threads per block and enough blocks to handle the $nTotal$ elements. This launch configuration is common for vectors. If we want to operate on a multi-dimensional array, we can simply increase the number of threads per block in the y and z dimension. The ceiling function is needed to

ensure that enough blocks are launched for an arbitrary number of elements. Once the kernel is launched with a specific configuration, the kernel has access to that specific number of threads. Threads, blocks and grids are used to index the elements of an array in a kernel. The common approach is:

```
Attributes(Global) Subroutine Increment(a,nTotal)
Implicit None
Integer, Value:: nTotal
Real:: a(nTotal)
Real, Parameter:: b = 1
i = (blockIdx%x-1)*blockDim%x+threadIdx%x
If (I >= 1 .and. I<= nTotal) Then
        A(i) = a(i) + b
End If
End Subroutine Increment
```

We need to prevent reading or writing past the actual number of elements in the actual kernel, this is done using the *if* statement shown above. The *blockDim%x* is the number of blocks launched in the first direction. The *blockIdx%x* is a unique value attributed to each block in the first direction. The *blockIdx%x-1* is different than in CUDA C because Fortran uses indices that start at one, not at zero. The *threadIdx%x* is the unique value for the thread in the block.

4.3 PARALLELIZATION STRATEGY OF THE SPH CODE

The basic strategy is to launch a thread for each SPH element that needs to be processed. For simulations with more than a few thousand elements, the threads are put in a launch queue and are executed in parallel as threads become available during run-time.

4.3.1 NEIGHBOR LIST DATA STRUCTURE

The fundamental idea behind SPH is to interpolate field variables on a point by point (collocation) basis by considering the interaction with a finite set of neighbors (j^{th} particles) that influence the point of concern (i^{th} particle). These neighbors need to be determined with a searching algorithm. How the neighbor list is laid out and organized can mean the difference between a simple parallel strategy and an incredibly cumbersome one.

In this work, the neighbors list is stored into a 2-D array. Other groups [246, 247, 251, 260, 269, 270] have used linked-lists that are slightly more efficient for memory use. The linked-list is a cumbersome and complicated data structure to modify and update. In the 2-D array, the first entry in each column is the ith particle, the second entry in each column is the number of neighbors for the ith particle and the subsequent entries are the neighbors of the ith particle. For example, in 3D, the maximum number of neighbors for an equally spaced grid is ~56. The array is setup with nTotal rows (total number of SPH elements in the simulation) by *nNeib_max+2* columns (*nNeib_max* is the maximum expected number of neighbors for any one particle in the active simulation). Two must be added since the first two slots in each column do not hold neighbor information. A schematic of the data structure is shown in Figure 4-6.

The array, *Neib(nTotal,nNeib_max)*, is allocated at the start of the simulation as an array of integers. This is a reasonable approach for computational solid mechanics problems because the number of elements does not change throughout the simulation. This is in contrast to a fluid simulation where the fluid will likely flow in and out of the simulations domain. This is not the conventional way to set up the neighbor list in most SPH codes. The reason for doing this is that this layout will lend itself very well for setting up an adaptive search approach (described in section 4.5).

(1,1)	(2,1)	(3,1)	...	(nTotal-1,1)	(nTotal,1)
(1,2)	...	...	...	...	...
(1,3)	...	...	...	...	...
...	...	...	...	...	...
(1,nNeib_max-1)	...	...	...	(nTotal-1,nNeib_max-1)	...
(1,nNeib_max)	...	...	...	...	(nTotal,nNeib_max)

Paticle 1	Paticle 2	Paticle 3	...	Particle nTotal-1	Particle nTotal
Number of neighbors for particle 1	Number of neighbors for particle 2	Number of neighbors for particle 3	...	Number of neighbors for particle nTotal-1	Number of neighbors for particle nTotal
First Neighbor	First Neighbor	First Neighbor	...	First Neighbor	First Neighbor
Second Neighbor	Second Neighbor	Second Neighbor	...	Second Neighbor	Second Neighbor
...	...	...	...	...	...
Last Neighbor	Last Neighbor	Last Neighbor	Last Neighbor	Last Neighbor	Last Neighbor
Padding if Number neighbors less than nNeib_max+2	Padding if Number neighbors less than nNeib_max+2	Padding if Number neighbors less than nNeib_max+2	Padding if Number neighbors less than nNeib_max+2	Padding if Number neighbors less than nNeib_max+2	Padding if Number neighbors less than nNeib_max+2

Figure 4-6 – Neighbor Array Layout, Neib(nTotal,nNeib_max)

4.3.2   STRUCTURE OF ARRAYS (SOA)

The way that the data is arranged in memory is very important for parallel programming in general and especially on the GPU. The GPU is set-up to fetch data from memory in chunks (up to 128 values at once, if the correct data layout is used). The idea is to try to arrange the data in a contiguous manner so that as the data is needed, it can be accessed quickly by the GPU. A common data layout is to organize using an array of structures (AoS). For example, the data for a particle could be setup as:

```
Type Particle
        Integer :: ID
        Real :: X
        Real :: Y
        Real :: Z
End Type
Type(Particle), Dimension(:), Allocatable:: SPHelement
```

The drawback of this approach is that the data will be organized in the memory resisters as shown in Figure 4-7. In this example, there are three SPH elements (SPHelement is allocated to 3). In comparison the data could be flipped around and stored as a structure of arrays:

```
Type Particle
        Integer, Dimension(:), Allocatable :: ID
```

```
        Real, Dimension(:), Allocatable :: X
        Real, Dimension(:), Allocatable :: Y
        Real, Dimension(:), Allocatable :: Z
End Type
Type(Particle):: SPHelement
```

Although the difference in data storage is subtle, the improvement is drastic for parallel programming. This allows the GPU to fetch large chunks of data at once since now the values are arranged contiguously in memory. Figure 4-7 shows a simple SoA for three SPH elements.

The SoA approach is used in this research project and is found to be at least 4x faster than AoS. For some data structures, SoA can be 8x or even 16x faster (depending on the data access required by the specific algorithm).

SPHelement Data Structure (AoS)

ID1	X1	Y1	Z1	ID2	X2	Y2	Z2	ID3	X3	Y3	Z3

SPHelement Data Structure (SoA)

ID1	ID2	ID3	X1	X2	X3	Y1	Y2	Y3	X1	X2	X3

Figure 4-7 – AoS and SoA data structures

4.3.3   SPH SUMS (REDUCTIONS)

The SPH sums are the backbone of the SPH method. They are essentially reductions performed for each set of particle interactions. Typically a reduction requires 30 to 60 operations per particle (depends on the ith particles number of neighbors). Since the number of elements in an SPH simulation is often many orders of magnitude greater than the number of neighbors for each particle, an attempt to perform the reductions in parallel would be inefficient.

The idea is that a tread is set up for each element. Within the thread, a loop is used to accumulate the quantity that is being calculated. Reductions are very tricky to run in parallel in CUDA; there are many

ways to do it, but for a large set of relatively small reductions, it is more efficient to perform the reduction

in series. An example of an SPH sum is shown in Figure 4-8.

```
Attributes(Global) SUBROUTINE SPH_HeatTransfer(dTdt,nTotal)

! SPH heat transfer based on Jubelgas
IMPLICIT NONE

Include 'mat_const.inc'

Integer, Value:: nTotal
Integer:: i, j, k, d
Real(fp_kind):: dx(Dim), dx_sq(Dim),r, r_sq, small_val, x_dot_dw
Real(fp_kind), Intent(OUT):: dTdt(nTotal)

small_val = 1.0E-06

i = (blockIdx%x-1)*blockDim%x + threadIdx%x

If (i >= 1 .and. i <= nTotal) Then
dTdt(i) = 0.0
x_dot_dw = 0.0
        Do k = 3, Neib_t(i,2)+2
            j = Neib_t(i,k)
            Do d = 1, Dim
                dx(d) = x_t(i,d) - x_t(j,d)
                dx_sq(d) = dx(d)*dx(d)
            End Do
            x_dot_dw = (dwdx_t(i,k,1)*dx(1) + dwdx_t(i,k,2)*dx(2) + dwdx_t(i,k,3)*dx(3)) &
                                        / (dx_sq(1) + dx_sq(2) + dx_sq(3))
            dTdt(i) = dTdt(i) + (1.0/Cp_t(i))*(mass_t(j)/(rho_ht_t(i)*rho_ht_t(j))) &
                                * ((4.0*k_ht_t(i)*k_ht_t(j))/(k_ht_t(i)+k_ht_t(j))) &
                                * (T_t(i)-T_t(j))*x_dot_dw
        End Do
End If

End Subroutine SPH_HeatTransfer
```

Figure 4-8 – SPH sum example

4.4 PERFORMANCE COMPARISON – HEAT TRANSFER SIMULATION

For this test, SPHriction-3D has been simplified to test the thermal solver only. The program consists

only of the following four main components:

1. Neighbor Search (brute force, $\mathcal{O}(N^2)$)

2. Determine surface particles and their normal vectors

3. Perform SPH heat transfer calculations (equation (3-99))

4. Perform explicit time integration using forward difference scheme (Euler),

Timers have been wrapped around each of the main subroutines in the code. A comparison between

the serial implementation of the code on the CPU and a parallel implementation on the GPU will be

performed. The system used to perform the simulations is summarized as:

- Intel I7-3630QM CPU

- NVIDIA GeForce GTX 660M GPU

- NVIDIA Tesla K20 GPU (NVIDIA test drive on a Windows Server Cluster)

- NVIDIA Tesla K40 GPU (NVIDIA test drive on a Windows Server Cluster)

Full specifications for the processing units can be found at

- http://ark.intel.com/products/71459/Intel-Core-i7-3630QM-Processor-6M-Cache-up-to-3_40-GHz

- http://www.geforce.com/hardware/notebook-gpus/geforce-gtx-660m/specifications

- http://www.nvidia.com/object/tesla-servers.html

Five different number of elements will be considered;

1. 2^{15} (32,768 elements)

2. 50,000

3. 2^{16} (65,536 elements)

4. 100,000

5. 200,000

For each of the element number cases, each set of tests will be carried out on the different processors as previously outlined. The 2^{15} and 2^{16} cases are chosen since the GPU is expected to be efficient at powers of two.

4.4.1 PERFORMANCE RESULTS

For the performance comparison, a comparison of the total simulation time for the various cases is compared in Figure 4-9. In general, a speed up of 10x to 20x is obtained compared to the CPU version with the GTX 600M. These performance values are conservative since the comparison is with respect to a relatively high speed CPU to a moderate GPU. The results from the Tesla K20 and K40 are significantly faster. The best speedup on the K20 is for the 65636 element case with a factor of 94.9x. The best for the K40 is found to be 127.3x. A graph showing a comparison of simulation times is shown in Figure 4-9. A comparison of the speed up factors is provided in Figure 4-10.
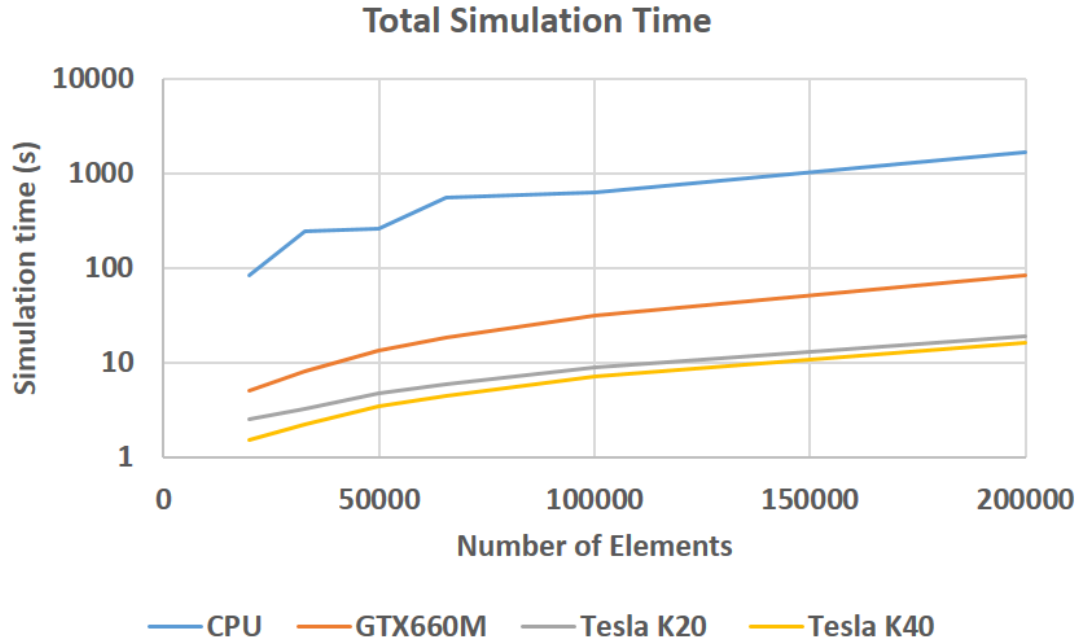
Figure 4-9 – Total simulation time comparison
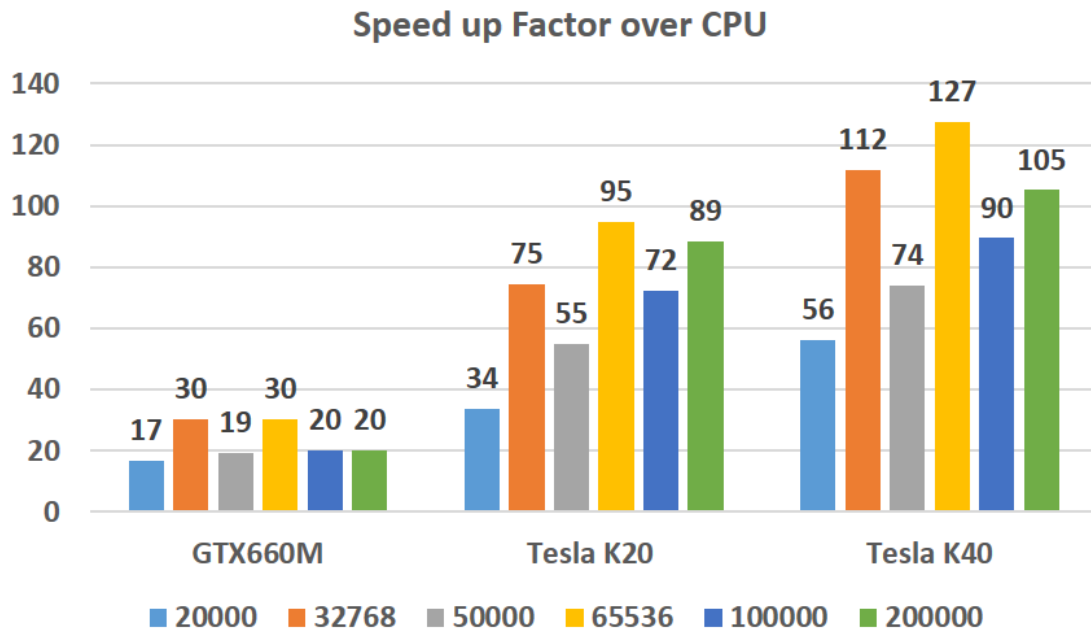
Figure 4-10 – Speed up factor comparison

The Tesla K40 GPUs are dedicated calculation units that have no graphics capabilities. Each card is priced at ~5000.00$. They are about 20 times more expensive than the GTX 660M. In a situation where

Kirk Fraser - April 2017

only the fastest simulation times will do, it is obvious that the increases in computing power (and cost) are well worth it. On the other hand, recently, NVIDIA has released the GTX 980Ti, it is ~850.00$ and is actually slightly faster than the K40 for single precision calculations.

4.5   ADAPTIVE NEIGHBOR SEARCH ON THE GPU

One of the most common searching methods involves binning the particles into cubes with length of 2h. The search is then accomplished by searching only to the neighboring bins of the bin in which the concerned particle resides. Figure 4-11 shows a schematic of a two dimensional domain that has been set up for a Cartesian grid for particle binning. In this situation, the search is carried out only with particles in the 8 neighbor bins in 2D or 26 in 3D. This method is a drastic improvement over the direct search without being overly complicated.

The implementation of the neighbor search on the GPU is slightly different because pairwise interactions cannot be leveraged. Typically, for a CPU algorithm, the neighbor list will only list the interacting particles once (this is done by requiring that j be greater than i in the search). As such only 13 cells must be searched in the cell search method. Conversely, the neighbor search on the GPU is performed for all the pairs; this means that all 26 neighboring cells are searched.
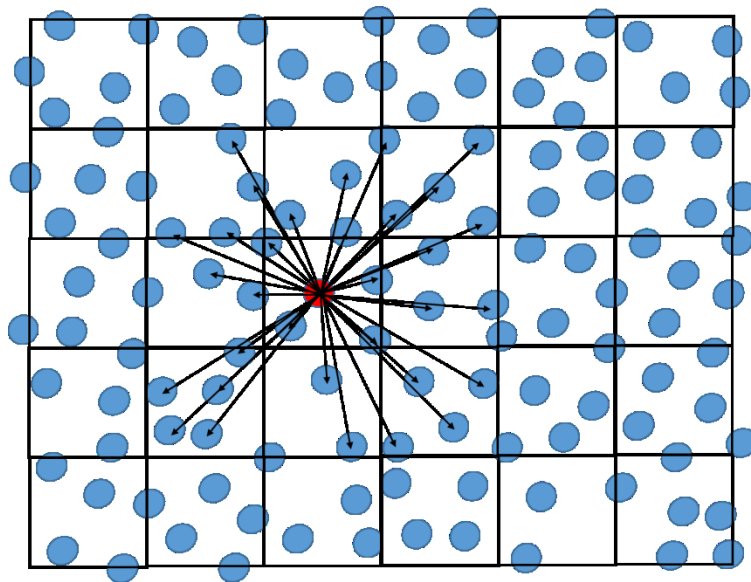
Figure 4-11 – Schematic of Cell Search Method

Typically, a solid mechanics model will contain regions in the domain that are either distinctly elastic or predominately plastic. For regions in the model with elastic strains, simulation time can be reduced by not evaluating the neighbors for the particles in the elastically deforming zone. Such a procedure sets up an adaptive approach where only the elements in the model undergoing significant plastic deformation will have their neighbors updated.

Very little work has been undertaken in the research community for adaptive neighbor searching. To our knowledge, the only group that has worked on this is Pelfrey and House [271]. They proposed an adaptive neighbor search for fluid dynamic simulations. Their algorithm is mainly intended to speedup SPH fluid graphics rendering (for video games or cinematography). One of the drawbacks of their method (and most likely, why very little work has been done in this field) is that for a typical fluid simulation, not re-evaluating the smoothing function for each particle can introduce significant error. They mention in their publication that the results are "visually" similar with the adaptive search. They do not compare their method quantitatively.

The adaptive neighbor search that is proposed is based on the cell search method. Some slight changes are needed to the standard algorithm. In this work, a linked list data structure is not used as this would be cumbersome and would slow down the algorithm. Insertion and deletion from a linked list is somewhat complicated. Instead a static 2D array with the ith columns listing the neighbors for the ith particle. This method is less efficient from a memory standpoint because we must fix the amount of memory allocated for the rows in the array (when $h_{scale} = 1.2$, enough memory must be allocated for at least 56 neighbors per particle). That being said, the adaptive algorithm can be modified to different data structures as well as for different search methods (Verlet, Octree…).

The adaptive search process is as follows:

➢ Perform a neighbor search over the whole domain for the first cycle. This involves

→ Finding the bounds of the simulation

→ Determining the number of cells in each dimension

→ Binning particles into their cells

- $\rightarrow$ Searching for neighbors
- ➢ Calculate conservation of mass, momentum, material models, external forces, etc
  - $\rightarrow$ The effective plastic strain increment is determined based on the material model
- ➢ Build a list of elements with an effective plastic strain ($\Delta\varepsilon^p$) that surpasses a threshold, adapt_thresh. This is done by:
  - $\rightarrow$ Cycle through the elements, if an elements $\Delta\varepsilon^p$ is greater than *adapt_thresh* place the element id in a list called *AdaptSearchLis*t
  - $\rightarrow$ Mark the element in another array called *AdaptNode*, this array can be used to show which elements are adapted in the post-processor
  - $\rightarrow$ Increment a counter (*nNodeAdaptSearch*) to keep track of how many elements are in the *AdaptSearchList*. On the GPU, we use an *atomicinc()* function to prevent incorrect incrementation of *nNodeAdaptSearch*.
- ➢ Pass *nNodeAdaptSearch* and *AdaptSearchList* to the cell search subroutine. The search is performed only for the nodes in the *AdaptSearchList*
  - $\rightarrow$ The neighbors are updated
  - $\rightarrow$ The smoothing function values are updated
- ➢ The neighbor list and smoothing function values are kept and re-used for the SPH elements that are not on the *AdaptSearchList*

The *AdaptNode* array is used to visualize which elements are processed in the adaptive search. This is an important consideration to ensure that the field of elements adapted is not overly discontinuous.

Important Notions that Make the Algorithm Efficient

There are a few concepts that lead to the efficiency of the algorithm:

1. Often, a large simulation domain in CSM problems must be used to model correctly the boundary conditions and other important aspects. There will typically be large portions of the domain that are responding elastically. This is a stark contrast to a fluid simulation where the whole domain is "active".

2. For infinitesimal strains, the Piola-Kirchhoff stress tensor is equivalent to the Cauchy stress tensor. This is important because the total Lagrangian formulation in SPH classically uses the Piola-Kirchhoff stress tensor.

3. In the elastic regions, the total Lagrangian formulation can be written using the Cauchy stress tensor and the deviatoric stresses can be found by integrating the Jaumann rate equation.

4. The region of the domain that is deforming elastically does not need their neighbor list or smoothing function updated; only the regions with plastic deformation.

The approach is equivalent to using the total Lagrangian formulation in the elastic zones. The elements that are undergoing plastic deformation will have their neighbors list and smoothing function re-evaluated. This is akin to using the Eulerian formulation in the plastic areas. This approach was presented at the 13th International LS-DYNA conference, full details and examples are available in the associated paper [272].

4.6 AN EFFICIENT PARALLEL SURFACE TRIANGULATION ALGORITHM

Free surface triangulation is an important realm of research in many industries. Of particular note is the video game industry where a set of points must be rendered as a surface using triangulation [273, 274]. Other industries include shipbuilding [275], machining [276], shoreline erosion [277], as well as many other fields [278-283]. In the case of this research project, surface triangulation is a powerful means to show the results from the meshfree simulations. The general triangulation approach can be summarized by first determining the particles located on the free surface (as described in section 3.9.3), then the particles are joined together to form three node triangular elements as shown in Figure 4-12.

Performing a true Delaunay triangulation is a complicated task, even more so in parallel. However, in the case of visualizing the meshfree results, a simple patchwork of triangles (non-Delaunay) is sufficient. In this sense, there will be redundant triangles, which effectively wastes memory and calculation time. For the simulation models in this project, the increased memory is negligible and the calculation time can be drastically reduced because creating such a patchwork of triangles can be accomplished very efficiently on the GPU (this is not the case for true Delaunay triangulation).
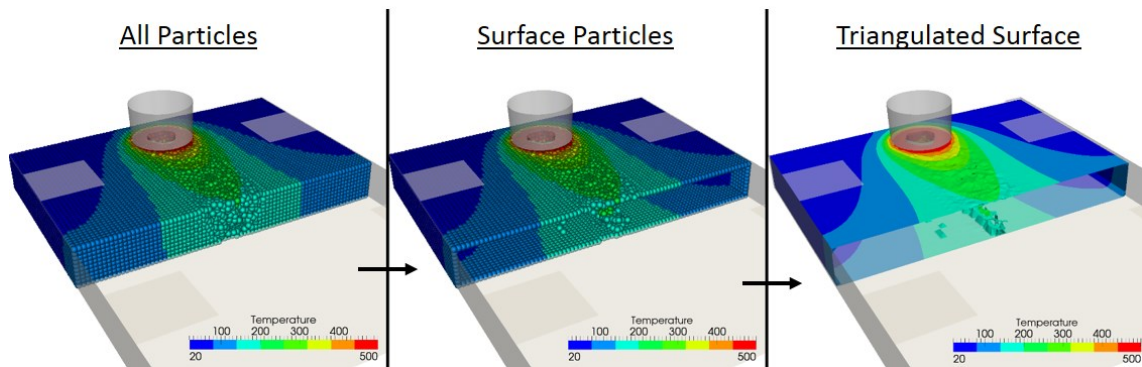
Figure 4-12 – Surface triangulation sequence

The triangulation algorithm uses the surface node and neighbor arrays. The idea is to connect the ith particle with its j neighbors that are within a cut off distance d_{cut}. To do this, a list of surface only neighbors, $NeibSurf$, is established from the full neighbors list. Next, the domain is split into a uniform grid (spatial hashing) of size $\text{SurfCell}_{\text{factor}}$. Each of the surface particles then have a unique hash ID, which allows for an efficient parallel implementation on the GPU. To create the surface triangles, the surface neighbor list is iteratively processed forming triangles from tuples of neighboring surface nodes. The algorithm is of $\mathcal{O}(\frac{1}{2}Nm!)$, where N is the total number of surface particles, and m is the average number of surface neighbors per particle. The factorial comes about because the neighbors list must be recursively traversed. The algorithm is rather complex in order to achieve good performance on the GPU and is provided in the appendices.

5 Modelling, Validation and Test Cases

In the previous sections, the groundwork was laid to afford a strong foundation for the numerical simulation of the FSW process. The simulations presented in this chapter are all performed using the developed meshfree code, SPHriction-3D. The code is the first of its kind, providing a robust and efficient means to solve complex, large deformation, multi-physics problems on the GPU. In each of the examples, the goals will be threefold:

1. Develop a phenomenological based model of different FSW processes capable of predicting temperature history and fields, deformation, defects, as well as force and torque

2. Show that the results from the numerical models closely fit those from the experimental tests

3. Provide valuable insight into the various processes, with attention primarily on the underlying physics

Four different FSW processes will be presented in this chapter: butt joint FSW, lap joint FSSW, complex joint FSW, and bobbin tool FSW of a hollow core section. The chapter is organized into the following sections:

Section 5.1 presents the numerical model of a butt joint FSW weld. The experimental setup is outlined, and results from the simulation are compared. Excellent correlation for temperature, deformation, and defects is obtained. The experiments for this work were carried out in the CURAL lab at UQAC.

Section 5.2 introduces a model of the friction stir spot welding process for a lap joint. Three test cases will be compared with experimental results for the tool force and torque. This set of experiments was performed at the "Centre de Soudage par Friction-Malaxage" at CEE-UQAC.

Section 5.3 provides one of the first fully coupled large deformation thermo-mechanics simulation of a complex FSW joint geometry. The joint is common for many structural applications such as bridge decks, train roofs, and aircraft fuselages. Three cases will be presented, the first case serves to validate the

model, and the remaining models are used to investigate viable options for reducing the quantity of flash. This set of experiments was performed at the "Centre de Soudage par Friction-Malaxage" at CEE-UQAC.

Section 5.4 offers a simulation model of the bobbin tool FSW process. A hollow core section is welded in the lab and the results from the simulation are compared. The focus in this model is the validation of the size and location of various defects common in the bobbin tool FSW process. This set of experiments was performed at the "Centre de Soudage par Friction-Malaxage" at CEE-UQAC.

Section 5.5 focuses on a series of parametric studies for the butt joint FSW model presented in section 5.1. These studies are designed to highlight certain aspects of the FSW process and shed light on the physics behind the scene.

Section 5.6 wraps up the chapter with a look at the residual distortion and the associated perceived stresses in a FSW weld. The simulation model goes through all the phases of the weld process from initial clamping to complete cool down of the plates.

Prior to working on simulating the FSW process, a large effort was undertaken to verify each of the various parts of the simulation code. A number of the verification test cases (elastic deformation, plastic deformation, heat transfer, momentum conservation, contact, etc.) are provided in the Appendices.

5.1   BUTT JOINT WELD – FRICTION STIR WELD

The majority of the research work focused on in this project is that of a butt joint weld of two 12.7 mm AA6061-T6 plates. This joint is easy to study in the lab due to its simplicity and is the subject of focus of many research groups. The experimental setup is shown in Figure 5-1; the top surface of the two plates is painted with a high temperature tolerant flat black spray paint. This helps to obtain good measurements with the infrared thermal camera (Thermal camera) because the paint has a constant emissivity of ~0.95. Note that the region where the FSW tool will progress is not painted, as this will negatively affect the friction between the tool and the work pieces. Furthermore, if the weld region were to be painted, the paint particles would become mixed into the weld joint, causing internal flaws. Thermocouples are embedded in the work pieces according to Figure 5-2. The thermocouples are located 2.15 mm from the

top surface of the work pieces. The experiments are performed in the CURAL lab at UQAC using FSW machine that was converted from a CNC.
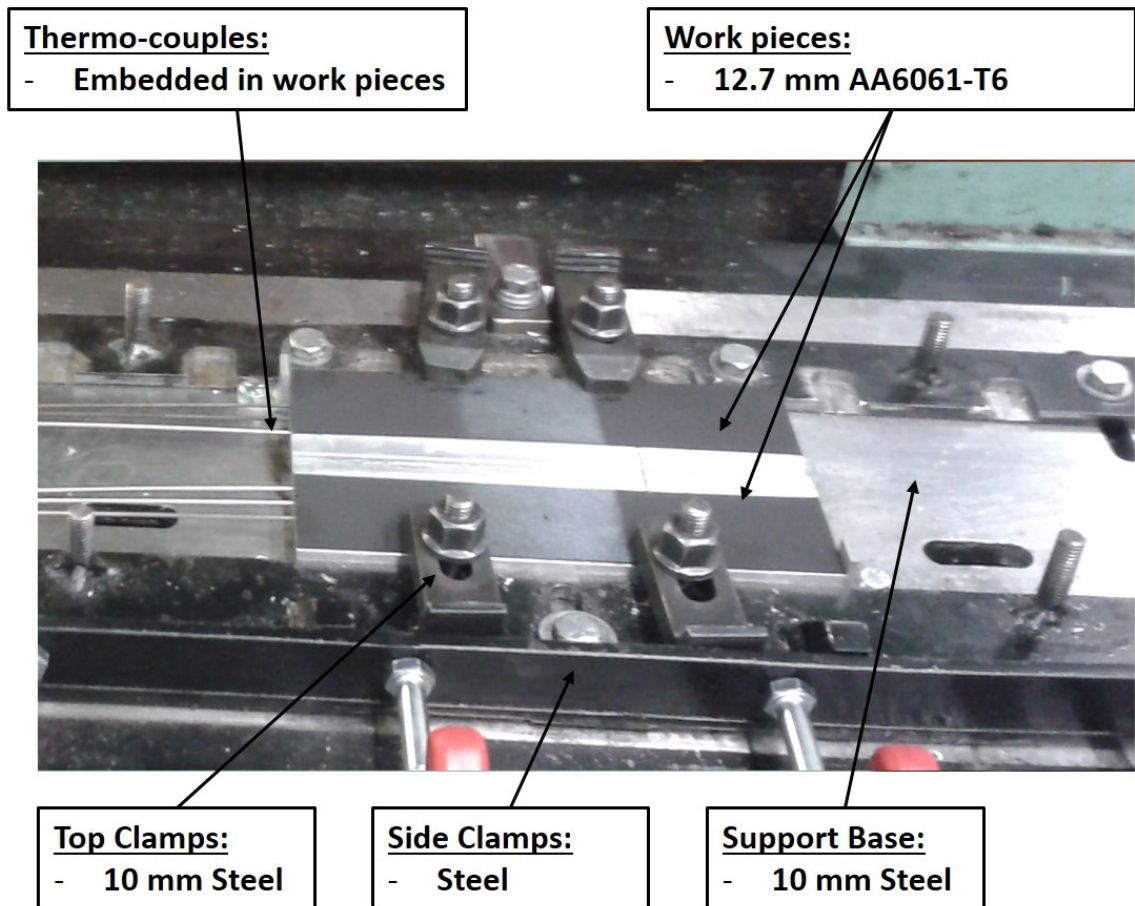
Figure 5-1 – Experimental setup for butt weld joint

The experimental setup allows for determining the temperature history within the plates as well as on the surface of the plates using a Thermal camera. The setup does not allow for determining force and torque. A method to find the residual stresses (discussed in Section 5.6) in the completed joints has also been developed using a non-destructive approach with digital image correlation (DIC).
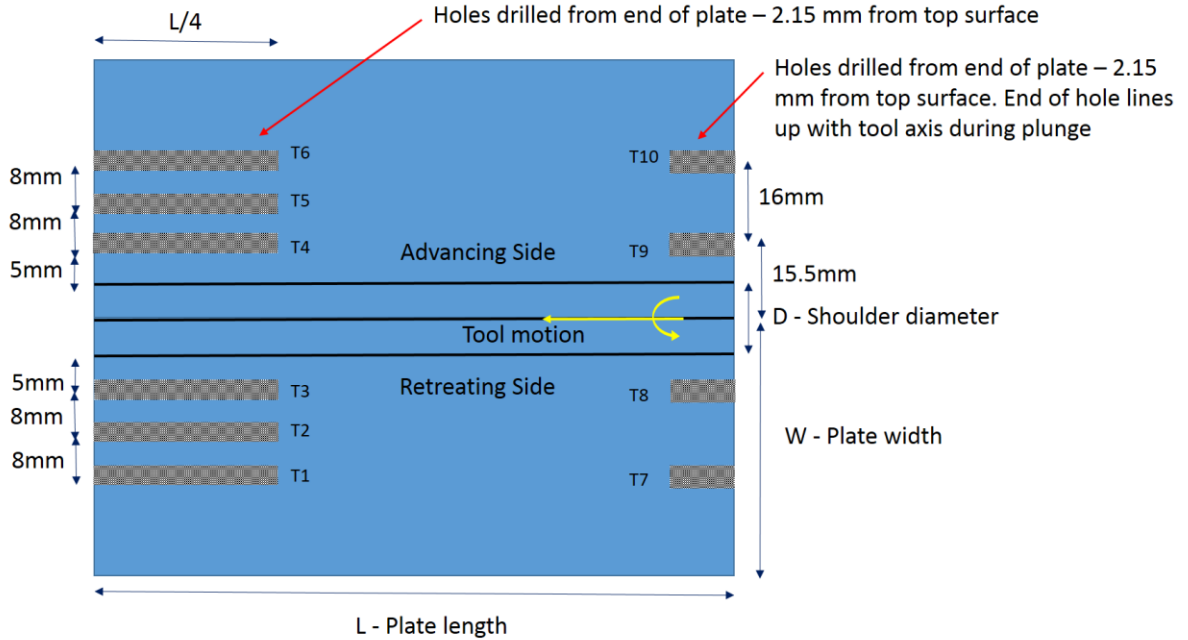
Figure 5-2 – Thermocouple arrangement for butt weld joint

The tool used is shown in Figure 5-3. The tool has a threaded design and a smooth concave (0.25 mm concavity) shoulder. The shoulder diameter is 21.6 mm, upper pin diameter is 11.0 mm, and lower pin diameter of 8.4 mm. The length of the pin is 10.8 mm. The pin is threaded with $p_{thread} = 1.25\ mm$, and the shoulder is featureless. The CNC machine is only capable of position control with constant tool velocity.

A number of tests have been performed in the lab. Three cases will be presented from the experimental results:

1- 800 rpm 305 mm/min advance

2- 800 rpm 660 mm/min advance

3- 800 rpm 1069 mm/min advance

In all cases, the plunge speed is 38 mm/min and the dwell time is 5.0 s.

Figure 5-3 – FSW tool for butt weld joint

A typical temperature history profile from the experiments is shown in Figure 5-4. Note that the data acquisition typically starts prior to the onset of the FSW process. This is done to ensure that important information is not missed during the tests. Because of this requirement, the experimental data is time shifted to ensure that the start of the temperature data corresponds to the start of the numerical simulation. The different phases shown in the graph are:

A   Plunge; pin engaged

B   Plunge; pin and shoulder engaged

C   Dwell

D   Advance

Although ten thermocouples were used in the experimental work, only the results from TC3, TC4, TC8, and TC9 will be used to compare results from the simulation models. Using more measurement points tends to overcrowd the graphs and reduces their overall readability. TC8 and TC9 are located close to the weld line at the plunge location. TC3 andTC4 are located ¾ of the way along the work pieces (in the weld direction, again close to the weld line).
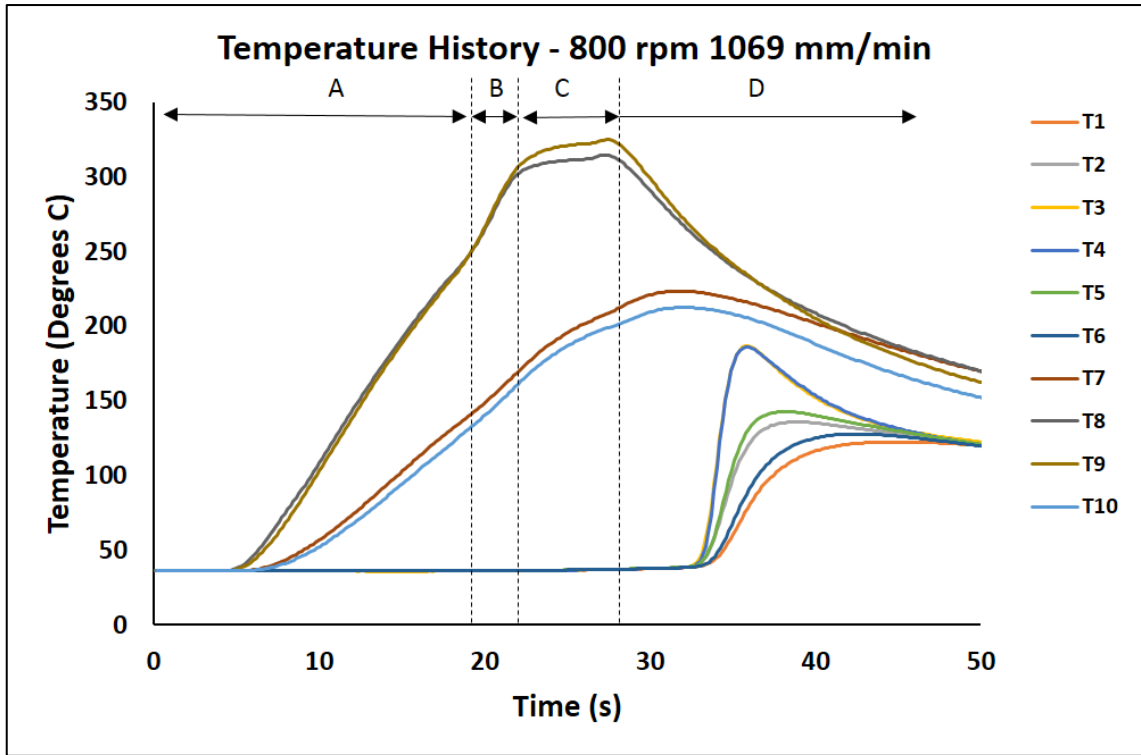
Figure 5-4 – Typical temperature history for butt weld joint

5.1.1   MODEL DESCRIPTION

The simulation model is composed of a combination of SPH elements and rigid finite elements (as shown in Figure 5-5). The two aluminum work pieces are modeled with elastic-plastic-thermal SPH elements. The tool and support base are modeled with zero thickness triangular plate rigid finite elements. A single layer of SPH elements is used to take into consideration the thermal mass of the tool and support base. This is a robust approach to get the thermal energy in the system correct. However, the method does not reproduce the temperature gradients within the tool and base. An equivalent thickness of 10 mm for the base, 4 mm for the pin, and 10 mm for the shoulder was used. This approach is not mandatory for the simulation code; in fact, the tool and support base could be entirely modeled with SPH elements. The trade off in temperature resolution would come in the form of increased computational time. Under certain situations, resolving the full thermal problem in the tool and base may be warranted. Such situations could be testing the sensitivity of the welding parameters on the base material and dimensions, or perhaps an investigation of the subsequent tool wear with different tool shapes.
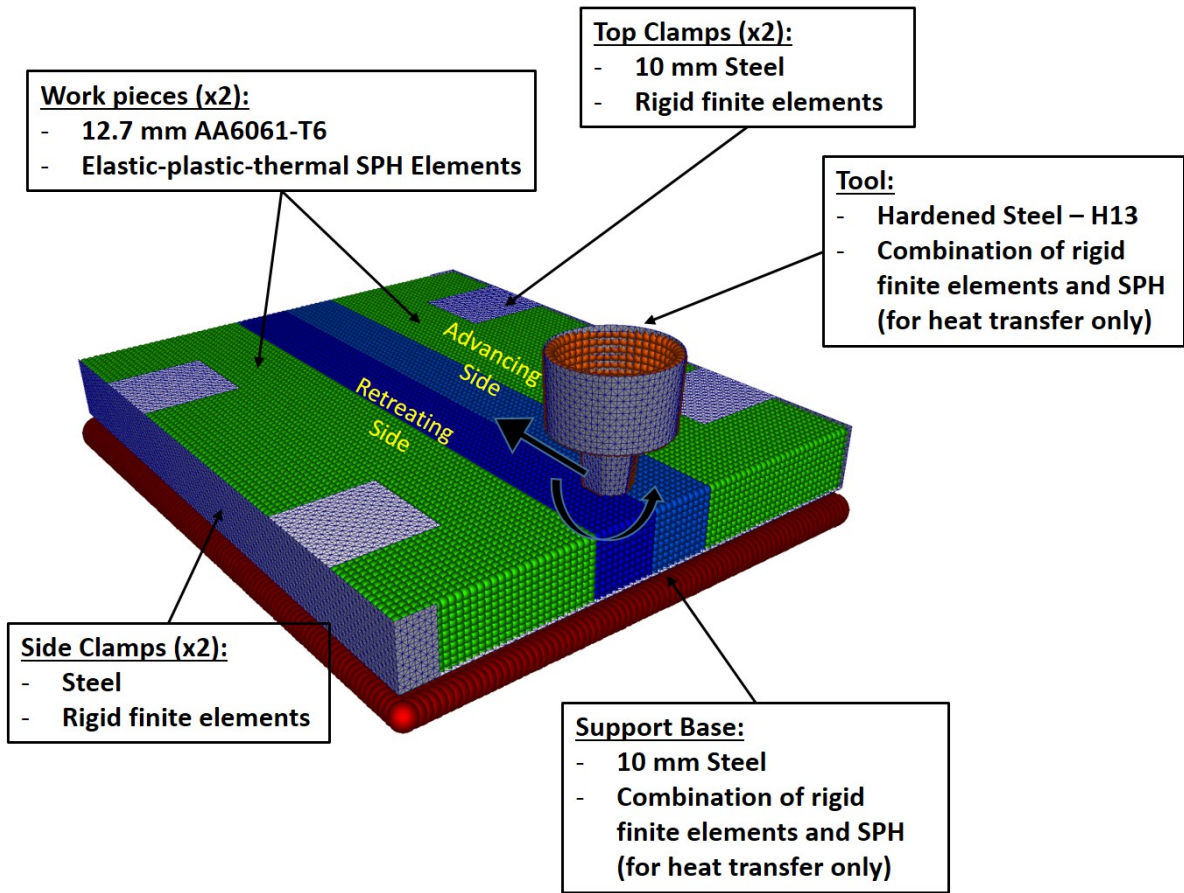
Figure 5-5 – FSW model of butt weld joint

The element spacing is setup on a uniform grid for the work pieces with an inter-particle spacing of 1.27 mm, this gives 10 elements through the thickness of the aluminum plates. The tool and base are meshed to provide an average finite element size of ~1mm; this is an efficient size for the contact calculations. A smaller tool mesh would increase the contact search time. Larger elements would have an effect on the number of contact pairs and lead to the need to perform the contact pair search more frequently.

The time step is controlled by the CFL condition, $CFL = 0.7$ and $dt_{min} = 2.07x10^{-7}$ s. $\tilde{V}_{SF} = 30$ (velocity scaling factor) has been used for this case study. The model is run in three distinct phases, plunge, dwell, and advance. In this case, the tool does not ramp up smoothly to full advance speed. The CNC machine is powerful enough to cause the tool to "jump" from not moving to full advance speed in a fraction of a second. The numerical model has been setup to mimic this "infinite acceleration" behaviour.

Four different material models will be used to test the ability of each to reproduce the correct material flow and temperature history. The material models are:

1   Johnson-Cook (JC) with perfect plasticity. Material parameters as shown in

2   The damage constants used in the cumulative damage friction model were listed in section 3.12.The various test cases were run in SPHriction-3D; the results will be presented according to the specific test cases (see section 5.1.1 for details of the cases). All graphs are made in GNUplot, and graphical results from the FSW models are presented using ParaView v3.12.0. The force and torque results are filtered using a low pass filter.

3   Table 5-2, B_{JC} and n_{JC} are set to zero

4   JC with strain hardening. Material parameters as shown in

5   The damage constants used in the cumulative damage friction model were listed in section 3.12.The various test cases were run in SPHriction-3D; the results will be presented according to the specific test cases (see section 5.1.1 for details of the cases). All graphs are made in GNUplot, and graphical results from the FSW models are presented using ParaView v3.12.0. The force and torque results are filtered using a low pass filter.

6   Table 5-2

7   Newly developed material model, FKS, with material parameters as shown in Table 3-2

8   A combination of JC strain hardening and the FKS (JC-FKS) strain rate and thermal softening form:

$$\sigma_y(\hat{\varepsilon}^p, \dot{\varepsilon}, T^*) = \left(A_{JC} + B_{JC}\hat{\varepsilon}^{p n_{JC}}\right)\Lambda(\dot{\varepsilon}, T^*)\Theta(T^*) \tag{5-1}$$

where $\Lambda(\dot{\varepsilon}, T^*)$ and $\Theta(T^*)$ were presented in eqn. (3-128) and (3-130). The resulting flow stress curves from eqn. (5-1) are shown in Figure 5-6. Notice that the strain hardening causes the flow stress to diverge away from the compression test results found experimentally.
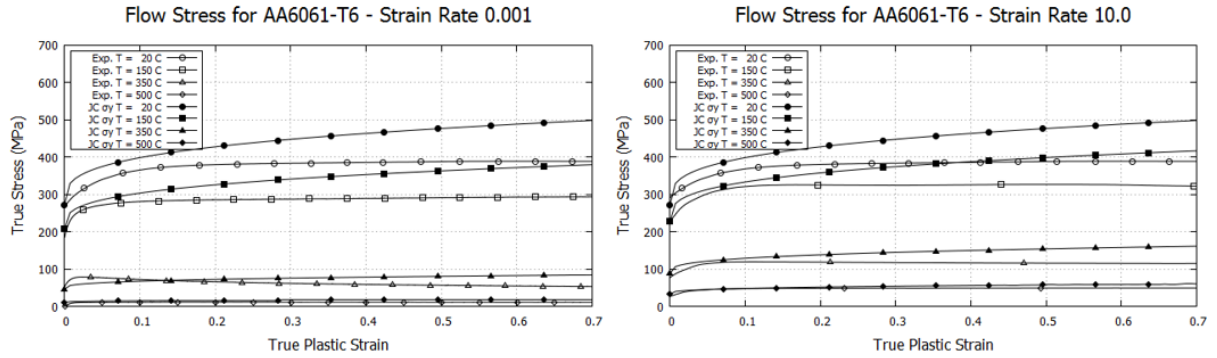
Figure 5-6 – Flow stress from JC-FKS (from eqn. (5-1)) compared to compression test

All the cases in this section use non-linear thermal material properties. The thermal conductivity and heat capacities are parabolic and quintic functions respectively (shown in Figure 5-7). The thermal properties are based on Mills [284]:

$$C_p(T) = 4.6 \times 10^{-10}T^5 - 4.9 \times 10^{-7}T^4 + 1.8 \times 10^{-4}T^3 - 2.8 \times 10^{-2}T^2 + 2.2T + 861 \tag{5-2}$$

$$k(T) = -4.3 \times 10^{-4}T^2 + 0.17T + 175 \tag{5-3}$$

Unfortunately, the thermal properties are not well known and the functions should be seen as approximate values. The modulus of elasticity is also taken as a parabolic function of temperature based on Eurocode 9 [285] (shown in Figure 5-7):

$$E(T) = -2.13 \times 10^5 T^2 - 3.0 \times 10^6 T + 7.0 \times 10^{10} \tag{5-4}$$
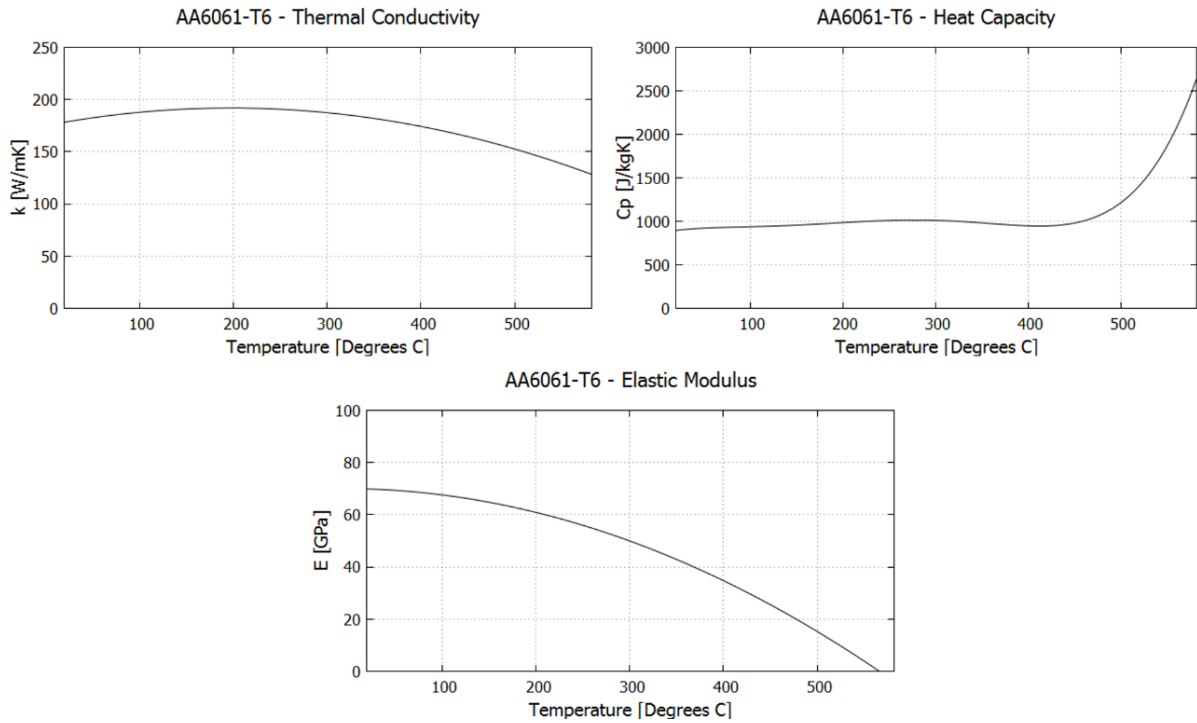
Figure 5-7 – Thermal properties of AA6061-T6

A series of simulations have been performed to highlight the code's ability to provide extensive insight into the effect of different process parameters, flow stress models, and friction models. The test cases are divided into three categories:

1   Material model testing

   a.   JC perfect plastic with stick-slip cumulative damage friction (CDF) – 800 rpm, 1069 mm/min

   b.   JC strain hardening with CDF – 800 rpm, 1069 mm/min

   c.   FKS with CDF – 800 rpm, 1069 mm/min

   d.   JC-FKS with CDF – 800 rpm, 1069 mm/min

2   Friction model testing

   a.   Slip friction (Coulomb) with JC-FKS – 800 rpm, 1069 mm/min

   b.   Stick with JC- FKS – 800 rpm, 1069 mm/min

   c.   Stick-slip, δ = 0.5, with JC- FKS – 800 rpm, 1069 mm/min

   d.   CDF with JC-FKS – 800 rpm, 1069 mm/min (same as case 1d)

3   FSW process parameter testing

    a.   800 rpm, 305 mm/min - CDF with FKS

    b.   800 rpm, 660 mm/min - CDF with FKS

    c.   800 rpm, 1069 mm/min - CDF with FKS (same as 1c)

Table 5-1 – Thermal-physical properties simulation model components

Work-Pieces Mechanical (AA6061-T6)   Work-Pieces Thermal

Parameter	Value	Units	Parameter	Value	Units
Density, ρ	2700.0	Kg/m³	Conductivity, k	Figure 5-7	W/mK
Initial yield, σ_{y_0}	276.0	MPa	Heat capacity, C_p	Figure 5-7	J/kgK
Shear modulus, G	26.3	GPa	Tool and Base Thermal (Hardened Steel)		
Room temperature, T_R	20.0	°C	Conductivity, k	55.0	W/mK
Melt temperature, T_{melt}	582.0	°C	Heat capacity, C_p	485.0	J/kgK
Speed of sound, c	4722	m/s	Density, ρ	7850.0	Kg/m³

The damage constants used in the cumulative damage friction model were listed in section 3.12. The various test cases were run in SPHriction-3D; the results will be presented according to the specific test cases (see section 5.1.1 for details of the cases). All graphs are made in GNUplot, and graphical results from the FSW models are presented using ParaView v3.12.0. The force and torque results are filtered using a low pass filter.

Table 5-2 – Johnson-Cook Parameters – AA6061-T6

Parameter	Value	Units
A_{JC}	276.0	MPa
B_{JC}	255.0	MPa
n_{JC}	0.3	-
m_{JC}	1.0	-

5.1.2 SIMULATION RESULTS – MATERIAL MODEL TESTING

The different flow stress models will be compared in this sub-section. The temperature history at TC3-4 and TC8-9 are shown in Figure 5-8. In general, the different models produce the same distinctive phases as found experimentally. In cases 1a, 1c, and 1d, we can see that the slope of the curve for TC3-4 is less than that of the experimental results. On the other hand, the rather stiff Johnson-Cook flow stress model (case 1b) follows almost exactly the experimental results. Clearly, the combination of a stiffer thermal softening model along with the JC strain hardening approach leads to an increase in the heat generation rate. This can be explained by an increase in friction heat in the stiffer model. This leads to a greater relative velocity between the tool and work-pieces. All the models other than 1c provide an over-estimate of the peak temperature at TC3-4 (located ¾ of the way along the weld line). Interestingly enough, although case 1b produces the highest temperatures, the model predicts the least volume of the internal defects as shown in Figure 5-12. The amount of plastic deformation in the model is greatest in case 1a since perfect plasticity is used.

A comparison of the temperature at the end of the dwell phase (at TC9) between experiment and simulation is shown in Table 5-3. The percent error is below 7% for cases 1a, 1c, and 1d. A similar comparison of the maximum temperature during the advancing phase (TC4) is presented in Table 5-4. Here the only case that provides a reasonable error is 1c (at 4%); all the other cases result in errors of 16-30%. This result further supports (and validates) the development of the FKS flow stress model and the cumulative damage friction law.
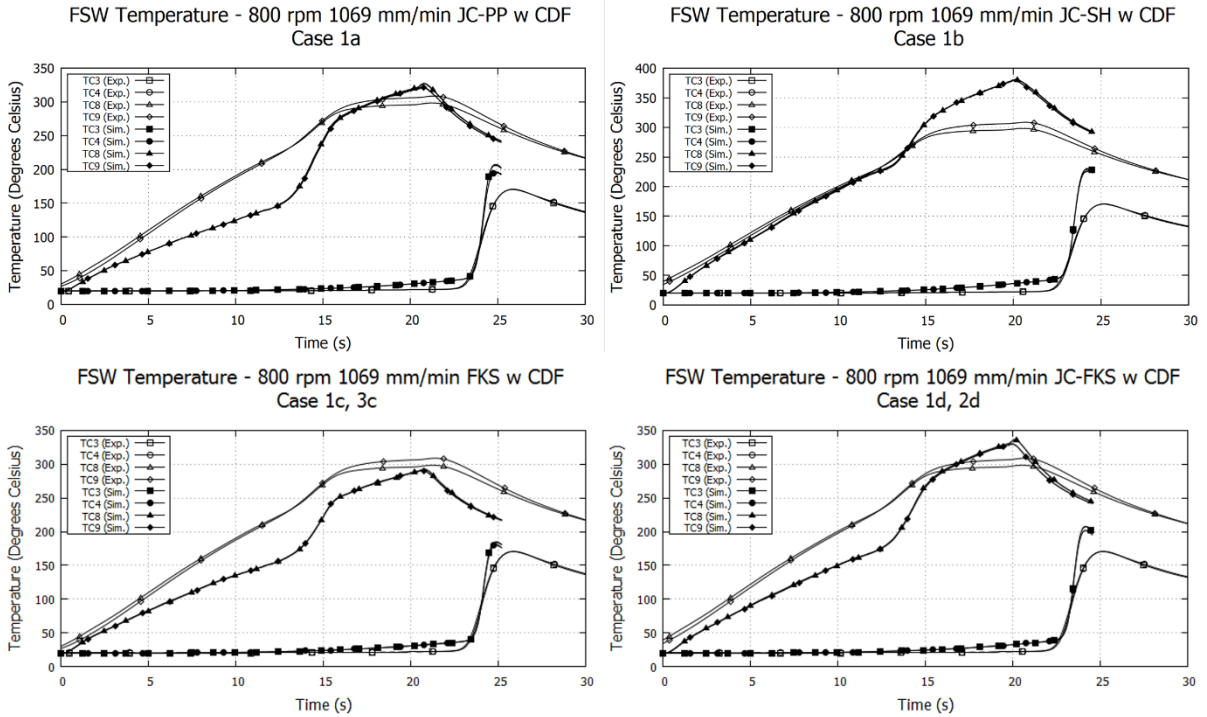
Figure 5-8 – Temperature history for case 1a, 1b, 1c, and 1d

Table 5-3 – Temperature comparison at end of dwell phase (TC9)

Case	Experiment [°C]	Simulation [°C]	% Error
Case 1a	309	322	4.2
Case 1b	309	380	23.0
Case 1c	309	291	5.8
Case 1d	309	330	6.8

Table 5-4 – Maximum temperature comparison during advance phase (TC4)

Case	Experiment [°C]	Simulation [°C]	% Error
Case 1a	171	207	16.9
Case 1b	171	230	30.0
Case 1c	171	184	4.0
Case 1d	171	208	17.5

The newly proposed flow stress model provides an excellent prediction of the internal defects as shown in the comparison of Figure 5-9. Here, the defects have been extracted using the free surface triangulation algorithm (see section 4.6) from the particles in the defect zones. Note the intermittent nature of the defects in both the simulation and experimental results. The weld shows a sizeable defect at the plunge/dwell site, and then the defects diminish in size and number until ~¼ of the way along the weld line. From ¼ to the end of the weld, the defects continue to increase in size. This is explained by the high temperature attained during plunge and dwell. Once the tool starts to advance, the peak weld temperature falls significantly (for the 1069 mm/min case). The average height of the defects in the simulation model is ~4.5 mm, compared to ~5 mm in the actual FSW joint.
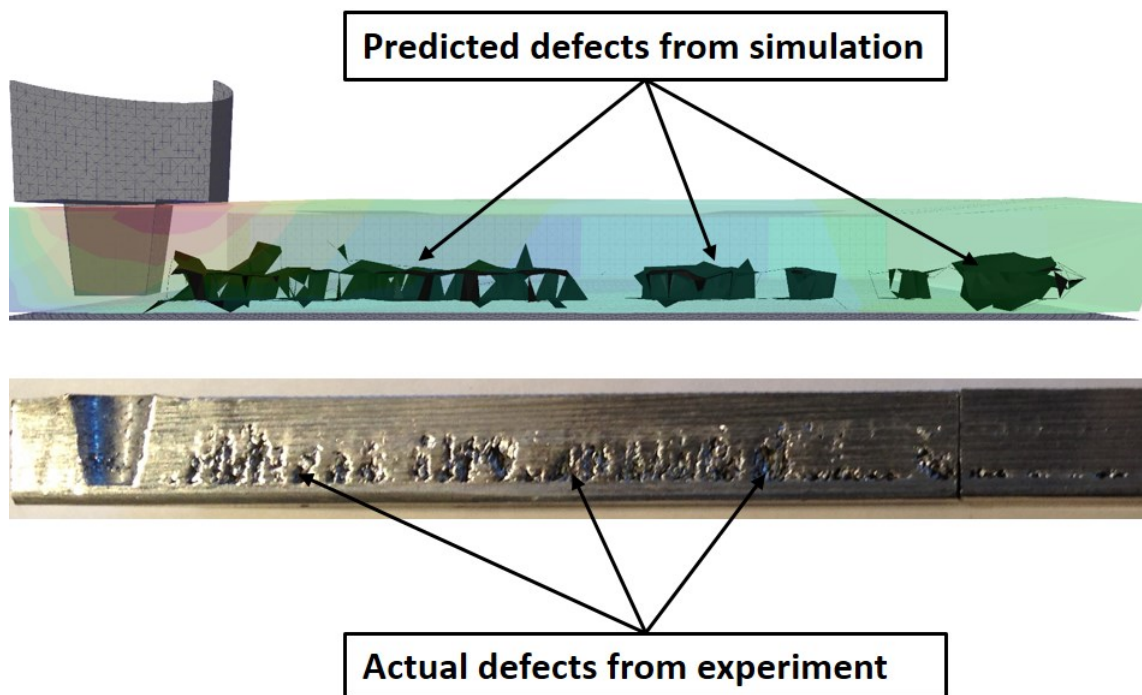
Figure 5-9 – Comparison of predicted defects (advancing side of case 1c and 3c) to experiment

A comparison of the spindle torque and forging force is shown in Figure 5-10. The highest torque and force is predicted in case 1b, this can be attributed to the increased stiffness of the Johnson-Cook model. No experimental results are available for the force and torque since the laboratory setup does not have a means to measure these quantities.
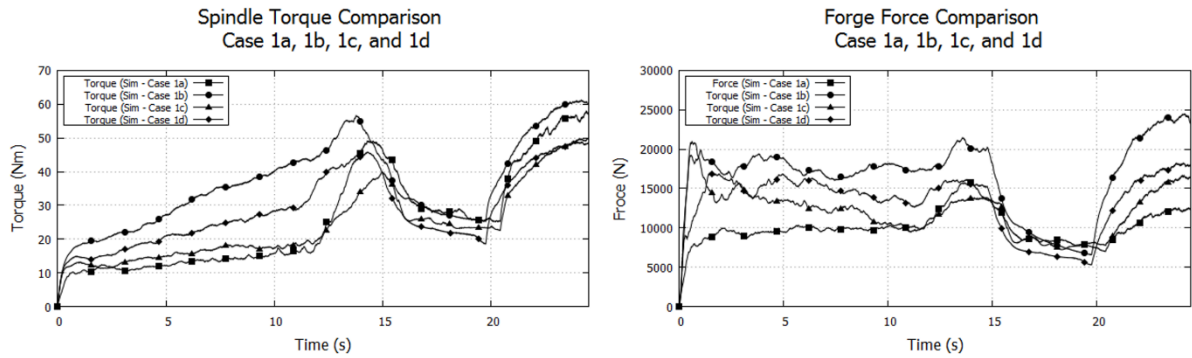
Figure 5-10 – Torque and forge force comparison for case 1a, 1b, 1c, and 1d
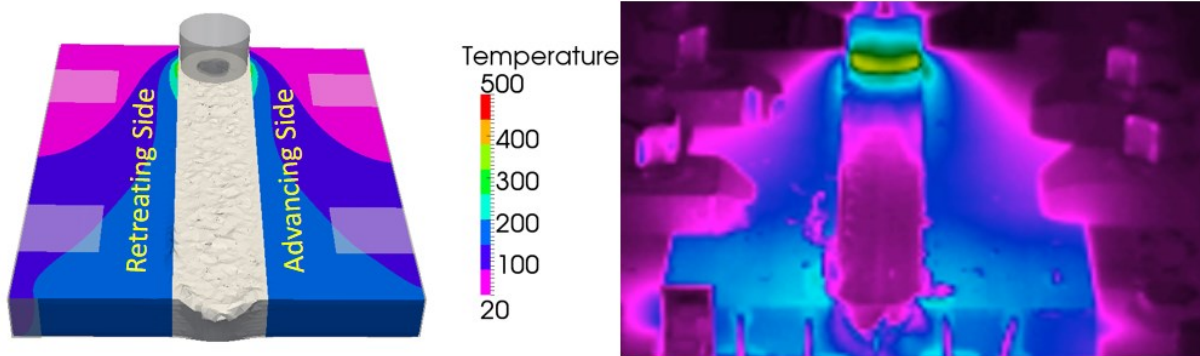
Figure 5-11 – Thermal camera image at end of weld for 800 rpm 1069 mm/min – Left: Simulation case 1c; Right: Experiment with Thermal camera. Temperature in [°C], same scale both images.

The true interest of the newly developed material model is that the temperature results correlate well with the experimental results, and at the same time, an excellent correlation of the defect size and location in the weld zone can be obtained. The other material models tend to over or under predict the defects. Although, none of the material models would leave any thought that using 800 rpm and 1069 mm/min would be acceptable. The JC model with strain hardening (case 1b) would suggest that a value close to 1069 mm/min could be acceptable due to the relative lack of defects present. The experimental results show the presence of extensive defects and tend to suggest that a speed no greater than 700-800 mm/min should be used.

The temperature profiles shown in Figure 5-12 further strengthen the case for the FKS flow stress model. Cases 1a, 1b, and 1d all show an over prediction of the weld zone temperatures. Case 1c has temperature contour profiles that correlate very well with an image taken using a Thermal camera (Flir:

http://www.flir.ca). Notice how the contours flare out from the tool and quickly curve back toward the weld line in Figure 5-11. Note that the temperature is not displayed correctly in the zones that were not painted (Thermal camera not calibrated to that emissivity), specifically the weld zone, the tool, and the support structure.
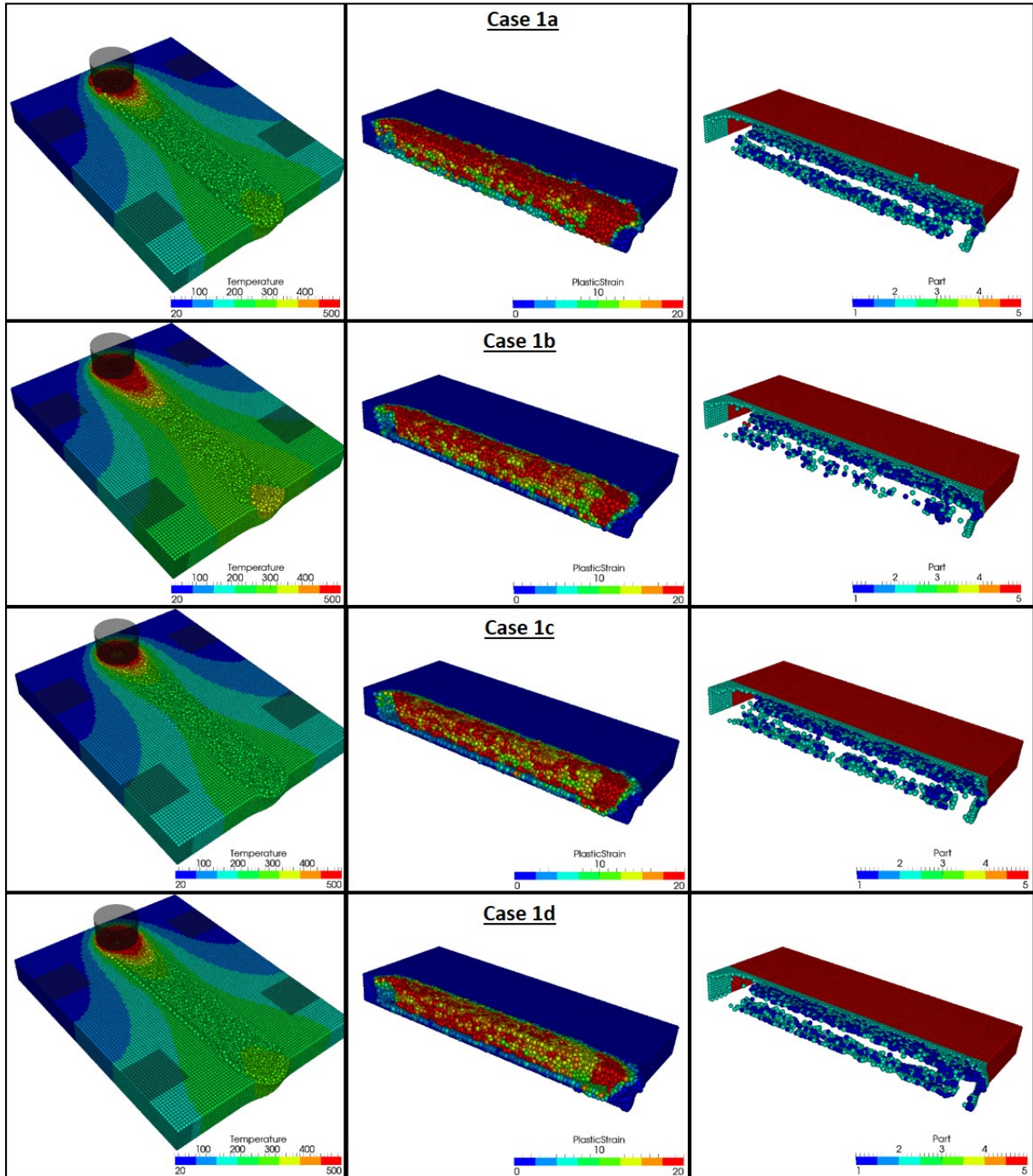
Figure 5-12 – Temperature (°C), plastic strain, and internal defects for cases 1a, 1b, 1c, and 1d

5.1.3   SIMULATION RESULTS - FRICTION MODEL TESTING

The temperature histories at TC3-4 and TC8-9 for cases 2a, 2b, 2c, and 2d are shown in Figure 5-13. Cases 2b, 2c, and 2d are almost identical. It will be assumed, that the heat generation is only moderately dependant on the friction model. Likely, the thermal mass of the tool and the base play a significant role. The slip model of case 2a (Coulomb) on the other hand does show a different trend. The heating rates for this case are distinctively different from those in the other cases. The stick model predicts the highest peak temperature; also, there is no perceivable difference between the temperatures on the advancing (TC4) or reversing side (TC3). Since the stick model will tend to transport material close to the contact surface, the heat will be transferred more uniformly, leading to similar advancing and retreating temperatures.
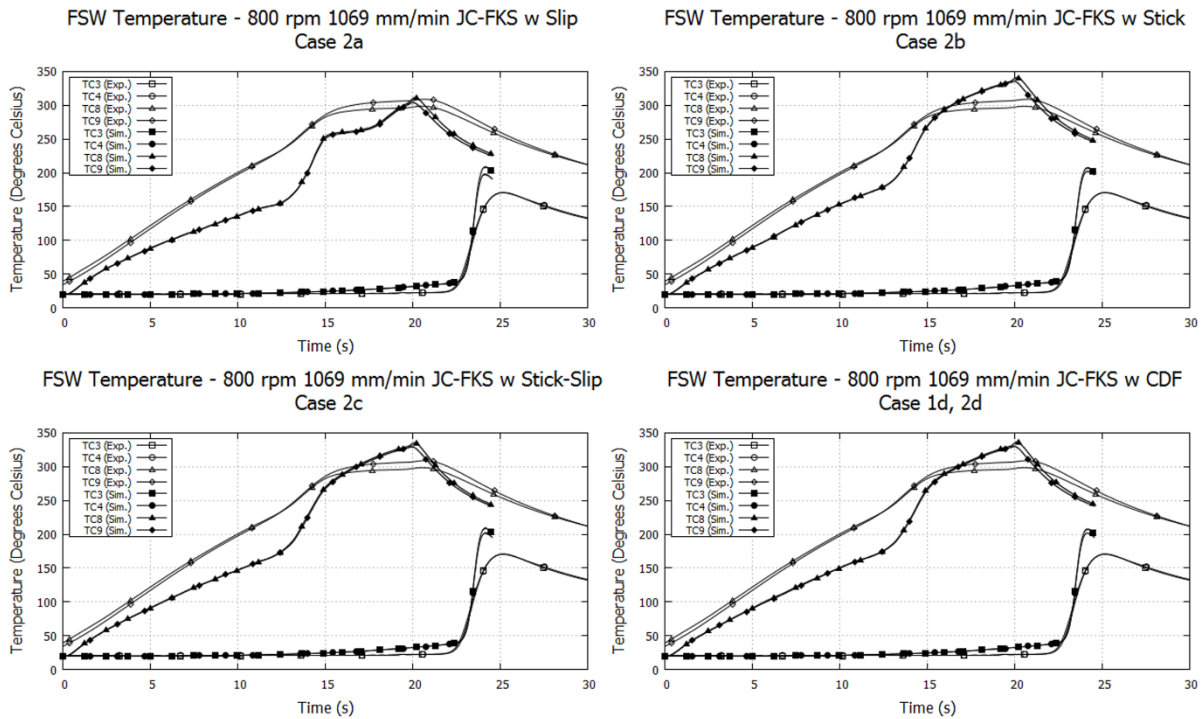
Figure 5-13 – Temperature history for case 2a, 2b, 2c, and 2d

The torque and force histories are shown in Figure 5-14. In a general sense, the peak torque attained during the dwell phase is uniform across the different models. During the advancing phase, the stick model predicts the highest forge force, whereas the slip (Coulomb) model predicts the lowest. Intuitively, the newly developed cumulative damage stick-slip model predicts torque and forces that are bound by

the stick and slip models. The pertinence of the individual friction models with respect to the torque and forces cannot be judged since no experimental data was obtained.
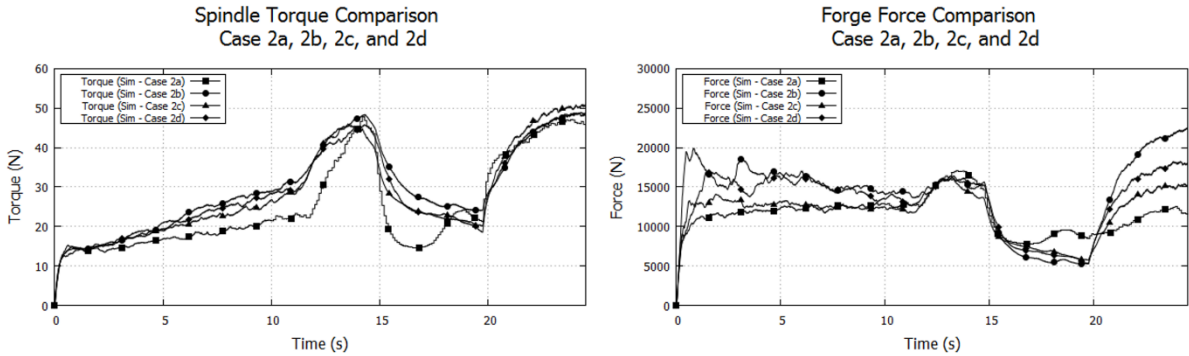
Figure 5-14 – Torque and forge force comparison for case 2a, 2b, 2c, and 2d

The form of the slip ratio (made to be a function of the cumulative damage) used in the new stick-slip model could be more suitably made a function of temperature and strain rate to take into account the likelihood that the aluminum material will be more adhesive at higher temperatures. More work would be necessary to formulate such a model. The improvement in the results may be worth the development.

5.1.4   SIMULATION RESULTS - PROCESS PARAMETER TESTING

The core reason for developing a sophisticated FSW simulation code is to be able to analyze the effect of different process parameters on the quality of the finished weld. In this sub-section, different advancing speeds will be investigated along with the subsequent weld quality. The effects of the different parameters on the tool and support will also be investigated.

The temperature history for the three cases is shown in Figure 5-15. The results are intuitive in the sense that the peak advancing phase temperatures increase as the ratio of rpm to speed of advance increases. Since each "material packet" will spend more time in contact with the FSW tool, there will be more heat generated, leading to increased peak temperatures. Again, excellent correlation is obtained between the numerical and the experimental results as shown in Table 5-5. The error in each of the three cases is below 4.3%, which is excellent. Here only the maximum advancing temperature is compared since the temperatures at the end of the dwell phase are the same in each model (same plunge and dwell parameters).
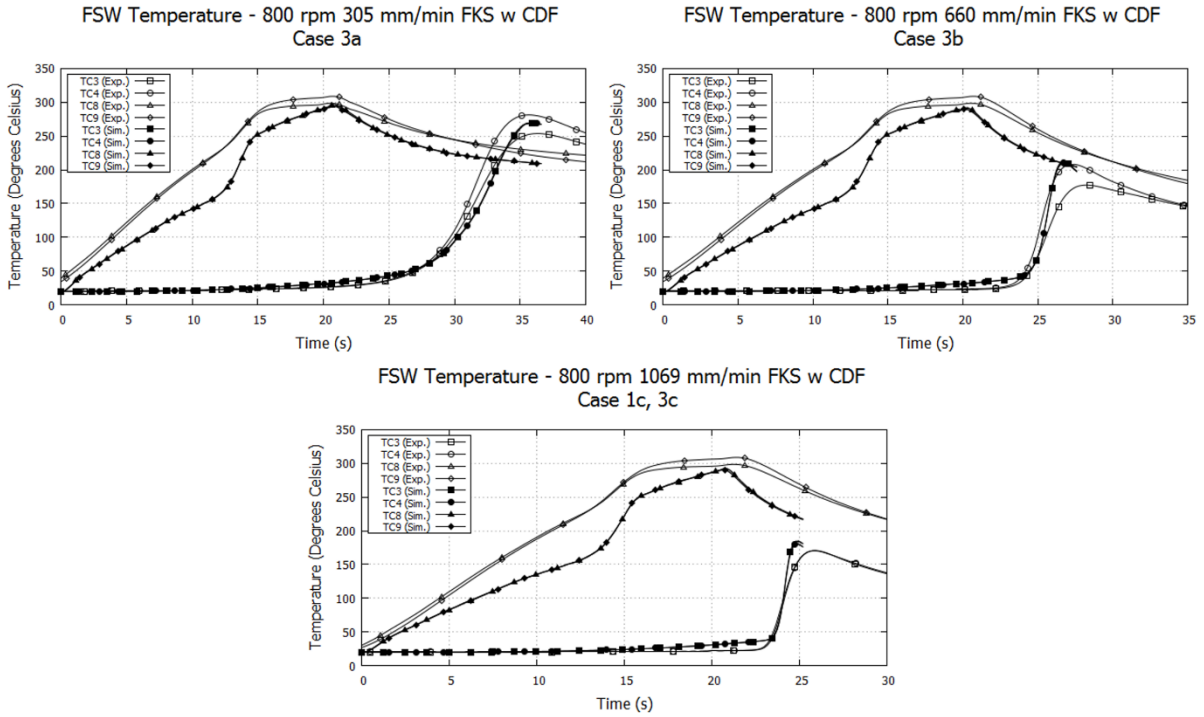
Figure 5-15 – Temperature history for case 3a, 3b, and 3c

Table 5-5 – Maximum temperature comparison during advance phase (TC4)

Case	Experiment [°C]	Simulation [°C]	% Error
Case 3a	281	269	4.3
Case 3b	208	210	1.0
Case 3c	171	184	4.0

A note on the relative precision of the thermocouple measuring approach; there are a number of sources of error and uncertainty. The first issue stems from the manner in which thermocouples are manufactured. The actual location of the measuring tip is unknown and can vary significantly. In addition, in some situations, the weld line was not perfectly straight, leading to a thermocouple on the advancing side reading a higher or lower temperature than on the retreating side. Another source of error can be associated to the thermal contact between the probe and the aluminum. A nickel-based grease was used at the interface between the two materials. There is no doubt that the grease improves the response

Kirk Fraser - April 2017

time of the thermocouples compared to air. However, there will still be a slight lag due to differences in the thermal properties.

The spindle torque and forge force is shown in Figure 5-16 for the three cases. For case 3c (fast advance, 1069 mm/min) the ratio of rpm to advancing speed is 0.75 which is simply unsuitable for a quality weld in 12.7 mm AA6061-T6 plates. The torque and force during the advancing phase surpass the plunge and dwell values. This is a telltale sign of a poor weld since the force and torque should decrease after the plunge/dwell phase, and then reach a steady state.
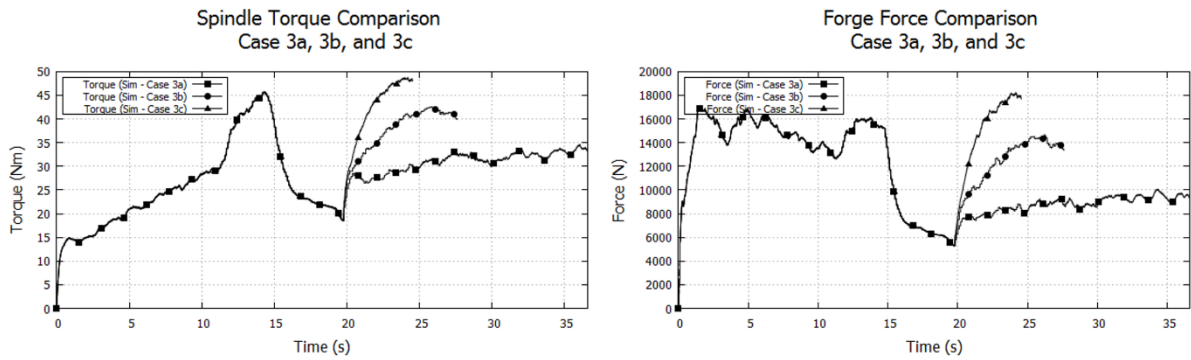
Figure 5-16 - Torque and forge force comparison for case 3a, 3b, and 3c

Certainly, performing the weld at 1069 mm/min will lead to detrimental effects on the welding equipment and the FSW tool. More importantly, this advancing rate leads to sizeable defects throughout the weld zone (as was shown in Figure 5-9). Both case 3a and 3b have more reasonable force and torque results and few to no defects in the weld zone.
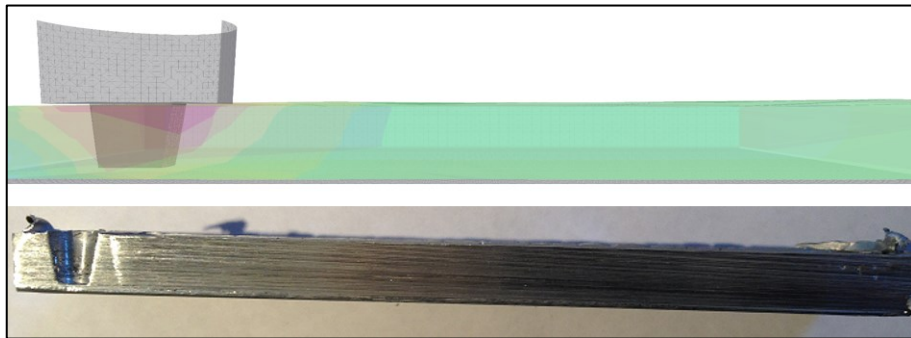
Figure 5-17 - Comparison of predicted defects (advancing side of case 3a) to experiment
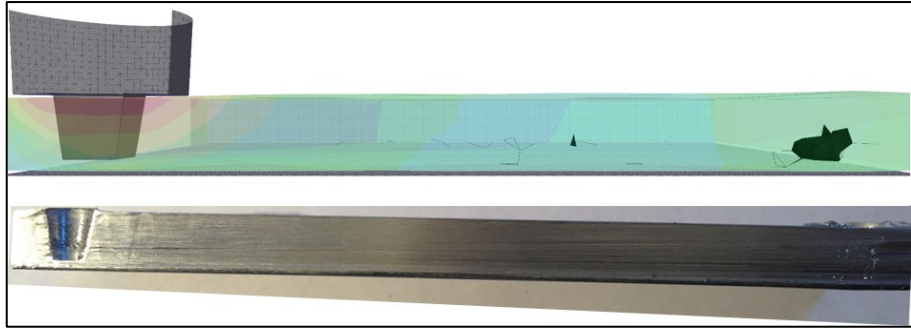
Figure 5-18 - Comparison of predicted defects (advancing side of case 3b) to experiment
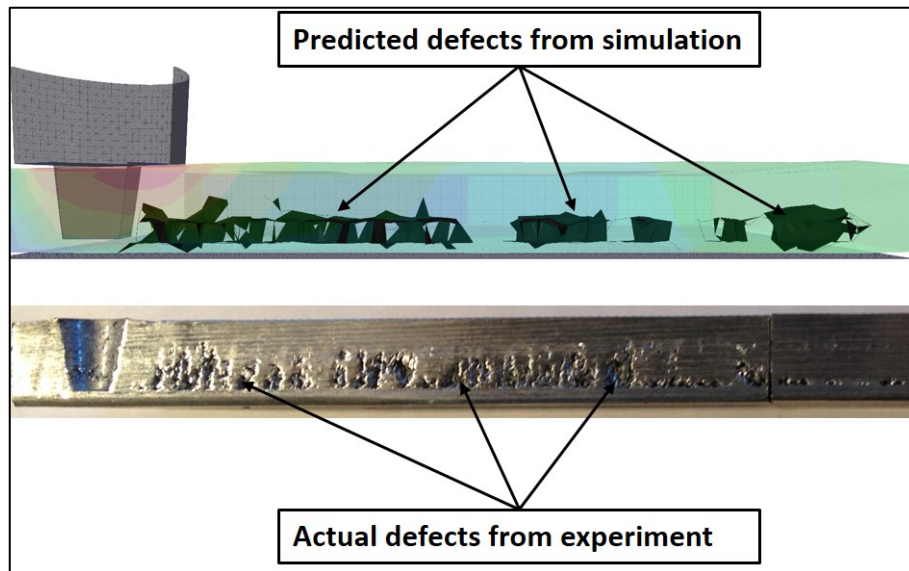
Figure 5-19 – Comparison of predicted defects (advancing side of case 1c and 3c) to experiment

The experimental welds showed small defects in the plunge/dwell region, followed by no defects in the advancing zone. Typically, the plunge/dwell region is discarded. For this reason, the defects are not considered significant. A comparison of the predicted defects for case 3a and 3b to experiment are shown in Figure 5-17 to Figure 5-19. No defects are predicted in case 3a; conversely, the experimental results show a slight defect at the start of the weld. For case 3b, some small defects are predicted at the start of the weld which is in good correlation with the experimental result. A Delaunay triangulation algorithm was used to join the particles making up the boundary of the internal defects.

The results of the tool wear analysis are shown in Figure 5-20. The predicted tool wear is highest for case 3a (19.16 µm) and lowest for case 3c (5.63 µm). The reason for this is that the relative velocity and time each "material packet" spends in contact is greatest in case 3a. In the future, we will investigate other formulations for predicting tool wear. Certainly, including the degradation of the tool hardness as a function of temperature will have an important effect on the results. It may be possible to include both the hardness of the tool and the aluminum in the formulation by using an appropriate averaging method. In this way, the wear rate will be more directly related to the hardness of the two materials at the contact interface.
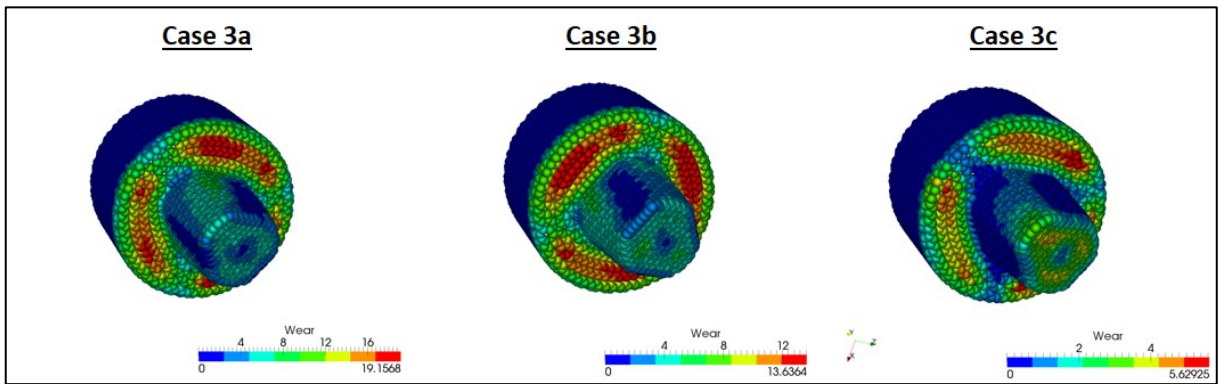
Figure 5-20 – Comparison of tool wear for cases 3a, 3b, and 3c

5.2 LAP JOINT - FRICTION STIR SPOT WELD

Lap joints are very economical and are used in many different applications. In many situations, they provide sufficient strength with far less welding time than a continuous weld. Two examples of parts joined by FSSW are shown in Figure 5-21. On the left side is a train roof that has spot welded ribs, and on the right side is a car door panel.
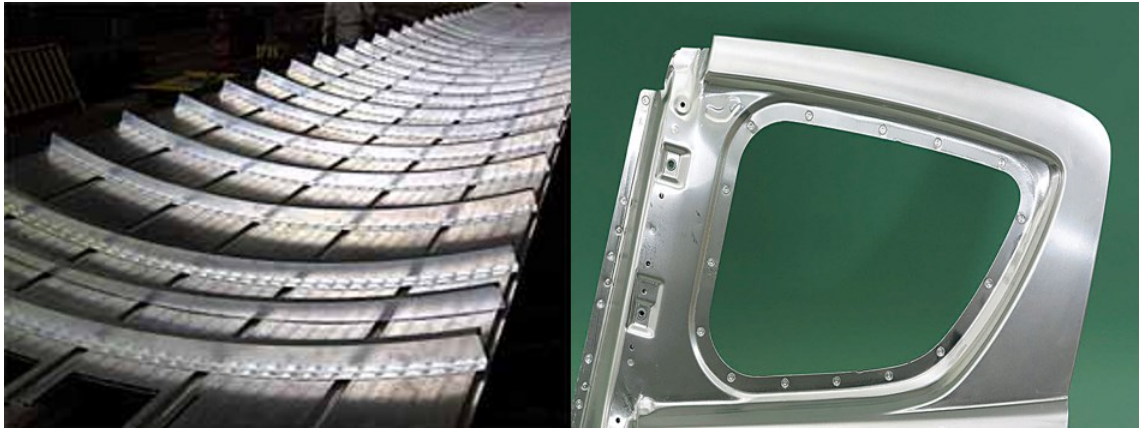
Figure 5-21 – FSSW examples; aluminum train roof [286] (left), aluminum car door panel [287] (right)

As discussed in section 2.3, the FSSW process will leave a hole when the tool is removed from the work pieces. This is acceptable in many cases where the weld will not be seen by the public. In this section, a FSSW simulation of a simple lap joint will be performed and compared to experimental tests. Three different cases will be investigated:

Case 1-   800 rpm, plunge at 100 mm/min, dwell 2.5 seconds

Case 2-   1000 rpm, plunge at 100 mm/min, dwell 2.5 seconds

Case 3-   1200 rpm, plunge at 100 mm/min, dwell 2.5 seconds

The experimental setup is shown in Figure 5-22. The surface of the upper plate in the lap joint has been painted a flat black to improve the image obtained with a thermal camera. Note that the area in the weld zone is not painted since this would adversely affect the friction between the tool and the work piece. Furthermore, if the weld region were to be painted, the paint particles would become mixed into the weld joint, causing internal flaws. Steel clamps are used to hold the plates firmly during the welding process. The lap joint is composed of a 102 mm x 102 mm x 6.4 mm lower plate and a 102 mm x 76 mm x 3.2 mm upper plate, both plates are AA6061-T6. The thickness of the plates was chosen to provide a good weld with the tool that was available (7.5 mm pin length). The Bottom plate is wider than the top plate because these widths of plates were available in the shop at that point in time. The tool used had a shoulder diameter of 21.5 mm, an upper and lower pin diameter of 9.0 mm and 5.0 mm. The shoulder is scrolled and the pin has threads with a pitch of 1.0 mm. The pin has a length of 7.5 mm, and the total

plunge depth was chosen to be 8.0 mm. With a plunge speed of 100 mm/min, the plunge duration is 4.8 seconds. Add to this a dwell time of 2.5 seconds makes the total time 7.3 seconds. Position control is used in all cases with constant tool velocity.
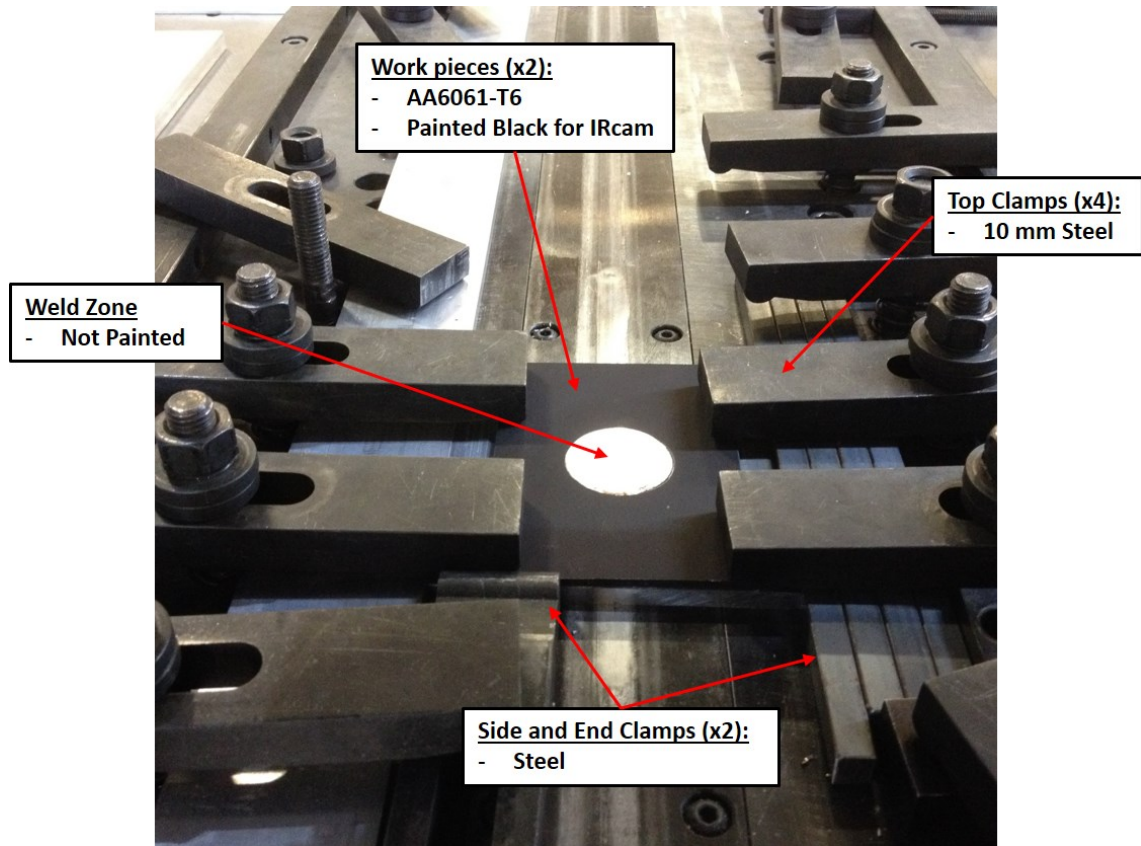
Figure 5-22 – FSSW experimental setup

This experiment was carried out in the technology transfer center of CEE-UQAC. The FSW machine, shown in Figure 5-23, is a gantry crane system that is capable of welding parts 18.0 meters long, 1.0 meters high, 3.5 m wide, and up to 15.0 mm full penetration in a single pass. The machine is state of art with the following features:

- Forge axis position and force control

- rpm control using a thermocouple embedded within the tool

- Spindle and tool liquid cooling for long, uniform welds

- Wide variety of tools included retractable pin tools, bobbin, convex and concave shoulder with and without threads

- Laser and camera lateral position control

- Deburring system (real time flash removal)

Figure 5-23 – CEE-UQAC FSW machine

5.2.1 MODEL DESCRIPTION

The simulation model, shown in Figure 5-24, is composed of elastic-plastic-thermal SPH elements for the upper (3.2 mm AA6061-T6) and lower (6.4 mm AA6061-T6) work pieces. The work pieces are modeled using the FKS flow stress model; material properties are provided in Table 3-2 and Table 5-1. The thermal properties are the same as defined in section 5.1.1. The FSW tool rotates counter clockwise in the various FSSW cases considered here. Contact between the tool and work pieces is incorporated following the presented models in section 3.11 and section 3.12. The damage parameters used are provided in Table 3-3. An initial uniform SPH element spacing of $\Delta s = 0.64$ mm is used for the work pieces. The average rigid finite element mesh size is 1.1 mm. The time step is controlled by the CFL condition, resulting in $CFL = 0.7$ and $dt_{min} = 9.0 x 10^{-8}$ s. $\tilde{V}_{SF} = 30$ (velocity scaling factor) has been used for this case study. Heat loss due to convection and radiation from the free surface of the work pieces is included according to sections 3.9.5 and 3.9.6. The convection coefficient is set to 20 W/m²K, and the emissivity is taken as 0.95. Heat transfer from the work pieces to the base is accounted for by increasing the convection coefficient by a factor of 30.0 for the SPH elements in contact with the finite elements. This is a reasonable approximation for the heat lost to the base since it is essentially a heat sink (very large thermal mass). On the other hand, the tool has a small mass, and as such, will steadily increase in temperature. This effect will be taken into consideration by using a variable tool temperature (function of simulation time, t) to approximate the heat loss (or gained once the tool is hot enough) between the work pieces and the tool. The relation used is:

$$T_{tool}(t) = \begin{cases} 3125.0t + 20.0\ ^\circ C, & 0 < t < 0.1536\ s \\ 500.0\ ^\circ C, & t \geq 0.1536\ s \end{cases} \qquad (5\text{-}5)$$

This relationship is adapted from the work of Fehrenbacher *et al.* [288]. In this manner, the increasing temperature in the tool will effectively cause heat to be drawn from the weld until the temperature in the work pieces surpasses the tool temperature. At that point, heat will be drawn from the tool in order to increase and stabilize the temperature in the weld zone. Two distinct phases, plunge, and dwell are simulated; the runtime for this model was 36 hours on a GTX 980 Ti. This model was not run on a CPU equivalent code due to an estimated run time of over a month (based on experience with other models).
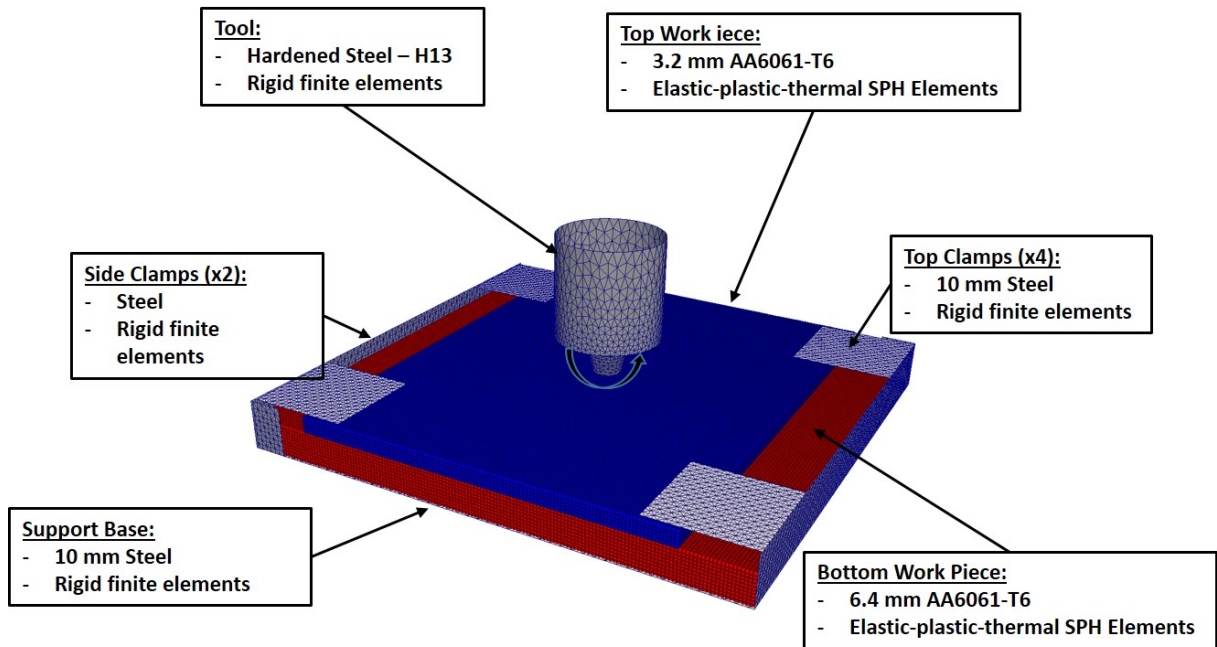
Figure 5-24 – FSSW simulation model

The finished welds, showing the flash formations, are shown in Figure 5-25 for the three cases. From the images, the weld with the most flash appears to be case 1. This is not, however, representative of the actual amount of material that was extruded past the edge of the tool. The size of the flash cannot be inferred from the finished welds because some was entrained around the rotating tool and ripped away from the plate. In actuality, the flash size should be greater with increased rpm. Temperature images were obtained using a FLIR Thermal camera. The images obtained allow a simple comparison

of the surface temperature between experiment and simulation. Temperature histories were not obtained for the FSSW test cases.

Figure 5-25 – Finished welds: 800 rpm (left), 1000 rpm (center), 1200 rpm (right)

5.2.2   SIMULATION RESULTS

The simulation cases were modeled and run in SPHriction-3D, the various results will be presented here. A comparison of the temperatures for the three cases is shown in Figure 5-26. The free surface triangulation algorithm (see section 4.6) is used in the images. The first row of images is from part way through the plunge phase when only the pin is engaged in the plates. The second row is at the point that the shoulder has made full contact. The last row corresponds to the end of the dwell phase. The maximum temperature has been capped at 500 °C for each of the cases to provide a means to compare. The differences in temperature profiles are subtle; close inspection shows that the temperature contours extend out further in case 3 than case 2, and case 2 extents further than case 1. This result is intuitive since the increased rpm is expected to cause an increase in temperature. Without exception, the hottest region in the model is directly under the tool shoulder. Since the temperature history was not obtained experimentally, these results cannot be compared.
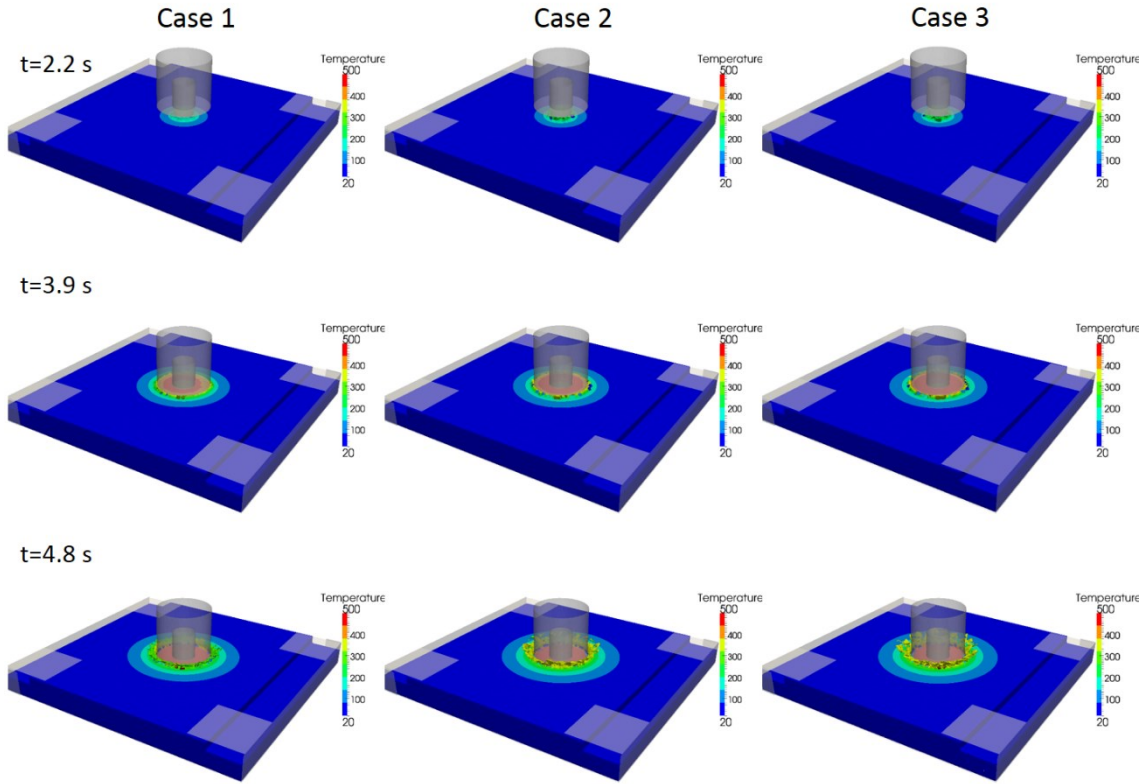
Figure 5-26 – FSSW temperature comparison for the three cases

Force and torque measurements were obtained since the FSW machine is equipped with the required sensors. Comparisons between the experimental and simulation results for the forge force (left) and spindle torque (right) are compared in Figure 5-27.
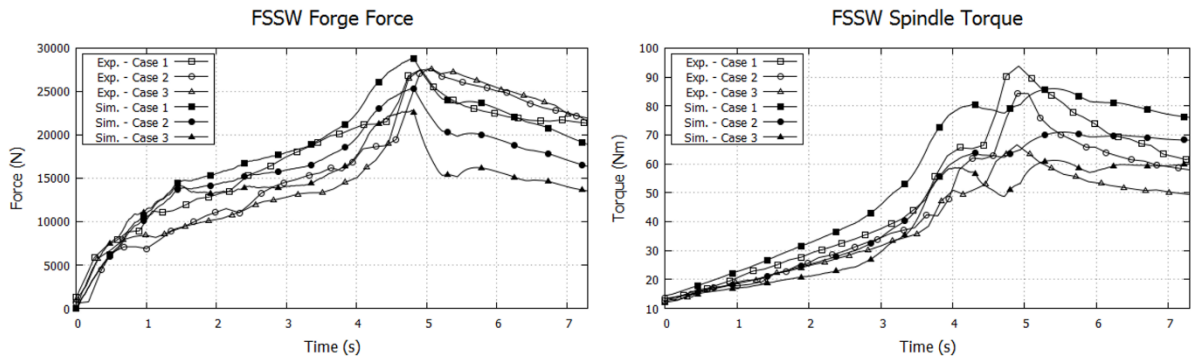
Figure 5-27 – Forge force (left) and spindle torque (right) comparison

A comparison of the plastic strain through the thickness of the welded plates is shown in Figure 5-28. The cross section is taken through the center of the tool. The image also visually shows the increased

amount of flash created in the 1200 rpm case. Due to the increased rpm, the weld is hotter, which will increase the amount of thermal expansion as well as cause the aluminum material to flow more. These reasons combine together to create more flash.
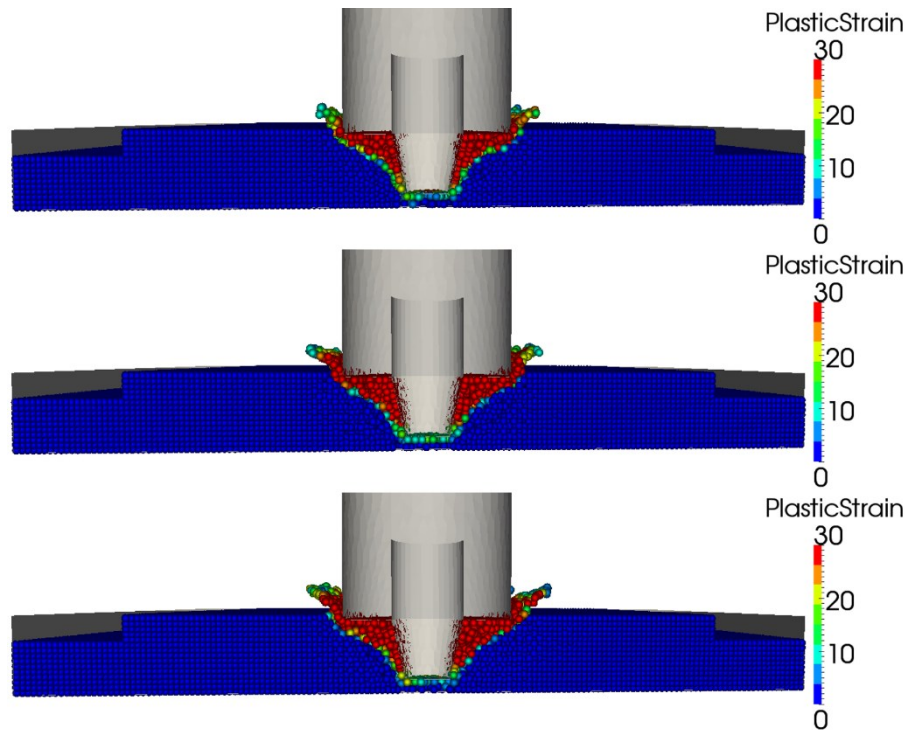
Figure 5-28 – Flash and plastic strain comparison – Top: Case 1 - 800 rpm; Middle: Case 2 - 1000 rpm; Bottom: Case 3 - 1200 rpm

5.3 COMPLEX JOINT GEOMETRY

To date, most of the work on simulating the FSW process has been focused on a simple butt-joint geometry model. Such a model is sufficient for academic research. However, for real engineering applications, the numerical model should be robust enough to be able to simulate complex geometries within a reasonable timeframe. In this section, the FSW simulation model and results for a complex geometry will be described. The case considered is of an aluminium alloy bridge-deck that is fabricated by extrusion in multiple sections and joined using FSW. The joint geometry can be seen in Figure 5-29. One of the drawbacks of using extruded sections is that the parts tend to fit together with some undesirable qualities for FSW. In this case, the two work-pieces join a ~0.5 mm step at the top surface of the joint (as shown in Figure 5-29). The left most work-piece is slightly thicker than the other piece,

and as such, poses a challenge for FSW. The tool will have to push down an extra 0.5 mm in order to come into contact with the lower of the two surfaces. This in turn causes the formation of significant flash on the thicker work-piece. The overall height of the joint is 100 mm, the three vertical members are 3 mm thick, the thicker plate (left side of step in image) is 3.7 mm thick, and the thinner plate is 3.2 mm thick.
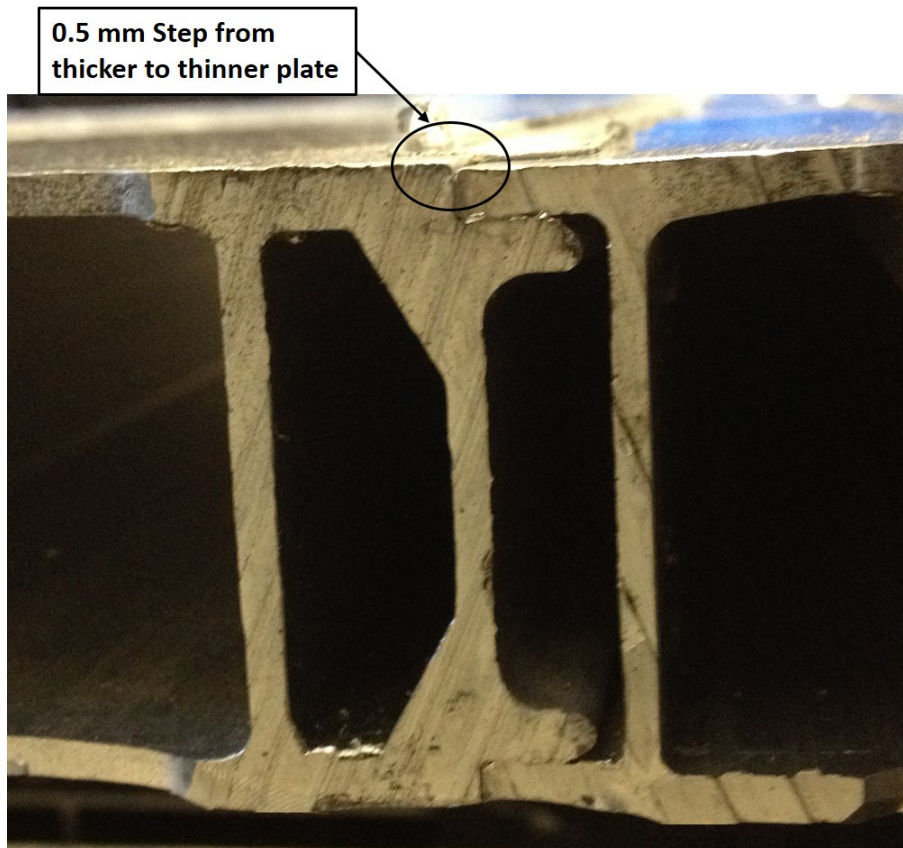
Figure 5-29 – Complex joint

5.3.1   MODEL DESCRIPTION

The complex joint geometry is modelled by a combination of SPH elements for the work-pieces and rigid finite elements for the tool. Since the tool is made of hardened steel, it can safely be approximated as a rigid body. The simulation model is shown Figure 5-30; the rigid tool and the two work-pieces including the step at the top surface are shown. The mesh size for the finite elements is 0.6 mm in the pin and shoulder region. Large elements are used outside of this region since contact with the work-pieces is only during flash formation.

The entire joint geometry is modelled with elastic-plastic-thermal SPH elements to allow for an improved prediction of the thermal expansion and the stresses in the joint during the welding process. The vertical member below the weld seam carries 90% of the forge force during the welding process. With this modelling approach, the stresses and the possibility of web collapse can be evaluated. The tool interacts with the work-pieces through a penalty based contact algorithm as described in sections 3.11 to 3.14. The tool has a shoulder diameter of 15 mm, an average pin diameter of 6 mm, and a pin depth of 3.8 mm.
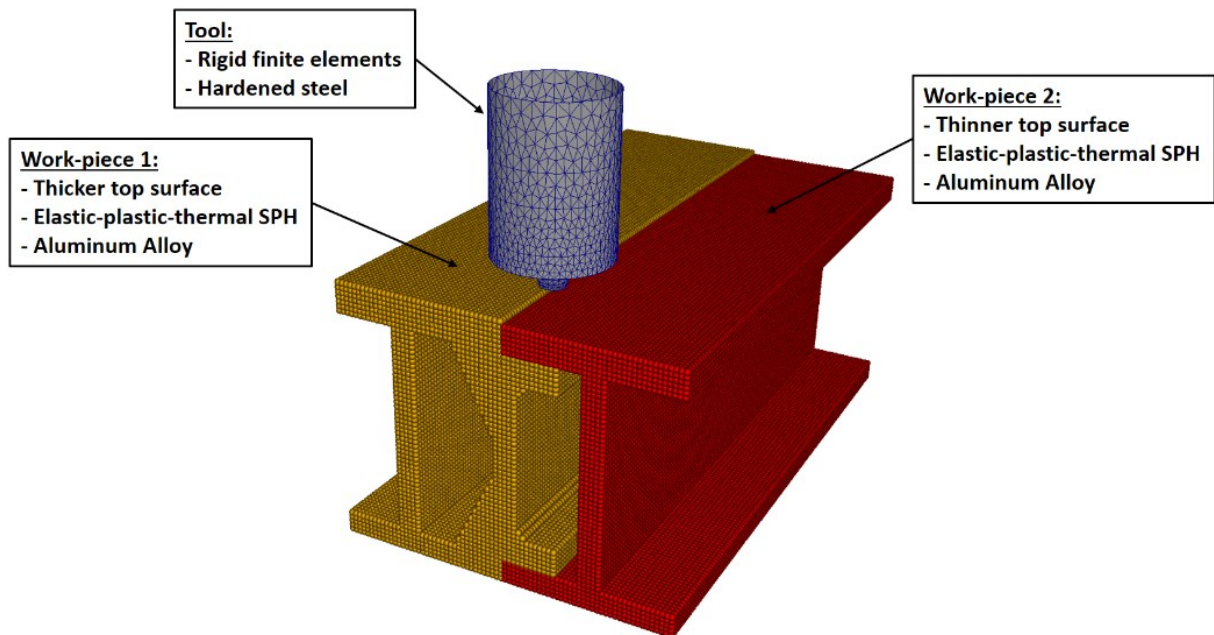
Figure 5-30 – FSW joint simulation model

The simulation model is composed of only a small region of interest of the actual bridge deck. Convection (10 W/m²K) is included in the model as well as radiation (the top surfaces of the work-pieces were painted black, emissivity of 0.95). The painted surface creates a less diffuse surface, making the thermal camera more precise. The adaptive thermal boundary condition algorithm as described in section 3.9 was used. The material parameters of the aluminium alloy used in the simulation are shown in Table 5-6 below. A uniform grid particle distribution of 0.6 mm to discretize the work-pieces has been used. This spacing allows for a sufficient number of particles through the thickness without incurring an excessive calculation penalty. The time step size is selected based on the Courant-Friedrichs-Lewy (CFL) criteria, $dt_{min} =$

Kirk Fraser - April 2017

$CFL[h/(v_{max} + c)]$. For this FSW model, $CFL = 0.7$, $dt_{min} = 9.8x10^{-8}$ s, and $\tilde{V}_{SF} = 30$. The small time step size is one of the major drawbacks of using a solid mechanics approach. Nevertheless, the time step size is required in order to capture the propagation of elastic stress waves within the aluminium.

Table 5-6 – Thermophysical properties of AA6005-T6

Mechanical			Work-Pieces Thermal		
Parameter	Value	Units	Parameter	Value	Units
Density, ρ	2700.0	Kg/m³	Conductivity, k	175.0	W/mK
Initial yield, σ_{y_0}	240.0	MPa	Heat capacity, C_p	895.0	J/kgK
Shear modulus, G	26.3	GPa	Tool Thermal		
Room temperature, T_R	20.0	°C	Conductivity, k	55.0	W/mK
Melt temperature, T_{melt}	605.0	°C	Heat capacity, C_p	485.0	J/kgK
Softening exponent, m	1.34	-	Density, ρ	7850.0	Kg/m³
Speed of sound, c	4722	m/s			

The model is run as two distinct phases: plunge and advance. The dwell phase was not part of the process as a ramp-up procedure to full advance speed was used in the experiment. A well-defined ramp-up is good practice to limit the forces and torque on the tool and can replace the dwell phase. The plunge speed is 25 mm/min and the full advance speed is 1250 mm/min with 2100 rpm. The ramp-up is performed linearly for an initial tool displacement of 40 mm; after this point, the tool speed is constant at 1250 mm/min.

Because of the 0.5 mm step, excessive amounts of flash are produced as the tool advances. The flash has to be removed following the welding phase and requires a significant amount of work for the welding technician. In order to attempt to reduce the quantity of flash produced, three cases are investigated. Case 1 represents the actual process parameters used in the experiment. This case is used to validate the tool force and torque, as well as the temperature distribution and history. The tool plunges full depth (4.3 mm) until the tool shoulder contacts the top surface of both work pieces. Case 2 and 3 are variations

on case 1. In case 2, an attempt is made to reduce the quantity of flash by reducing the plunge depth (4.2 mm as opposed to 4.3 mm). This will have the effect of limiting the volume of material that is sheared off the top surface of the thicker plate. In case 3, the flash formation will be reduced by operating the FSW tool with a clockwise rotation. This results in the advancing side being on the surface of the thicker plate. This will increase the weld temperature and help to move more material to the lower side of the step, ultimately creating a superior weld compared to case 1 and 2.

5.3.2 SIMULATION RESULTS

The three cases were run in SPHriction-3D; in this section, the results from the three different cases will be presented. The production process parameters correspond to case 1 and are used to validate the model. A video of the results for the three cases is available here: https://www.youtube.com/watch?v=eLOQILkUx-A.

The temperature distribution results for the three cases are shown in Figure 5-31 at different times during the simulation. The maximum temperature for case 2 is lower than for the other two cases. This is because the tool plunges 0.1 mm less; in turn decreasing the forge force and the heat generated. The ultimate result is that the quality of the weld in case 2 is significantly lower than in the other cases. Of the three cases, the best weld quality is obtained from case 3. Since the tool rotates clockwise, the advancing side is on the surface of the thicker work-piece. This helps to move the hot material to the thinner work-piece at the front of the tool. This is a favourable situation compared to having the hot material move around the back of the tool (as in case 1 and 2). This causes the work-pieces to heat up more uniformly than is possible in either case 1 or 2.

Four measurement points (TCs) for the temperature distribution have been used in the three simulation cases as shown in Figure 5-32. TC1 and TC2 are placed at the middle of the work-piece (along the weld direction). TC3 and TC4 are placed in line with the tool axis during the plunge phase. The four TCs are at the surface of the work-pieces and located 11.5 mm from the interface of the two work-pieces. MTC1 is a moving temperature measurement point that is located on the underside of the tool and follows the tool as it rotates and advances. MTC1 is located 6 mm from the tool axis on the underside of the tool shoulder.

The temperature was measured experimentally at two points on the surface of the thicker work-piece (at location TC1 and TC3) for Case 1 only using data obtained from a thermal camera. Due to the filming angle available with the thermal camera (restricted access to work area); temperatures on the thinner work-piece could not be evaluated (TC2 and TC4). Figure 5-33 shows that there is a good agreement between the experimental and simulation results at TC1 and TC3. The simulation results for MTC1 (moving thermocouple) cannot be compared to experiment since the tool was not equipped with a thermocouple. The simulation model has a tendency to over predict slightly the temperature.

The stress-strain curves for AA6005-T6 were not available; as such, engineering judgement was used to scale the flow stress curves found for AA6061-T6. Certainly, the uncertainty of the true nature of the flow stress for the AA6005-T6 alloy leads to error in the predicted temperatures. Furthermore, the heat capacity and thermal conductivity of the aluminium alloy at high temperature are not well known. These parameters play an important role in the coupled thermal-mechanical model. The relative difference between the TC's on the thicker and thinner plates gives a good means of diagnosing the quality of the weld. If there is a large difference in the temperature readings, one can conclude that the pressure is higher on one side of the weld then the other. This leads to an un-favourable temperature distribution and the weld quality suffers. Case 2 is an excellent example of such a situation. Notice the large difference in temperature in TC3 and TC4. Since the plunge depth was insufficient, there is not enough pressure on the thinner plate, leading to a decrease in temperature.
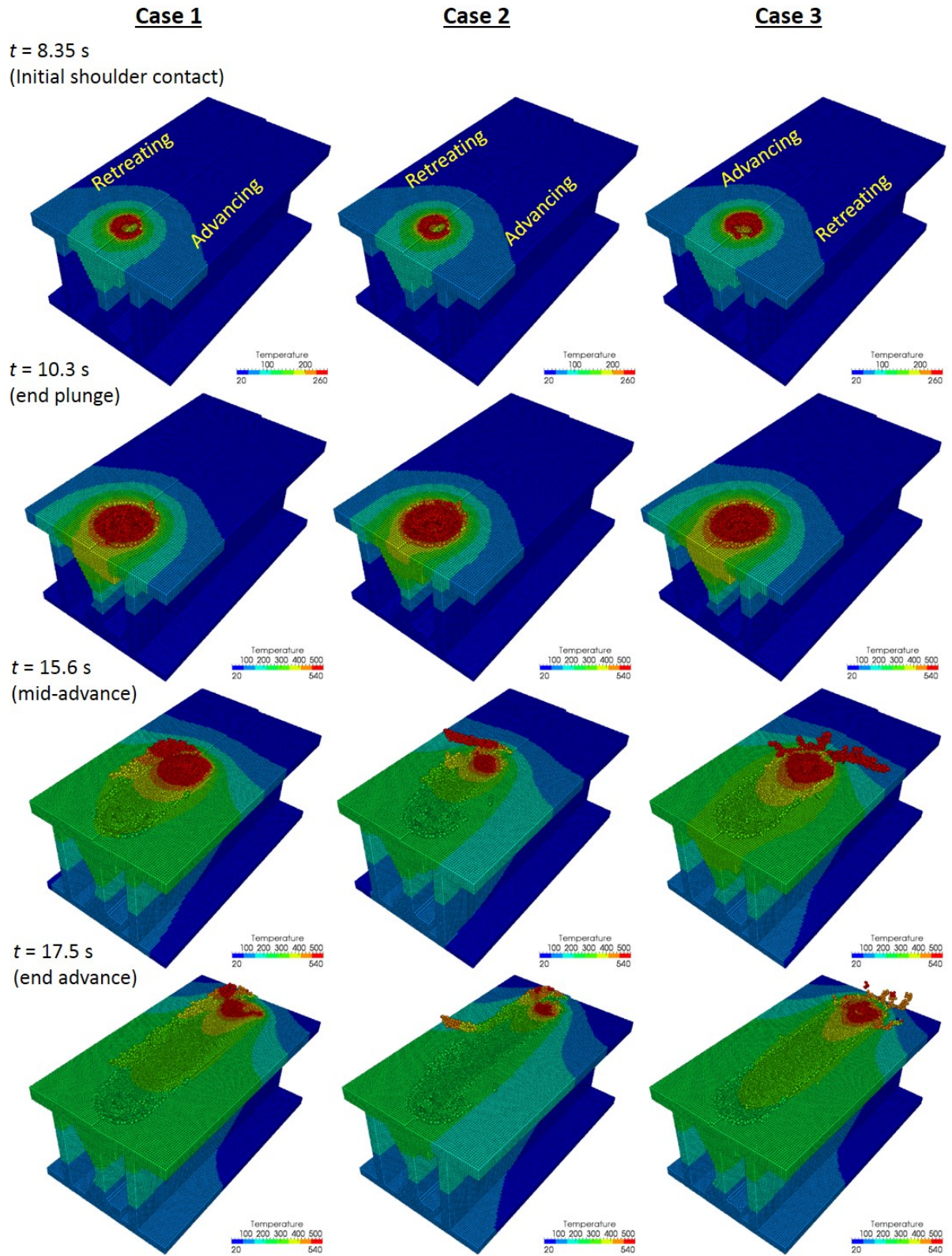
Figure 5-31 - Temperature and deformation results for the three cases
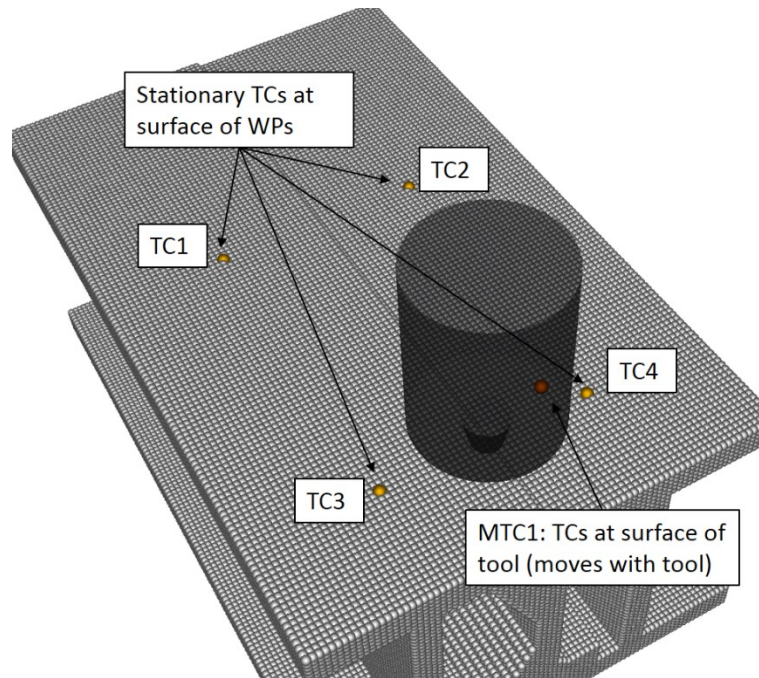
Figure 5-32 - Temperature measurement points in the simulation model

The temperature results at MTC1 are also an excellent indication of the weld quality. Since MTC1 follows the tool as it rotates and advances, large temperature fluctuations are suggestive of inadequate process parameters. The variation in temperature at MTC1 for case 3 is significantly less than in other two cases. The experimental setup that was used did not allow for embedding thermocouples in the work-piece or in the tool. Using a thermal camera is beneficial in cases like this since holes do not need to be drilled in the aluminium or the tool. The surfaces to be filled should be painted a light coat of flat black paint that can easily be removed with light buffing following welding. Temperature measurements with a thermal camera provide a very powerful diagnosis tool in the lab or in the hands of a FSW technician at a commercial company.
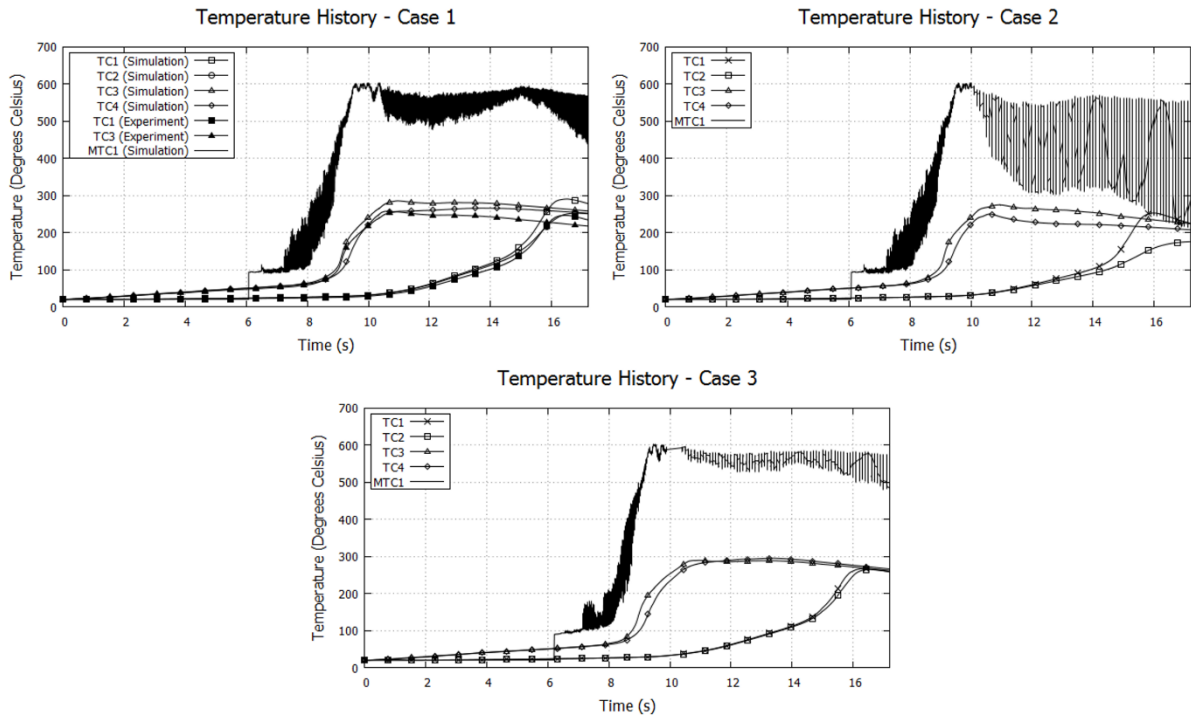
Figure 5-33 - Temperature history results for the three cases

The images obtained can help the technician or engineer to understand if their chosen process parameters are adequate and if not, give good hints as to why. For example, if the thermal camera shows a significantly higher surface temperature on the advancing side than the retreating side, the tool is likely advancing too fast for the chosen rpm. During the plunge phase, the thermal camera can again be used to determine if the plunge speed is too high (surface temperature too low) or low (surface temperature too high).

Of particular interest is the strong oscillation at MTC1 for case 2. Near the end of the simulation, there is a peak-to-peak temperature change of over 300 °C. The temperature on the thinner plate is too low to allow the aluminium material to flow and the weld is essentially incomplete. This can be verified by investigating the plastic strain contours in the weld zone as shown in Figure 5-36.
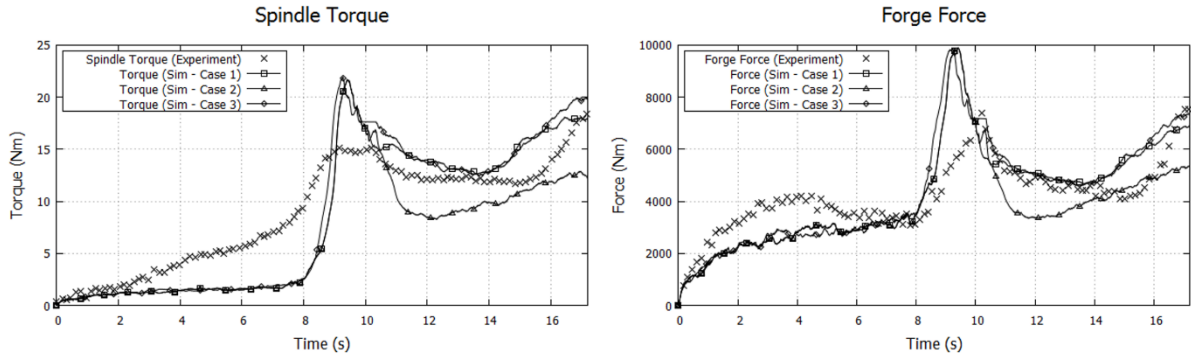
Figure 5-34 - Spindle torque and forge force comparison

Case 3 is the only one of the three in which the mechanically affected zone spans the entire diameter of the tool. In case 1, the welded zone gets narrower as the tool advances. For case 2, the welded zone spans no more than half the tool diameter from the edge of the tool pin on the thinner plate into the thicker plate.
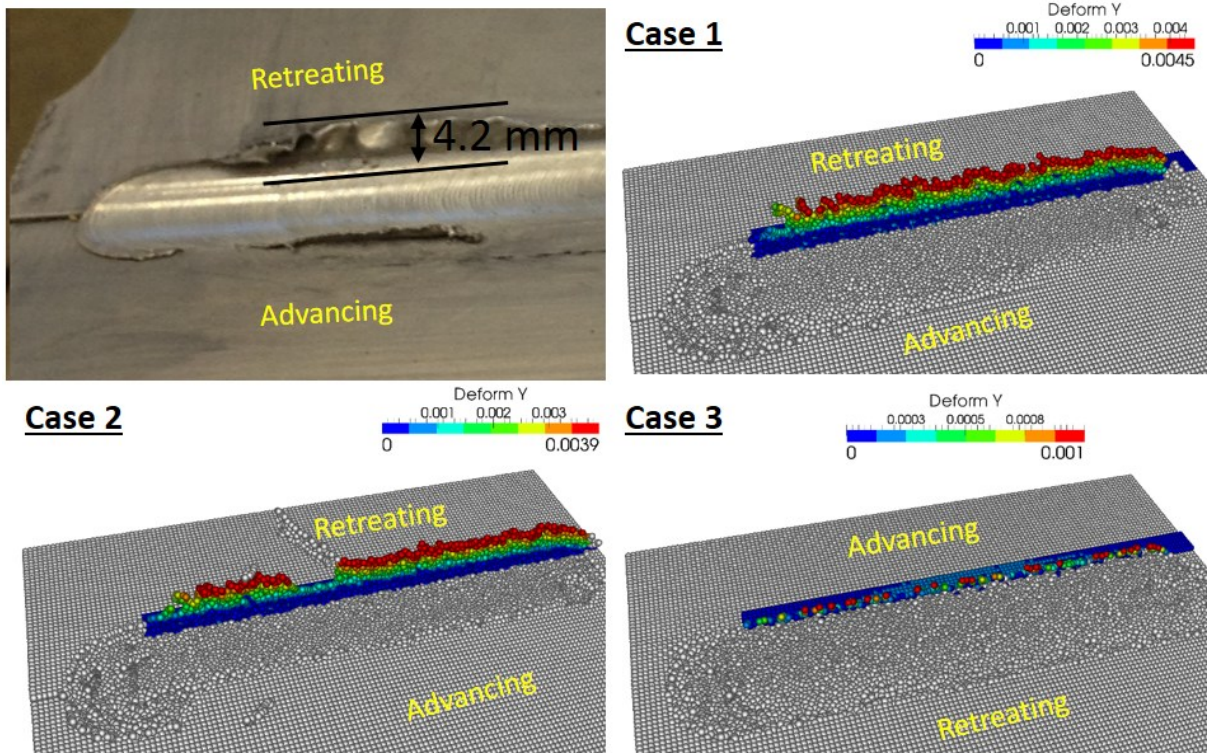
Figure 5-35 - Flash height comparison at end of advancing phase

A comparison of the spindle torque and the forge force is shown in Figure 5-34. The inertia of the spindle plays a strong role in the experimentally measured torque. Because the plates being welded are very thin, the max process torque does not exceed 25 Nm and the average torque during the advancing phase is ~20 Nm. However, the no-load torque measured was ~10 Nm, accounting for almost half of the typical process torque. In the simulation models, the inertial effects of the spindle are not taken into consideration because the tool is modeled only as a rigid surface to provide contact between the tool and the work pieces. Estimating the inertia of the tool and spindle is complicated since the machine drawings were not available. The simulation torque is calculated by taking the cross product of the contact forces and the distance vector between tool axis and an SPH element.

A good correlation between the forge force from experiment and simulation was obtained. The inertial effects do not play an important role here, leading to a better prediction than was obtained with the torque. Other factors lead to a reduction in the precision of the predicted torque and forge force. These effects are; the thermo-physical properties of the material, the chosen friction law, differences in how the FSW machine and simulation model control the position, and rpm of the tool, as well as discrepancies between the actual geometry of the work-pieces and the tool compared to their idealization in the simulation model.

Nevertheless, the simulation model provides an excellent understanding of how a change in process parameters affects the torque and forces. We can see that the tool torque and forge force for case 2 is lower than that of case 1 and 3. This is an intuitive result as the plunge depth is shallower, leading to less contact pressure.
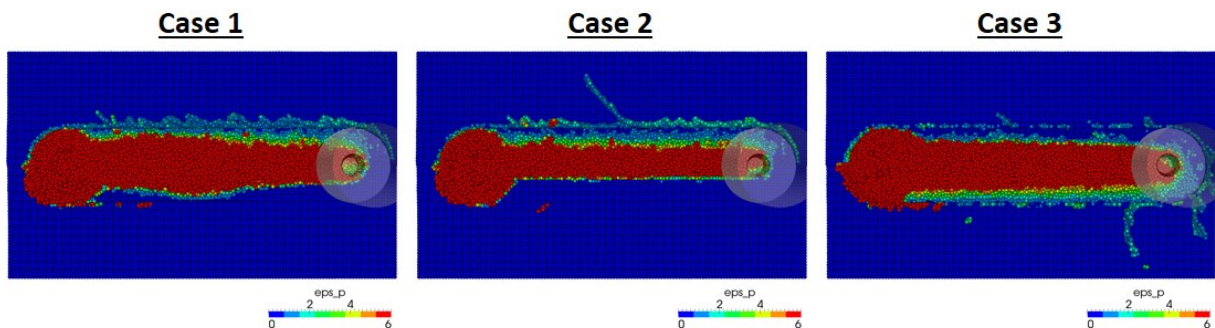
Figure 5-36 - Plastic strain at end of advancing phase showing the effective weld zone

A flash height of 4.2 mm was measured experimentally; case 1 predicts a flash height of 4.5 mm, 3.9 mm for case 2, and less than 1 mm for case 3. The flash heights are shown in Figure 5-35; notice that the wavy pattern of the flash is well represented in the simulation model for case 1. Clearly, the flash produced in case 3 is significantly less than in the other two cases. The reason is entirely due to the change in tool rotation. Flash lines will most commonly be laid down on the retreating side of the weld. By ensuring that the advancing side is on the thicker side, the material is "ripped" from the thicker side and transported to the retreating side. Because of the height change, the flash is not able to attach to the thinner side and creates intermittent "flakes" that can be removed in less time than is possible in the case of a continuous flash line on the thicker side (as in case 1 and 2).

One of the major attributes of a Lagrangian based simulation method is the ability to follow the evolution of any of the material points in the model. Eight points were selected to show how the material moves in the weld zone:

- Shoulder edge at top of plate (points 1 and 5)

- Union between the shoulder and the pin at top of plate (points 2 and 6)

- Weld centerline at top of plate (points 3 and 7)

- Weld centerline at mid plunge depth (points 4 and 8)

The initial location (before the tool interacts with them) of the points is shown in Figure 5-37. Once the tool starts to pass by the points, they will follow distinct paths that are indicative of the level of mixing.
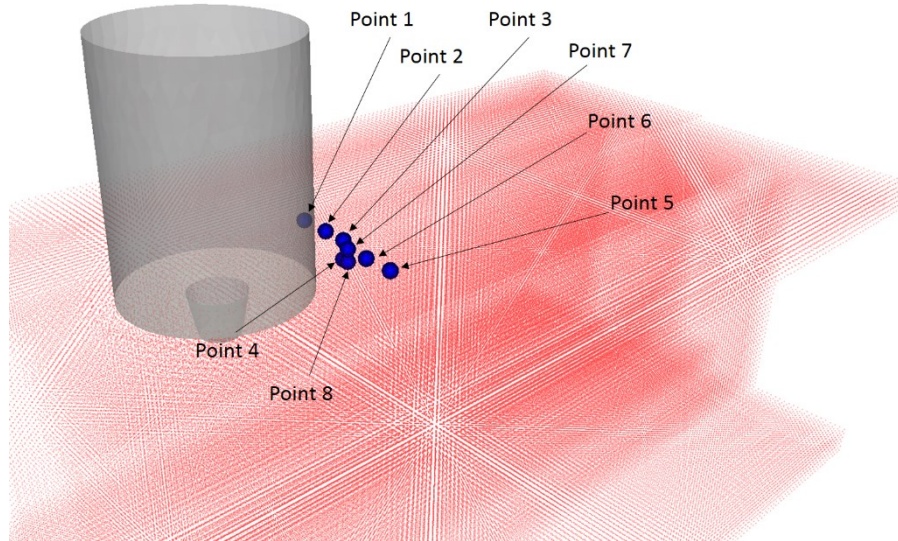
Figure 5-37 – Initial particle positions used in path line analysis

The path lines for the eight particles for the three different cases are shown in Figure 5-38. In Case 1 and Case 2, particles 1, 2, and 3 (initially at plate top, retreating side) all are transported into the flash. This result shows that the surface layer is completely expelled from the weld zone on the retreating side. Also in Case 1 and Case 2, point 4 (initially at weld centerline at mid plunge depth on the retreating side) merely is transported in a semi-circular arc of ~160°. The path lines for the particles on the advancing side for Case 1 show significant travel distances. They follow a path that is no less than two full rotations around the tool. This is in contrast to Case 2; little to no travel distance for the particles on the advancing side was found, although point 6 does make almost a full revolution.
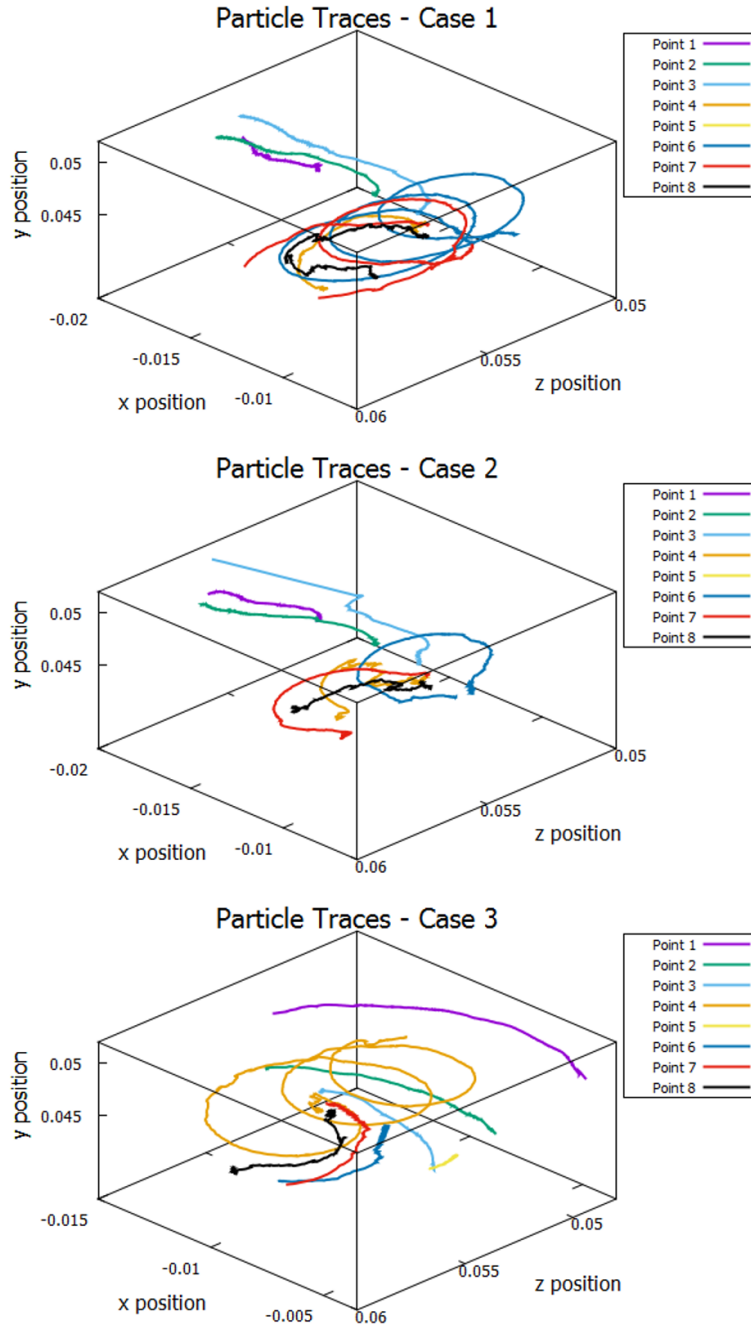
Figure 5-38 – Particle path lines for the three cases

Case 3 shows significantly different path line behaviour. The results would lead to the presumption that there is not as much mixing (particle travel) as would have been expected. Only point 4 travels around the tool (~3 full rotations). All the other points follow semi-circular arcs of varying degree. However, the actuality is that the results are dependent on the choice of tracer particles. The mixing results (Figure

5-39) shows that the level of mixing is as good, if not better, in Case 3 than Case 1 in both the surface mixing and cutaway views. Material mixing will be discussed in more detail in Section 6.1.2 when a mixing metric is developed.
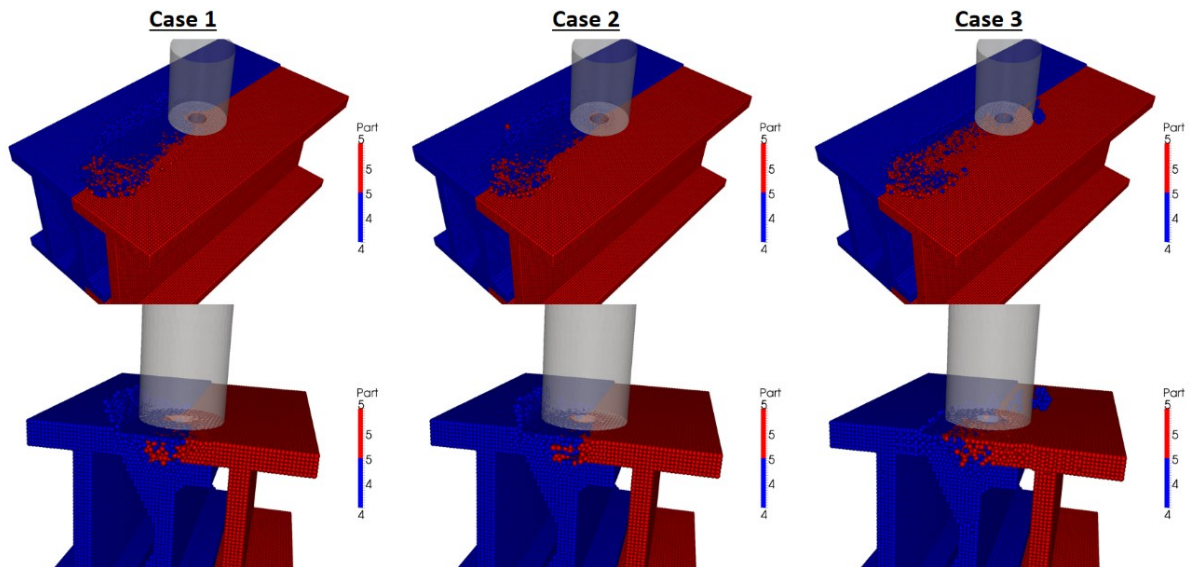
Figure 5-39 – Mixing results for the three cases

5.4   HOLLOW CORE JOINT - BOBBIN TOOL

Bobbin tool FSW is a relatively new variant of the standard FSW process, also called self-reacting FSW. The bobbin tool is composed of an upper and lower shoulder (as shown in Figure 5-40) that squeezes the workpieces to create the required forge force. This additional shoulder attached to the extremity of the pin replaces the anvil at the bottom of the weld. The bobbin tool always rotates perpendicular to the plate to be welded. During the welding process, the material is pressed between the two shoulders and the vertical force (normal to the welded surface) applied on the tool is equal to zero. Consequently, the bobbin tool FSW requires less machine stiffness in the forging direction as compared to standard FSW. However, the stress state in the pin is increased and can be very high. Bobbin tool design is very important.

The bobbin tool FSW has been used to weld hollow core, closed profiles or for application where placing an anvil under the work pieces is difficult. Although bobbin tool FSW looks very similar to FSW, several differences exist between them. The presence of the additional shoulder has a major effect on the heat

generation. Greater heat is generated at the two shoulders and the thermal boundary conditions are modified by the absence of the anvil, which normally acts as a cooling surface. The material flow, the grain size, and orientation are also modified, affecting weld quality.

The tool entry procedure is also different in bobbin tool FSW. Two possible entry procedures are used. In the first one, the distance between the two shoulders is adjusted to be slightly less than the material thickness and the pin of the bobbin tool is pressed against the edge of the material to be welded. At the beginning of the weld, only half of the shoulder surface is in contact with the work piece. In the second entry procedure, a hole is predrilled to ensure that the two shoulders are in full contact with the material to be welded. The bobbin tool can be designed to allow the distance between the shoulders to be variable and to keep the force constant during the whole welding. In this condition, the welding procedure is named adaptive self-reacting friction welding. The entry procedure has a major impact on the weld quality. If the tool is pressed against the edge of the material, the heat generation is greatly reduced as only a section of the two shoulders is touching the surface of the material at the beginning of the weld. The thermal cycle during the process is crucial for the mechanical properties of the weld. The starting procedure must be adjusted to make sure that the tool and work pieces are hot enough before the tool can advance at full speed.

The simulation of bobbin FSW has received much less research attention than that of standard FSW. A transient thermal model featuring the complete moving geometry was developed on Matlab scripting and COMSOL [137]. In this approach, a moving geometry model was implemented using an ALE style approach and the thermal history of a point was predicted. Other thermal 3D models were developed using thermal pseudo-mechanical heat sources, including tool rotation, analytic shear layer model, and ambient heat sinks to simulate the welding equipment and surrounding air [289]. Experimental validation was conducted and the model's prediction was found to be in good correlation for the peak temperature.

The effects of pin features and dimensions of scrolled shoulder bobbin FSW were investigated experimentally and the data obtained were used to develop a conceptual theory representing the underlying physics of the FSW process [290]. A kinematic approach in which the flow process is decomposed into several simple flow components was also developed and applied to the self-reacting

FSW process to explain the basic structural features of FSW welds. A good correlation between the lateral bulging of the nugget, the grain-refined region in the weld center, and the strength of the weld was found [224].

5.4.1   MODEL DESCRIPTION

The simulation model is shown in Figure 5-40; the work pieces are modeled with elastic-plastic-thermal SPH elements and the tool is modeled as a rigid body with zero thickness triangular plate elements. The work pieces are held fixed on their edges (shown in blue in Figure 5-40); this conditions mimics the support arrangement in the lab. The two work pieces are shown as green and yellow particles with a single layer of red at the interface between the two plates. Using a unique color for the interface allows the determination of the evolution of this interface during and at the end of the welding process. A uniform SPH element grid of 1 mm was used for the work pieces; the average finite element size was 0.8 mm for the tool.

The bobbin tool has a scrolled and convex upper and lower shoulder. The diameter of the upper shoulder is 32 mm, and the lower is 25 mm. The pin is a tapered and threaded design with a major and minor diameter of 10 mm and 8 mm. The gap between the upper and lower shoulder is ~6 mm. The experimental setup is confidential and cannot be shown. The tool starts completely out of the work pieces, then advances into the plates and starts to form the weld. The rpm was constant at 600 rpm and the advancing speed was 800 mm/min. The thermophysical properties are the same as in the butt joint weld model. The time step is controlled by the CFL condition which was set to $CFL = 0.7$, leading to $dt_{min} = 1.2x10^{-7}$ s, and $\tilde{V}_{SF} = 16$ (velocity scaling factor) has been used for this case study. Heat loss due to convection ($h_{conv} = 20\ W/m^2K$) is taken into consideration according to sections 3.9.5. The plates were not painted black in this experiment and as such, radiation can be neglected without any loss of precision. Because of the tensile stress field present in the entry region, artificial stress is used with $\epsilon_{ASM} = 0.1$ (see section 3.8.2). The model is run in SPHriction-3D; the runtime for this model was 4 hours on a GTX 980 Ti compared to 132 hours (~5-½ days) on a 4 core Intel Xeon E3-1235 (speed up of 33x). The model runs considerably faster than the FSSW and complex joint models since only the advancing

Kirk Fraser - April 2017

phase needs to be modeled. In the other models, the plunge and dwell phases account for a significant amount of calculation time. In addition, the number of elements in this model is less than in the others.

Work-pieces:
- **6.4 mm AA6061-T6**
- **Elastic-plastic-thermal SPH**

Tool:
- **Rigid finite elements**
- **Hardened steel**

Retreating

Advancing
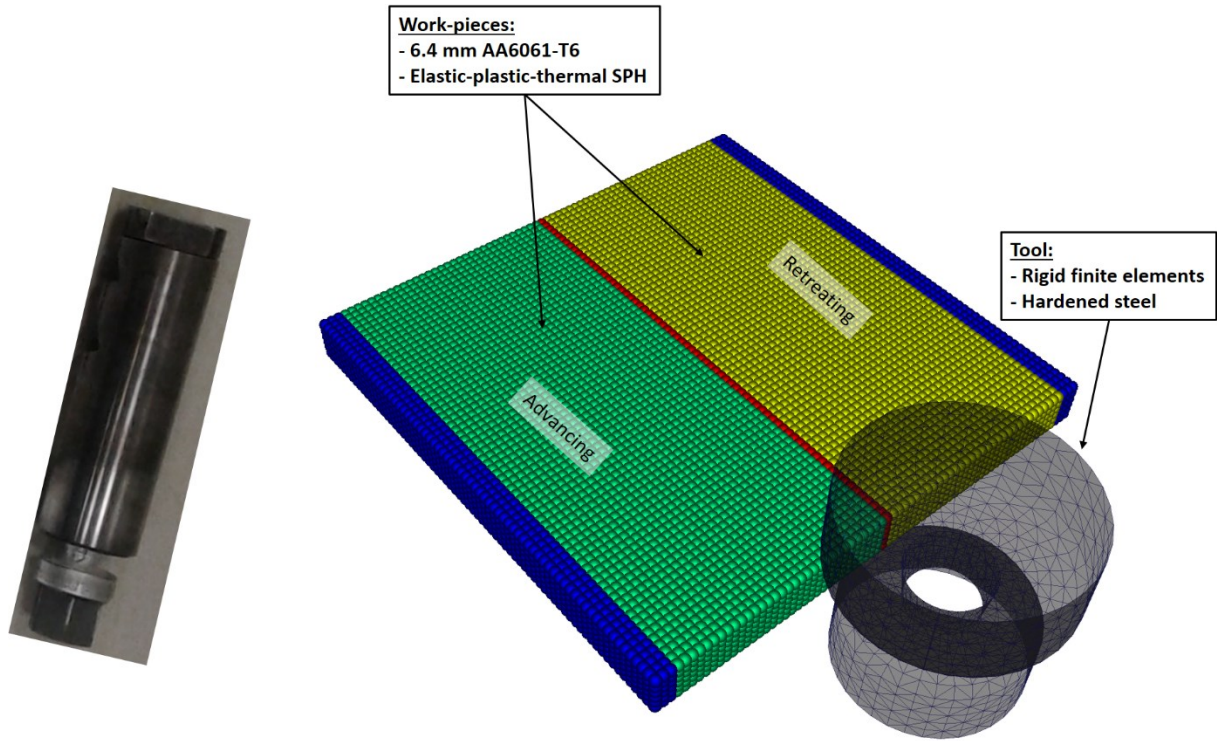
Figure 5-40 – Bobbin tool and simulation model

5.4.2   SIMULATION RESULTS

A comparison of the predicted torque from simulation and experiment is shown in Figure 5-41. Once the tool starts to engage into the work pieces, the torque rapidly increases. At the point that the tool is completely embedded in the work pieces, the torque stabilizes and remains steady until the end of the weld. The tool then breaks out of the work pieces and creates the telltale end defects. Good agreement between the experimentally measured and calculated torque from the simulation model is obtained.

Globally, the calculated torque is slightly higher than the experimental results. This can be explained by likely differences between the tool entry procedures. In the experiment, the tool enters the work piece in a staccato manner where by increasing velocity plateaus are used to increase gradually the tool advance speed. However, in the simulation, the tool enters into the work pieces more quickly. This would have the effect of causing a lower predicted entrance zone temperature, leading to an increase in the predicted

tool force and torque. Another possible discrepancy could be the manner in which heat is transferred from the tool to the work pieces (and vice-versa). Additionally, the uncertainty of the thermal properties of the aluminum alloy at elevated temperatures will play an important role as well.
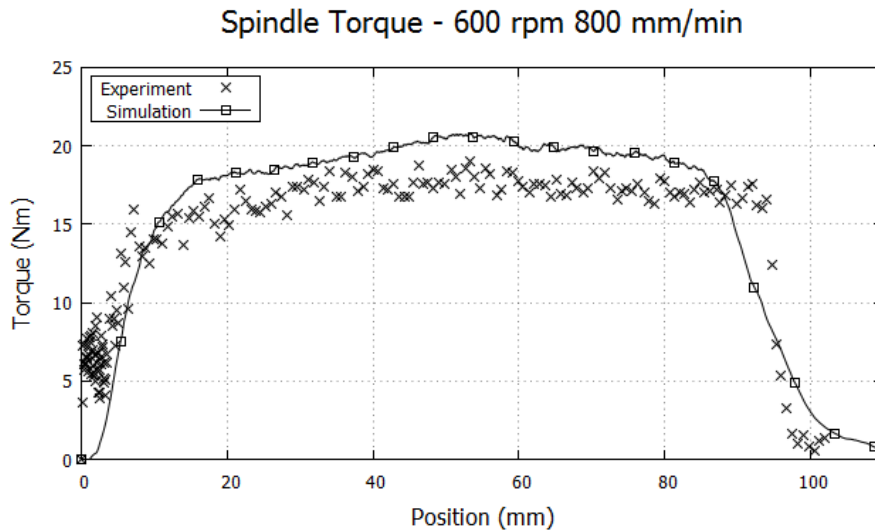
Figure 5-41 – Bobbin FSW spindle torque

A sequence of images at t = 1.92 s, 4.0 s, and 8.3 s (end of weld) is shown in Figure 5-42. On the left side are the temperature results and on the right are the mixing results. The temperature results are shown by performing surface triangulation (see section 4.6) on the work pieces overlaid on top of the particle results. The experimental setup did not allow for temperature measurement, so no comparison is possible. Nevertheless, both the temperature and mixing results are logical. The hottest zone is under the tool as expected, and the temperature profile is asymmetrical (higher on the advancing side).

Excellent correlation between experiment and simulation was obtained for the size and location of the defects as can be seen in Figure 5-43 and Table 5-7. These results show the power of the proposed meshfree method for FSW. No other numerical method could capture these levels of defects with such precision. The size and location of the defects is excellently represented in the simulation model as can be seen in Table 5-7. Of particular interest is that the errors are under 20% for the defect predictions. The distance to form the weld shows a large error; however, measuring this distance is difficult due to the uncertainty of where the weld fully started to form.
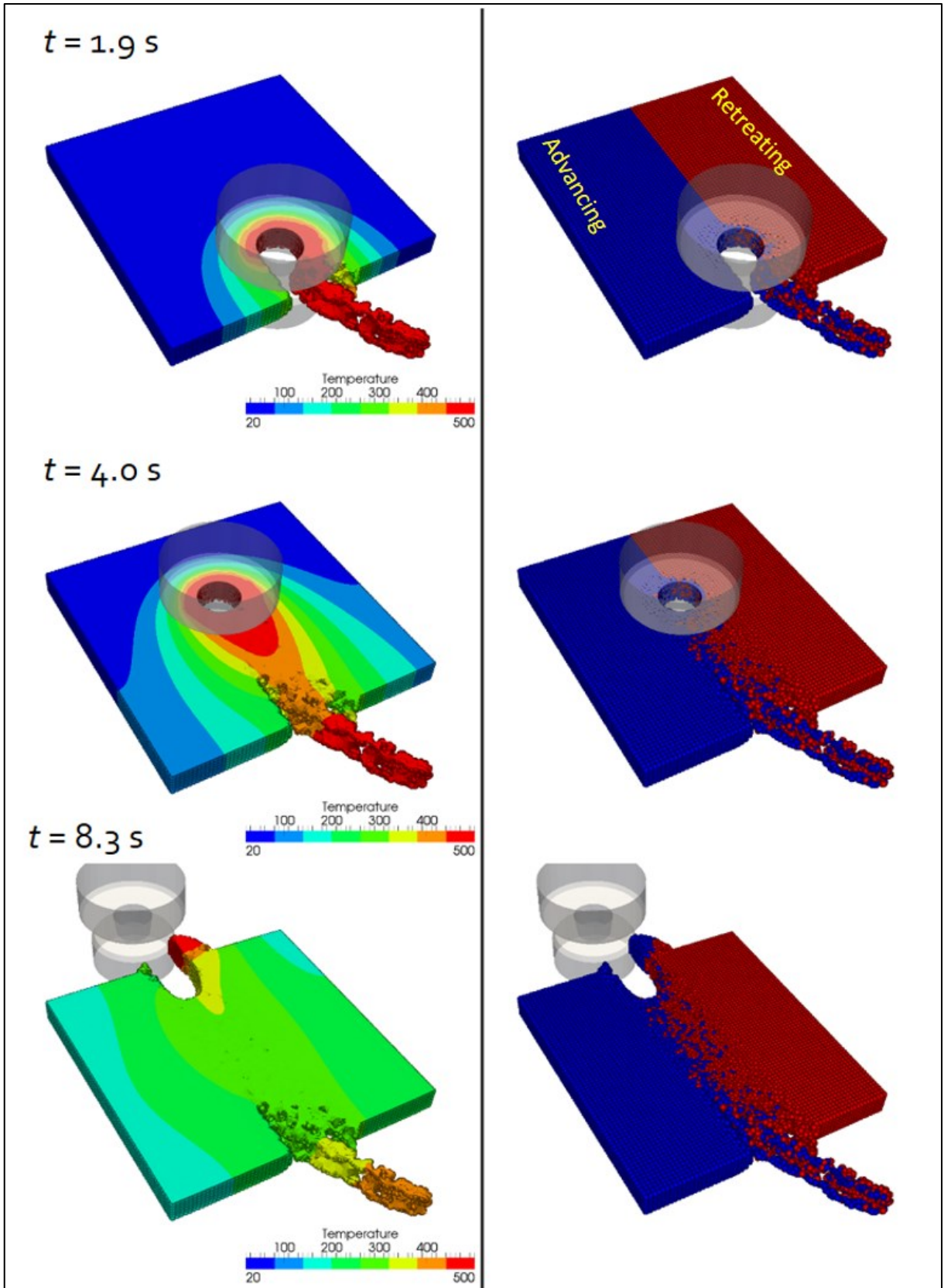
Figure 5-42 – Bobbin FSW temperature (left) and mixing results (right)

The maximum torque error is 5.6%, which is excellent for such a complicated non-linear multi-physics problem. The shape of the tail is different in the simulation model since it does not curve as found experimentally. There are a few possible reasons for this result. The first possibility would be due to the incomplete interpolation found in SPH. The stress at the surface will not be correct, which would lead to an incorrect prediction of the curvature of the tail. In high-speed cutting, Limido *et al.* [176] noted that the curved chip result can only be obtained with the use of a kernel gradient correction, KGC (see section 3.8.1). Attempts to use the KGC in the bobbin tool model lead to divergence of the explicit solver. The reason is entirely due to the large plastic deformation and mixing present in the model. Although the shape of the tail is not well represented, the length is on the other hand. Ultimately, the volume of material expelled from the weld track is more important than the actual shape.
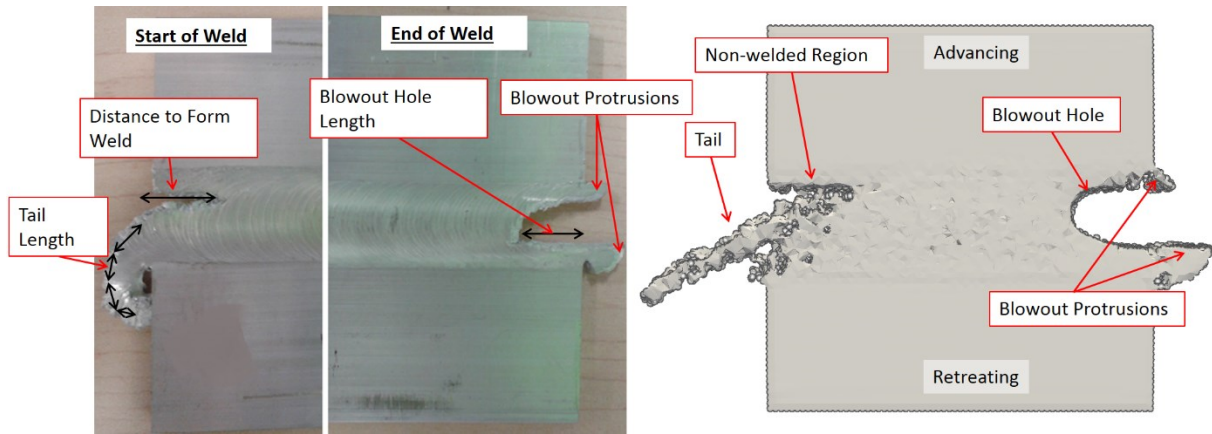
Figure 5-43 – Comparison of predicted and experimental defects

Table 5-7 - Material Flow Comparison between Experiment and Simulation

	Simulation	Experiment	% Error
Tail Length (mm)	50.0	45.0	11.1
Blowout Hole (mm)	20.0	18.0	11.1
Blowout Protrusion (Advancing side, mm)	18.0	15.0	20.0
Blowout Protrusion (Retreating side, mm)	7.0	6.0	16.7
Max Torque (Nm)	20.0	18.9	5.6
Distance to Form Weld (mm)	35.0	20.0	75.0

Another issue with the bobbin simulation with SPH is the tensile instability problem described in section 3.8.2. In the entry region, there is a strong tensile stress field on the advancing side. This causes some unwanted numerical fracture (not physical, see section 3.8.2) to occur. The remedy to this problem would typically be to increase the value of ϵ_{ASM}. However, larger values of ϵ_{ASM} lead to un-desirable behaviour (local spikes in energy) and was deemed to be unacceptable.

Using the wear model described in section 3.14, the wear profile can be predicted. Figure 5-44 shows the wear contours. The highest level of wear is confined within the inner half of the upper and lower shoulder. This result is due to the convex nature of the shoulders. The pressure is going to be higher closer to the pin since the convex profile causes the gap to be smaller in this region.
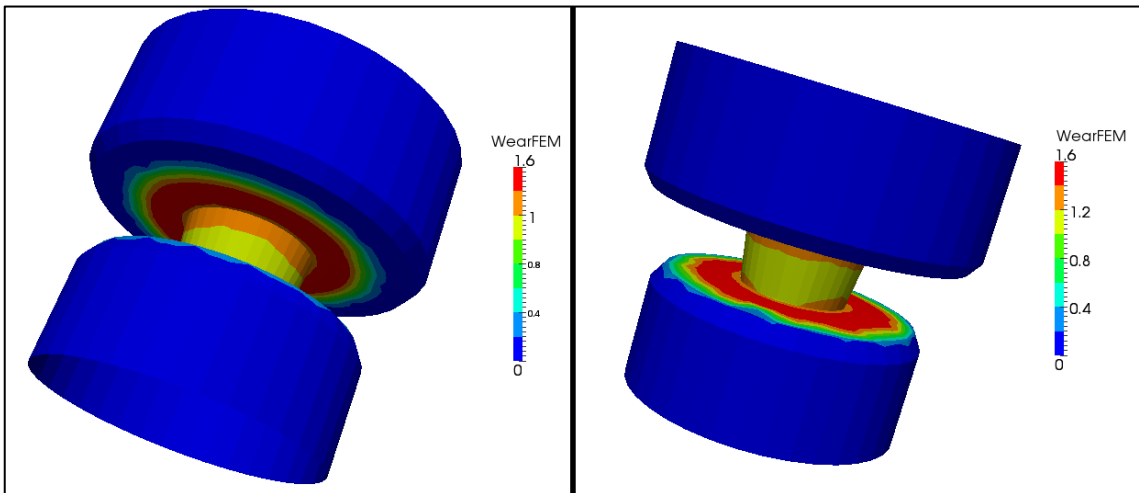
Figure 5-44 – Bobbin tool wear prediction from simulation model

5.5 PARAMETRIC TEST CASES

The results presented in the last section focused on testing the effect of changing significantly important parameters in the simulation model. Now, in this section, the focus will turn towards investigating the effects of other parameters. The following parameters will be investigated:

Smoothing length scale factor - This parameter can be thought of as being the meshfree equivalent of a mesh convergence study in the grid-based methods. Different scale factors will be compared.

Smoothing function - Many smoothing functions can be used for SPH, however, some of the options are not well suited for FSW simulation. Different functions will be investigated and compared.

Velocity scaling factor test - The magnitude of the scaling factor will have an effect on the contact behaviour and the overall system kinetic energy. This test will show how the scaling factor affects the run time and the precision of the solution.

CFL number test - The effect of higher and lower CFL number will be evaluated compared to the typical value of 0.7 that is used in most of the simulation models.

XSPH test - The XSPH method is used to improve the stability of the SPH code. The algorithm essentially adds an attractive force for the elements on the surface of the solid boundary. Different settings will be investigated to show how they affect the FSW results.

Thread pitch testing - A comparison between no threads, thread pitch 0.8 mm, and 1.25 mm will be compared.

Support base material test – During FSW, a large quantity of heat transfers from the work pieces to the support base. Typically, a steel support base is used; in this test case, an insulating and a highly conductive base will be considered.

In some of the above cases, the parameters are needed to ensure that a stable and robust result is obtained in the code. In other cases, the parameters are physics based. In all cases, the parameters are simply settings that can be changed in the code. The results are at best slightly changed for a wide range of some of the specific parameter values. However, some parameters can lead to drastic changes in the simulation results. For the sake of simplicity, the advancing phase of the butt joint model (described in section 5.1) will be used. All simulations are restarted from the final state that was saved from the end of the dwell phase. All of the test cases will use case 3c, 800 rpm 1069 mm/min (see section 5.1.1) unless otherwise specified.

For the smoothing length scale factor, velocity scaling factor, and the CFL number comparison cases, the run time (CPU/GPU time) will be affected. The run time comparisons are made using GTX Titan

Black GPUs. The run times would be ~40% - 60% less had the GTX 980 TI been used. However, the relative difference in timing between the different parameter runs would likely have been similar.

All simulations are performed with a counter-clockwise tool rotation, leading to the advancing side being on the right and the retreating on the left side of the weld line.

5.5.1   SMOOTHING LENGTH SCALE FACTOR

The number of elements that are included in the influence domain is controlled by the smoothing length scale factor, h_{scale} (as described in section 3.2). The greater the value of h_{scale}, the greater the number of neighbors a material point will have. In 3D, an element will have approximately 32, 56, and 80 neighbors with h_{scale} equal to 1.1, 1.2, and 1.3 respectively. As could be expected, the calculation time will be proportional to the number of neighbors used in the SPH equations. A comparison of the runtime is shown in Figure 5-46. The increase in calculation time is linear, $\mathcal{O}(h_{scale})$. Often, one could presume that the precision of the method would increase with increasing h_{scale}, however, this is not the case when the goal is defect prediction. The reason is that increasing the number of elements used when interpolating leads to an over smoothing of the field variables. This is also detrimental in the presence of a surface discontinuity (typical of an internal defect). With a larger smoothing length, elements on the other side of the defect could be used when they rightly should not be. This effect can be visualized in the lower row of images in Figure 5-45. As h_{scale} is increased, the level of predicted defects decreases. In fact with $h_{scale} = 1.3$, almost no defects are predicted. The volume of defects found experimentally is significant (shown in section 5.1.4), and as such, higher values of h_{scale} are not recommended. This phenomenon has been noticed by other researchers [291, 292] working on fracture and fragmentation of solid bodies. For this reason, a value of $h_{scale} = 1.1$ has been used throughout this work. Other values lead to reduced precision of the predicted defect size. Interestingly enough, the temperature (shown in the upper row of images in Figure 5-45) and stress results (not shown) are essentially identical as the scale factor is increased.

$h_{scale} = 1.1$ $h_{scale} = 1.2$ $h_{scale} = 1.3$
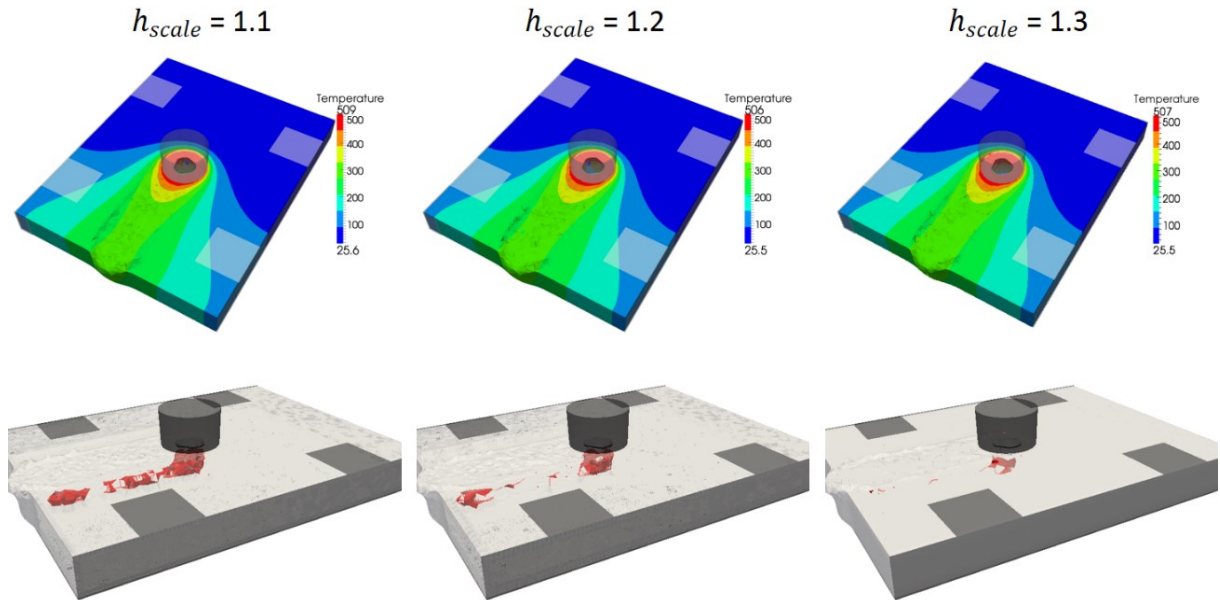
Figure 5-45 – Comparison of the temperature field and defect prediction; $h_{scale} = 1.1$ **(left),**

$h_{scale} = 1.2$ **(center),** $h_{scale} = 1.3$ **(right)**

	1.1	1.2	1.3
Run Time	13095	15300	17460
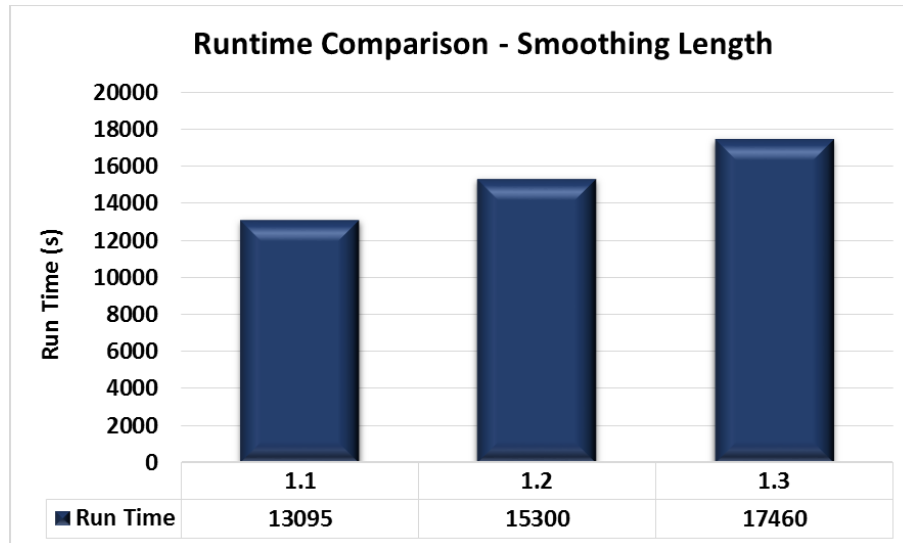
Figure 5-46 - Runtime comparison for the smoothing length scale factor test

5.5.2 SMOOTHING FUNCTION

In recent decades, many different smoothing functions have been developed for the SPH method. Each kernel has its own merit for different types of simulations. The question to decipher is then: which kernel

is best suited for simulating the FSW process: From a close look at the physics of FSW, the smoothing function should be at least continuously smooth up to the first derivative since the conservation equations are formulated using the first spatial derivative of the kernel. Also, due to the predominate compression stress field in FSW, the smoothing function should be well suited to compression dominant problems. The following kernels will be evaluated:

- Cubic B-spline (Monaghan [179])

- Quadratic function (Johnson and Beissel [193])

- Hyperbolic spline (Yang *et al.* [194]) ; this is the kernel used in all other simulations in this work

- Quartic function (Liu and Liu [182])

- Quintic function (Wendland [191])
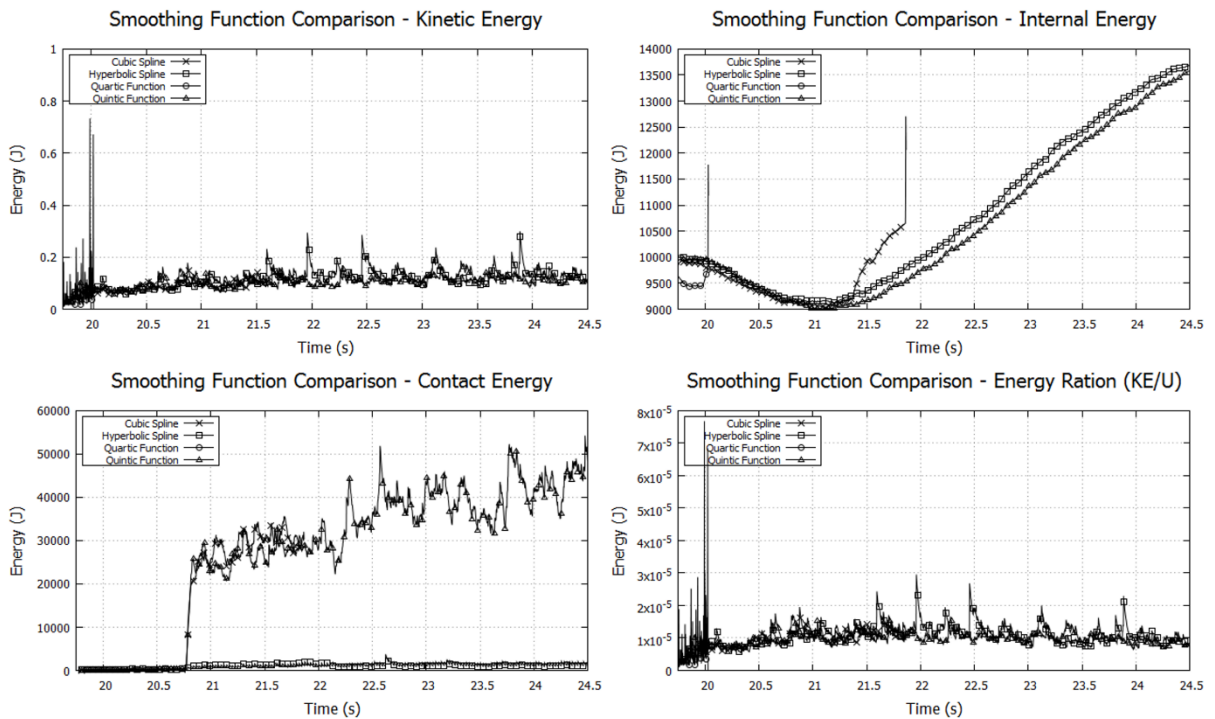
Figure 5-47 – Comparison of system energies for the smoothing function test: Kinetic energy (top left), internal energy (top right), contact energy (bottom left), and energy ratio (bottom right)

The main concern with this comparative study is to evaluate the stability of the solution. The total internal energy, kinetic energy, contact energy, and the ratio of kinetic to internal energy are compared between

the cases in Figure 5-47. The cubic b-spline, quadratic function, and the quartic function kernels all lead to a divergent solution. The point of divergence is signified by the early termination of the curve in the graphs. In fact, the quadratic function kernel leads to nearly immediate solution divergence. This kernel is well known [178] to give poor precision for disorganized particles (as found in FSW). Because of this, not enough data could be collected to graph the energy results for this kernel.

The top left graph shows the total kinetic energy for the system. The energy for the quartic function is erratic, until divergence occurs at $t = 20$ seconds. The energy for the cubic spline stays reasonably bound until divergence at $t = 21.8$ seconds. The energy for the hyperbolic spline and the quintic function are as expected throughout the entire simulation.

The top right graph provides the internal energy results. The divergent behaviour is the same for the cubic spline and the quartic function. Note that the steady increase in internal energy is expected for a FSW simulation due to the continuous input of thermal energy and the continuous increase in the number of elements with plastic strain. The results from the contact energy shed a great deal of light on to the reason why the cubic spline and quintic function were not used in this research project. Because of the high level of compression under the tool shoulder, the contact energy becomes unstable for the kernels that have their first derivative falling to zero as the particle distance decreases to zero. As the particles get closer to one another (compression), the kernel become unstable since decreasing resistance would be present. This sets up a strong overshoot situation in the contact algorithm. The hyperbolic spline kernel has a non-zero first derivation at zero particle spacing. This leads to stable contact behaviour.

The first derivatives of the smoothing functions were previously shown in Figure 3-4, and are repeated in this section for clarity in Figure 5-48. The graph brings to light the underlying reason the hyperbolic spline is well suited for FSW simulations with strong compression fields. Although this kernel was developed to improve the behaviour of water droplet formation (surface tension, a compressive field), there is no doubt that it is an excellent choice for other types of problems as well.
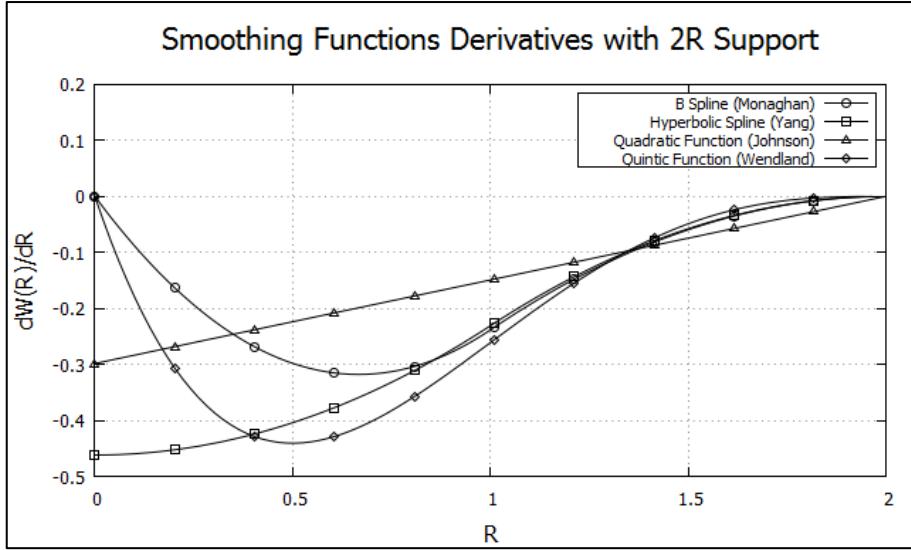
Figure 5-48 - Comparison of smoothing function derivatives with 2R support (repeat of Figure 3-4 for clarity)

5.5.3   VELOCITY SCALING FACTOR

Due to the computationally intensive nature of a full-coupled large deformation meshfree simulation of the FSW process, some form of time scaling is necessary to obtain results within a reasonable period. As discussed in section 2.10, mass scaling is not considered a viable option. As such, velocity scaling is used, which effectively reduces the calculation time by artificially increasing all terms with units of time in the denominator. The choice of velocity scaling factor, $\tilde{V}_{SF}$, is problem specific and requires careful consideration. In this parametric study, $\tilde{V}_{SF}$ is varied between 15 and 40. The convergence of the average defect height is shown in Figure 5-49. The graph shows that the average defect height converges towards the measured value of 5 mm as $\tilde{V}_{SF}$ is decreased. On the other hand, the simulation time is adversely affected by such a decrease in the scaling factor. Figure 5-50 shows the CPU/GPU calculation times at different values of $\tilde{V}_{SF}$.
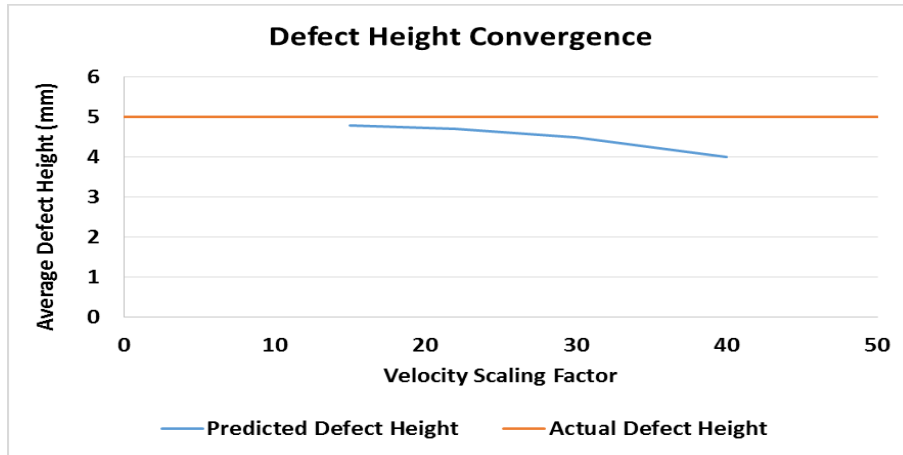
Figure 5-49 – Average defect height convergence as a function of velocity scaling factor

The runtime shows a quadratic dependence on the scaling factor, $\mathscr{O}(\tilde{V}_{SF}^{2})$. Certainly, the quadratic nature is unexpected; common sense would have led to the assumption that the relationship would be linear. A possible explanation could be due to a slight increase in compression with lower $\tilde{V}_{SF}$, leading to more neighbors for the elements in the weld zone (increasing the neighbor search time and the calculation time). Likely the quadratic relationship is due to other factors that are difficult to discern.
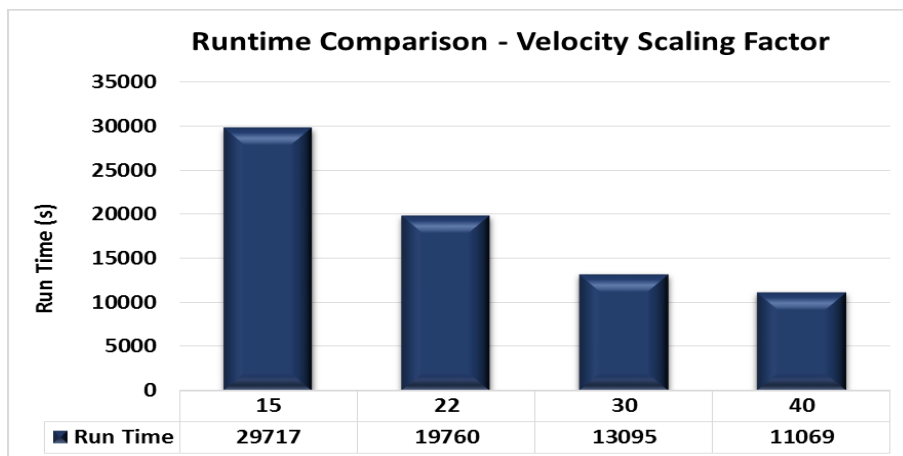
Figure 5-50 - Runtime comparison for the velocity scaling factor

The common procedure for a mass or time scaled simulation is to verify that the ratio of kinetic to internal energy remains low. Figure 5-51 shows the energy ratios for the various values of $\tilde{V}_{SF}$. Clearly, the increased value of the scaling factor increases the energy ratio, however, the increase is not drastic and can still be considered acceptable. In fact, higher values of $\tilde{V}_{SF}$ would be admissible for this specific

model, at the detriment of the precision of the prediction of the size of the defects. In many cases, the design engineer may be interested to increase the scaling factor during initial model development and for preliminary design verification. Then, $\tilde{V}_{SF}$ can be decreased to improve the precision of the simulation for the final runs of the model.
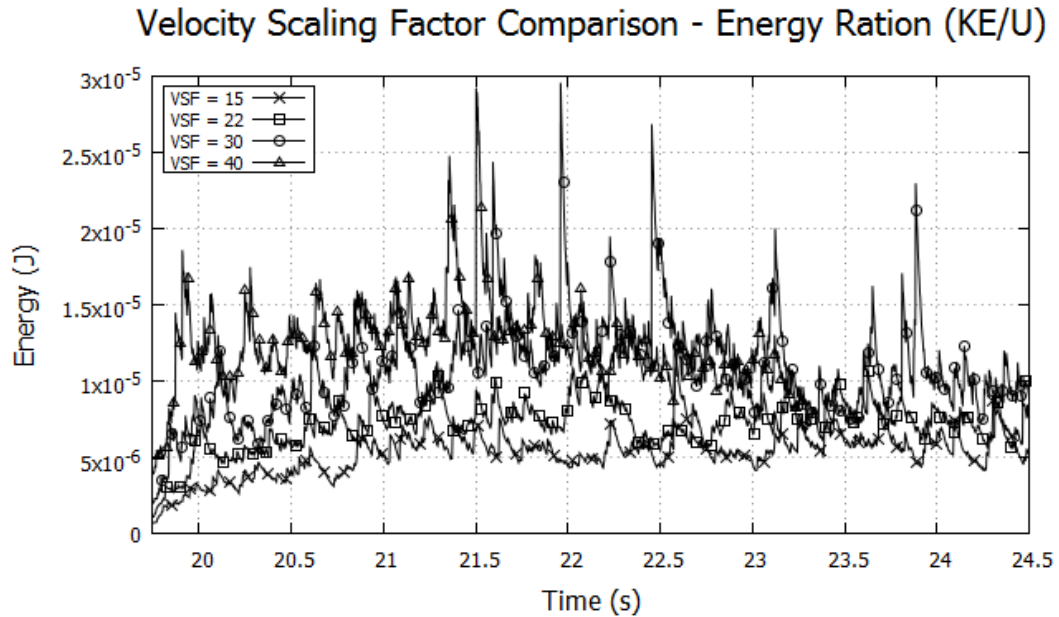
Figure 5-51 – Comparison of kinetic to internal energy ratio for the velocity scaling test

5.5.4   CFL NUMBER

The stability of the meshfree numerical method is directly affected by the CFL number (see section 3.15). As the CFL increases, the solution will become increasingly unstable. Typically, for a formulation that takes into account elastic strain using a full explicit temporal integration scheme, the value of CFL should not surpass 1.0. In this work, a modified temporal integration scheme is used that makes the update of the velocity and position semi-implicit in the stress variable. Because of this, for problems that do not involve contact from one body to another, the CFL number can be greater than 1.0. For example, in a simple elastic-plastic tension test example (see section 9 Appendices for details), SPHriction-3D is able to achieve precise results with CFL $\leq$ 1.5. Certainly, the value of the CFL number will have also have a strong effect on the calculation time. For smaller values of CFL, the solution will take longer. Figure 5-52 shows how the run time is affected by the CFL number. The dependence of the runtime on the CFL

number shows a cubic relationship, $\mathcal{O}(\text{CFL}^3)$. After $CFL = 0.7$, the runtime does not improve significantly. This dependence can be explained by the decreasing level of stability for the contact condition. Although the time step size increases for larger CFL numbers, the time spent in the contact algorithm increases drastically. Because of this, there is very little change in the runtime for $CFL > 0.7$. Better insight into the effect of the CFL number on the contact stability can be seen in Figure 5-53. The kinetic and internal energy history traces show that there is not a significant change with increasing CFL. However, the contact energy history shows a strong change. The contact energy becomes erratic at values of $CFL > 0.8$. The reason is that larger penetrations can occur at larger time steps. Since the resisting force is proportional to the penetration depth, the SPH element will be pushed with very large force, causing the contact behaviour to be sub-optimal. The big spikes in the contact energy graph would be due to overly large penetrations, followed by large push back forces.
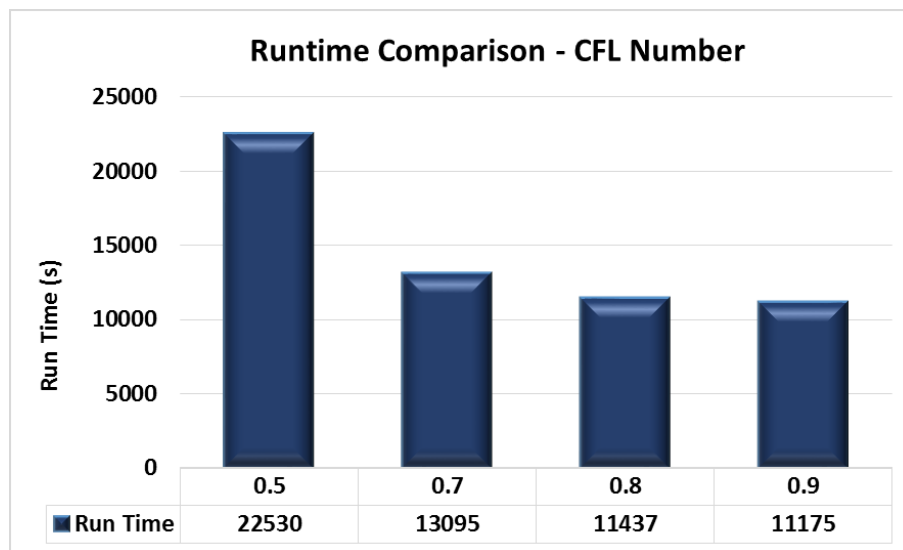
Figure 5-52 – Runtime comparison for the CFL number test

It turns out, in the case of the FSW process, the contact condition will generally override the CFL condition, leading to the need to set $CFL \leq 0.7$. For this reason, $CFL = 0.7$ has been used for the rest of the FSW simulation in this work.
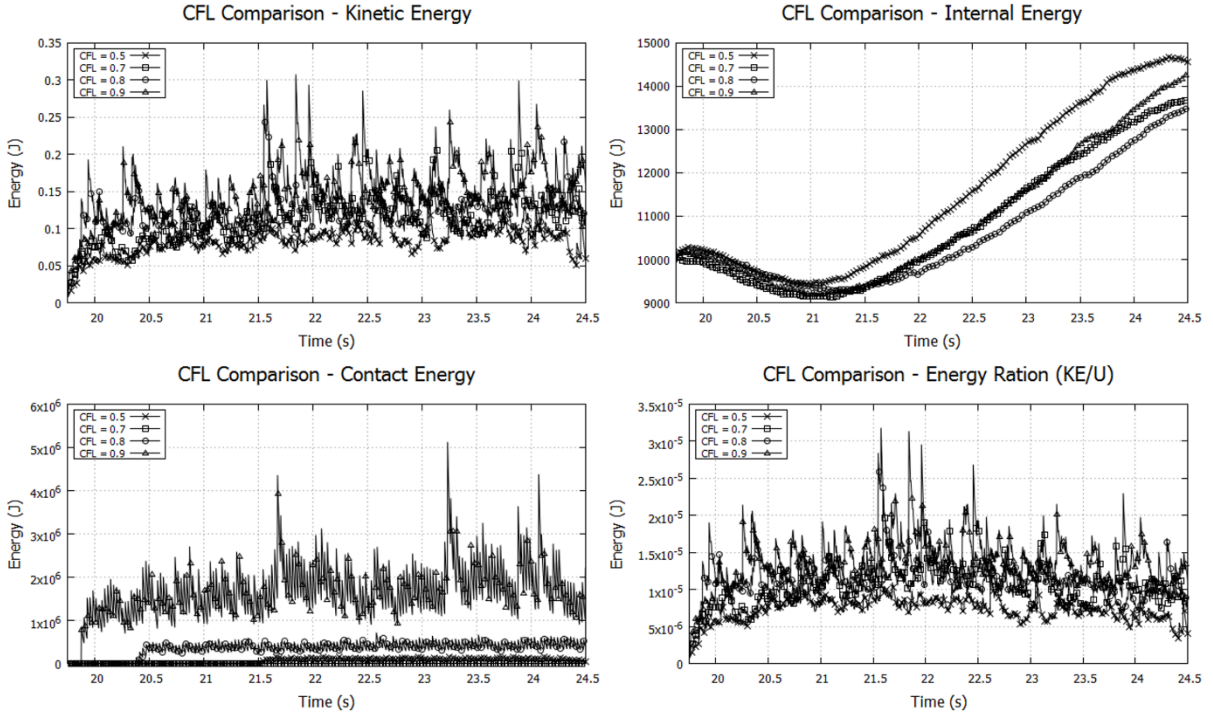
Figure 5-53 - Comparison of system energies for the CFL test: Kinetic energy (top left), internal energy (top right), contact energy (bottom left), and energy ratio (bottom right)

5.5.5 XSPH

The SPH method was originally developed for astrophysics problems where the domain was essentially infinite. For this reason, the SPH method was developed by neglecting the surface integral terms in the conservation equations. Monaghan [206] introduce the XSPH approach for problems with finite domains. The method also prevents one SPH element from penetrating within another element. In this section, the value of ζ_{XSPH} (see section 3.7) is varied to investigate the effect on the FSW simulation results. The XSPH approach in essence helps to prevent an SPH element located on the free surface from "flying off". The XSPH approach can be seen as adding a surface tension like force to the free surface elements. This additional resisting force can lead to reduced precision regarding the size of the predicted internal and surface defects. Figure 5-54 shows the predicted internal defect volume. $\zeta_{XSPH} = 0.05$ (top left) provides an over estimate of the volume of the internal defects (see Figure 5-19 for size of experimental defects). On the other hand, $\zeta_{XSPH} = 0.15$ (bottom left) and $\zeta_{XSPH} = 0.20$ (bottom right) lead to underestimates of the predicted defect volume. A value of $\zeta_{XSPH} = 0.10$ (top right) provides excellent

agreement of the predicted defect size and location. This value is used for all the simulation models in this work.
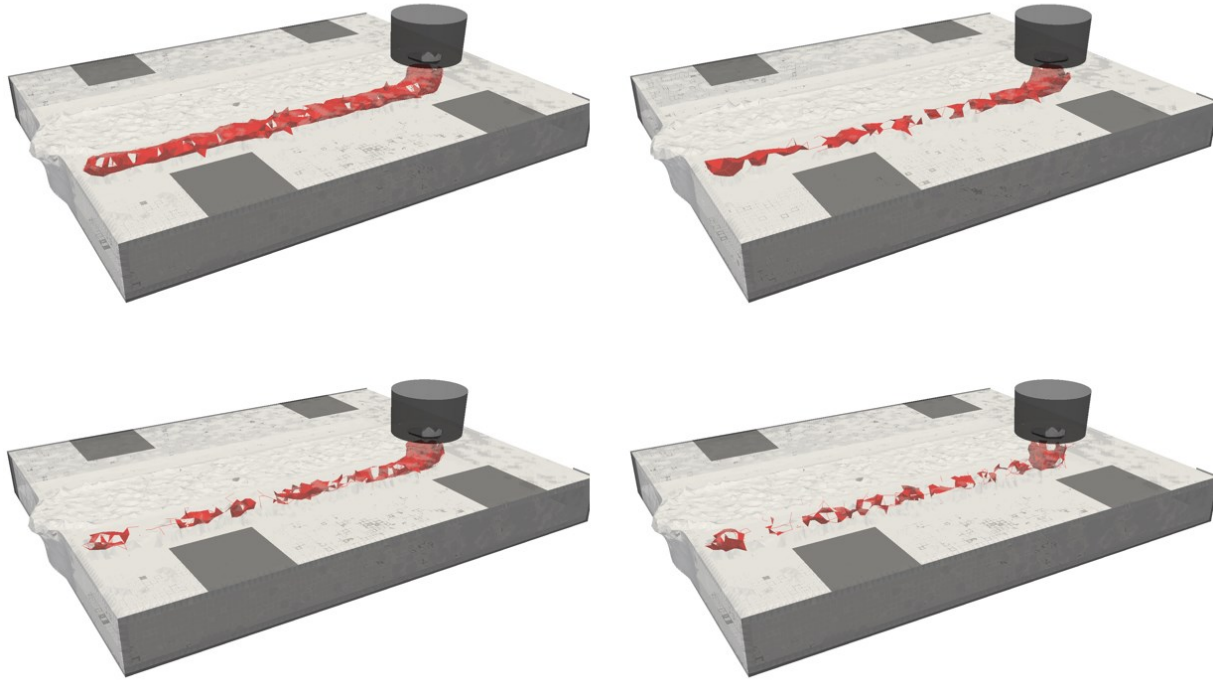
Figure 5-54 – Defect volume comparison for the XSPH test case: $\zeta_{XSPH} = 0.05$ **(top left),** $\zeta_{XSPH} = 0.10$ **(top right),** $\zeta_{XSPH} = 0.15$ **(bottom left),** $\zeta_{XSPH} = 0.20$ **(bottom right)**

5.5.6   THREAD PITCH

This investigation is focused on assessing the effect of changing the thread pitch parameter, p_{thread}, in equation (3-171). Three cases are used; $p_{thread} = 0$ (no threads), $p_{thread} = 0.8\ mm$, and $p_{thread} = 1.25\ mm$ (actual tread pitch for the tool used in the experiments). The test case used in this investigation is that of case 3b (see section 5.1.4). This case set is used since the expected result is to predict no defects in the advancing portion of the weld. In this manner, the effect of $p_{thread} = 0$ can be highlighted. Figure 5-55 shows a cross section view (with the tool just past the cut) for the three cases. The weld zone is significantly hotter for the case of $p_{thread} = 0$. This can be explained by the inability to conduct heat through the gap at the bottom of the weld due to the defect. The weld zone is in essence thermally isolated, and heat is trapped in the weld zone. The image shows the relative size of the predicted defect;

the overall height being the greatest for the $p_{thread} = 0$ case. The threads have the effect of transporting the hot material from under the shoulder down towards the bottom of the joint. This will tend to distribute the thermal energy uniformly in the weld joint, improve the material flow, and increase the overall level of plastic deformation in the weld zone.

Large Void (2.1 mm)

Medium Void (1.3 mm)

No Void
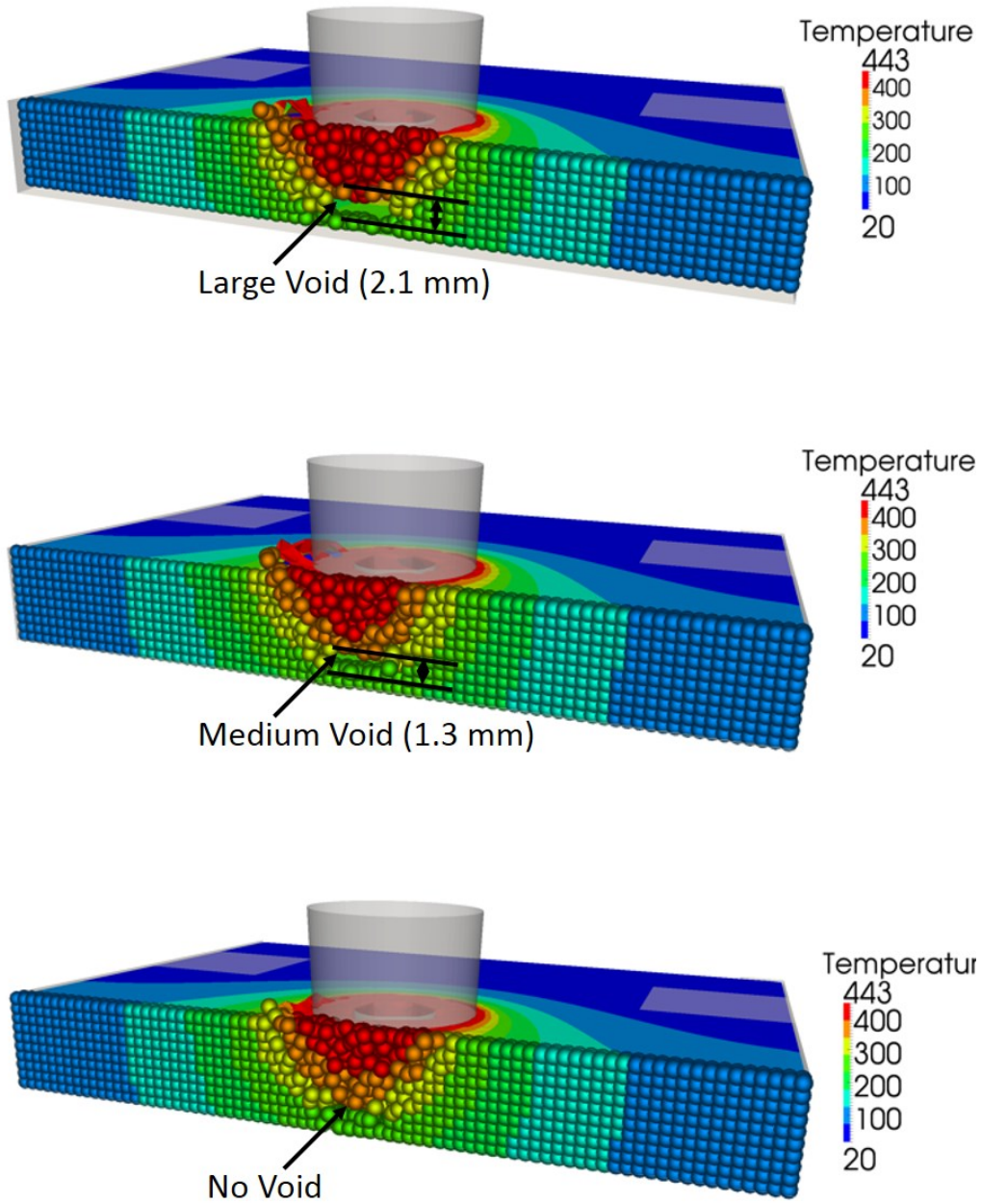
Figure 5-55 – Cross section cut through thickness of work pieces: $p_{thread} = 0$ **(top),** $p_{thread} =$ $0.8\ mm$ **(center), and** $p_{thread} = 1.25$ mm **(bottom)**

Figure 5-56 shows a comparison of the predicted internal defect volume for the different values of p_{thread}. The $p_{thread} = 0$ and 0.8 mm cases both show a large volume of predicted defects. The $p_{thread} = 1.25$ mm case predicts the correct volume of defect (as was shown in Figure 5-18). Certainly, the reason for the increased level of defects predicted with less thread pitch is evident; during mixing, the threads incur a downward movement of the material that helps to close in the weld cavity to create a defect free weld. In the FSW case studied here, the threads provide an efficient means to perform the weld at a higher rate of advance without the presence of defects throughout the weld zone.
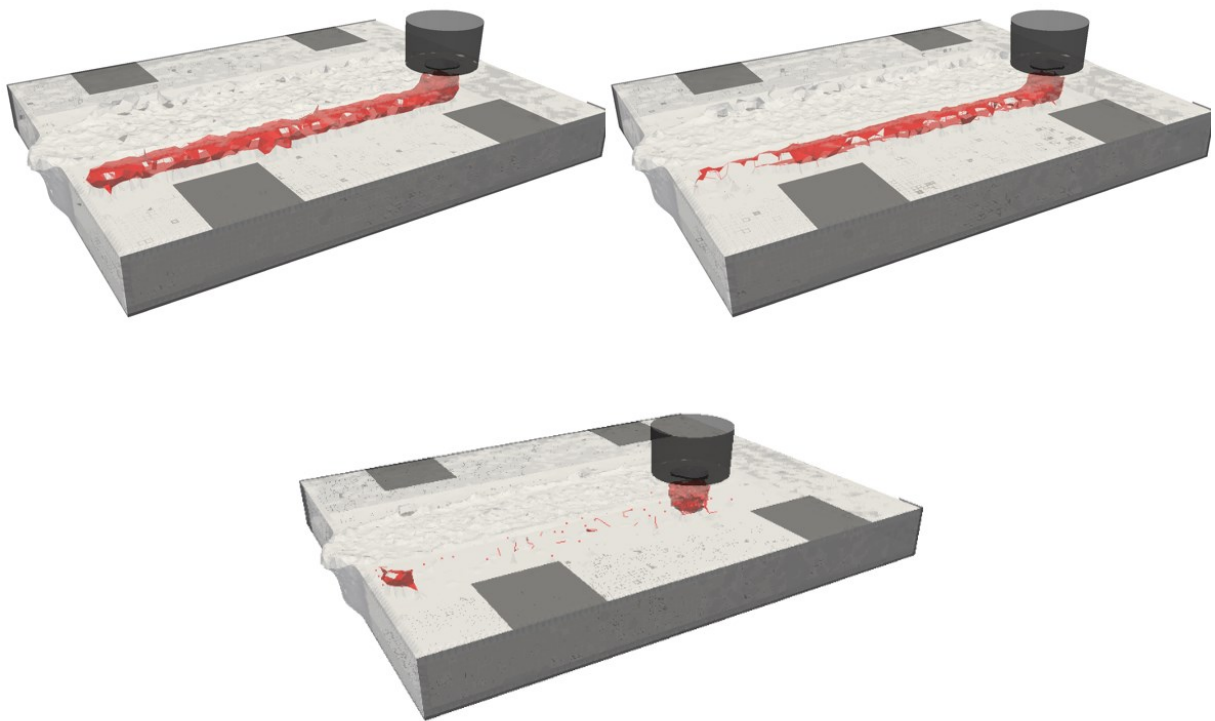
Figure 5-56 – Defect prediction comparison: $p_{thread} = 0$ **(top left),** $p_{thread} = 0.8$ mm **(top right), and** $p_{thread} = 1.25$ mm **(bottom)**

5.5.7   SUPPORT BASE MATERIAL

The thermal properties of the support base play a vital role in the quality of the finished weld. The principle of this test is to show the importance of the material properties by using extreme values for the thermal diffusivity. An insulating material with a low thermal diffusivity (fireclay brick) and a strong conductor with high thermal diffusivity (copper) will be compared with a steel support base. The thermal properties of the cases considered are shown in Table 5-8.

Comparisons of the temperature and defect results are shown in Figure 5-57. The surface temperature is shown in the top row. Notice that the overall surface temperature is higher in the copper base case. This can be explained by the increased diffusivity of the copper material. Since the simulation model uses a finite sized base, the temperature saturated more quickly in the copper base, leading to a marked increase in the temperature of the bottom of the workpieces. The change in the defect volume (middle row of Figure 5-57) is an important result of the different base materials. The brick support base incurred significantly less defect volume than the other cases. The cross-section view of the temperature profile in the bottom row of Figure 5-57 helps to understand this phenomenon. The cross section is taken half way along the weld line at the instant when the tool has just passed that point. Notice that the thermal energy is "trapped" closer to the weld zone in the brick base case. This helps to increase the overall weld zone temperature, leading to more favorable welding conditions.

Table 5-8 – Thermal properties of base materials used in test case

	Fireclay Brick	Steel	Copper
ρ – Density [kg/m^3]	2645.0	7850.0	8930.0
C_p – Heat capacity [J/kgK]	960.0	485.0	385.0
k – Thermal conductivity [W/mK]	1.0	55.0	400.0
$\alpha = k/\rho C_p$ – Thermal diffusivity [m^2/s]	3.94E-7	1.46E-5	1.16E-4

From these results, one could infer that using an insulating base is beneficial to the weld quality for a certain set of weld parameters. More importantly, it can be concluded that by decreasing the thermal

diffusivity of the base, a higher advancing speed could be used than would have been possible with a moderate to high conducting base.

Often, the FSW community has been interested in the percentage of thermal power input due to surface convection, radiation, plastic deformation, tool pin base, tool pin side, and tool shoulder. Such a comparison is provided in Figure 5-58 (instantaneous power results) and Figure 5-59 (average power). The power results are evaluated based on the equations developed in Sections 3.9.1 (plastic work), 3.9.2 (friction work), 3.9.5 (convection), and 3.9.6 (radiation).
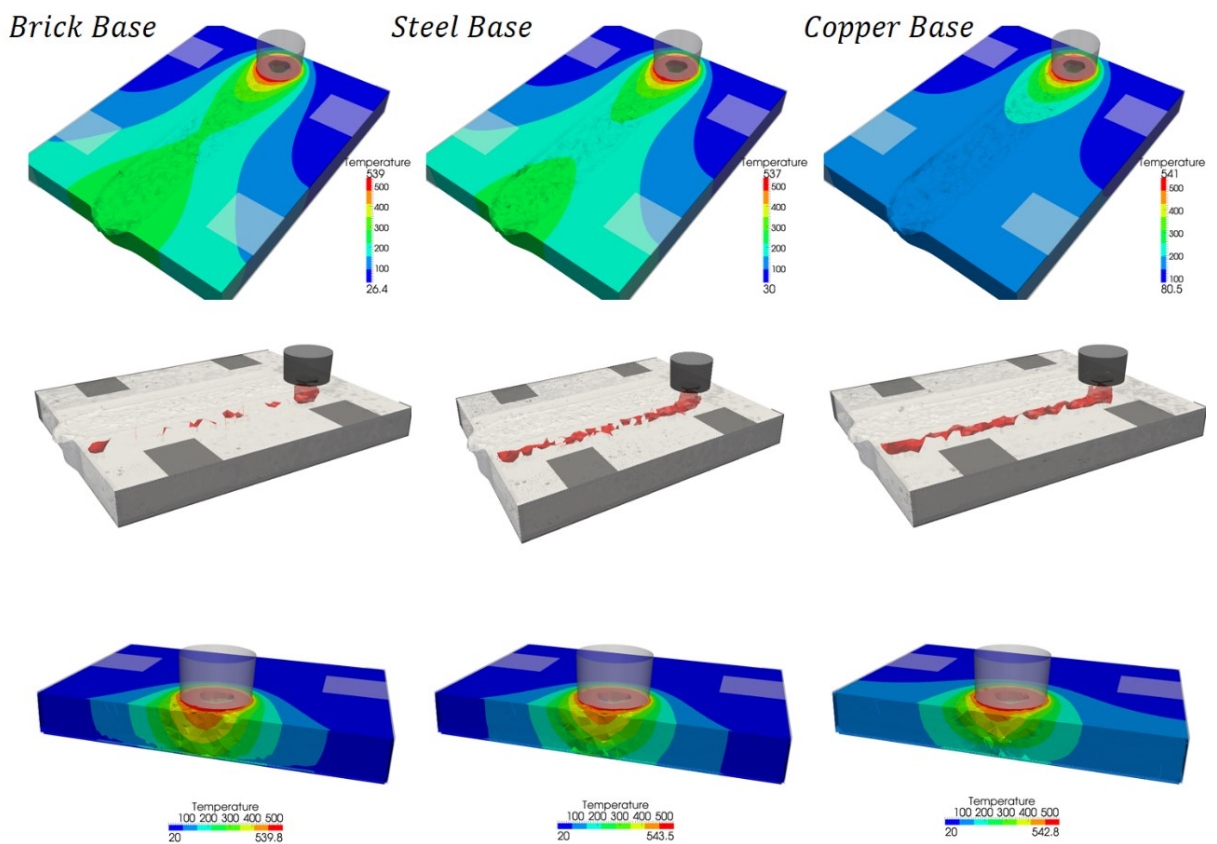
Figure 5-57 – Comparison of the temperature and defect results for the base material test: surface temperature (top row), internal defects (middle row), and cross section temperature (bottom row)

The overall system values for the individual heat source terms are found by accumulating (summation) the value at each element in the work pieces and converting to power by multiplying by the element volume. From the instantaneous power results, the pin base results show the largest variation between

the different base types. The pin base power is highest in the copper base model since the heat is evacuated more quickly. This leads to an increased pressure condition between the tool pin base and the work pieces, causing the pin base power to increase. The other results are relatively similar; however, the radiation results are also informative. The highest radiation power is found in the brick base model. This can be explained by the fact that the heat is more concentrated in the weld zone, leading to an increase in radiation effects.
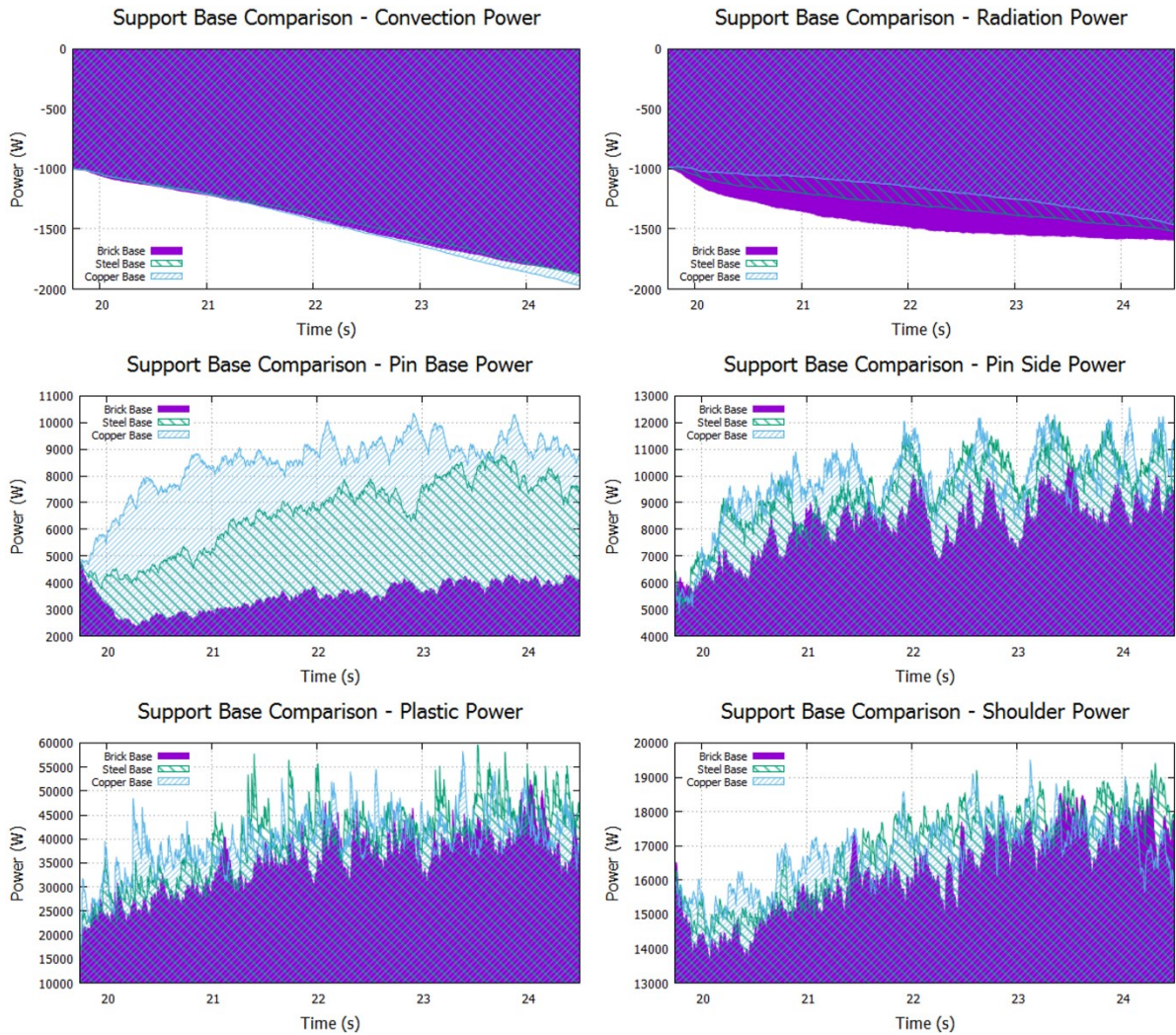
Figure 5-58 – Support base material power comparisons

The average power graphs in Figure 5-59 provide insight into the relative importance of the plastic power, and the tool pin base, side, and shoulder. In general, for the case considered here, the plastic deformation power accounts for ~55%, the tool pin base ~8%, the tool pin side ~13%, and the shoulder

~24%. Certainly, these percentages are problem and welding parameter specific. The plastic deformation will tend to increase as the weld pitch decreases (more rpm per advance) due to an increase in the heat generated from friction. The total tool power input is shown in the bottom right of Figure 5-59. Here the brick support base case shows the lowest total tool power, followed by the steel base, and the copper base with the highest power. The reason is as previously mentioned; the increased average temperature in the weld zone reduces the contact pressure, reducing the tool power.
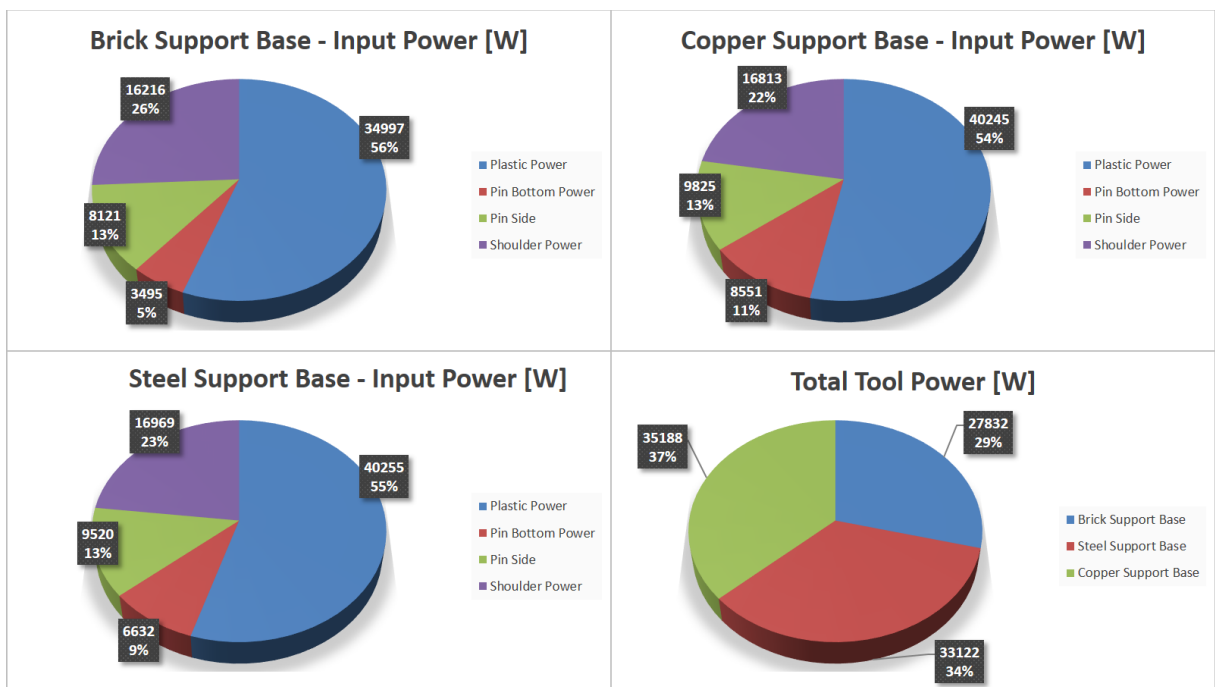
Figure 5-59 – Support base material input power comparisons

5.6   COOLING AND DISTORTION OF A BUTT JOINT WELD

The simulation code can be used to determine the cooling and distortion of the butt joint weld in section 5.1. The new adaptive thermal boundary conditions will allow heat to be removed from the free surface of the work pieces and predict the final deformed shape and the locked in stresses.

5.6.1   MODEL DESCRIPTION

The FSW process can be explained by the four main phases of the process:

Phase 1 – Plunge: The FSW tool is pressed into the work pieces (WPs) with a fixed rpm

Phase 2 – Dwell: The FSW tool remains stationary with a fixed rpm. This phase serves to continue to heat the WPs

Phase 3 – Advance: The FSW tool advances with a constant velocity and rpm. The weld is formed during this phase

Phase 4 – Cooling: The tool is retracted and the completed joint cools to room temperature

During Phase 1 and 3, enormous levels of plastic deformation occur in the aluminum in the region close to the tool (mechanically affected zone). These phases are responsible for creating a high strength weld. As the WPs cool in Phase 4, the locked in stresses are redistributed. The final deformed shape of the plates and the residual stresses are found when the WPs have completely cooled.

Table 5-9 – Material properties used in cooling analysis

	6061-T6 (Work Piece)	Base and Tool (Steel)
k (W/mK)	175	55
ρ (kg/m³)	2700	7850
C_p (J/kgK)	895	485
E (GPa)	$f(T)$ see Eq. (5-4)	210
σ_{y0} (MPa)	276	N/A
T_{melt} (°C)	582	N/A
T_{room} (°C)	20	N/A

The simulation starts from the final state of a FSW simulation (end of Phase 3) with 800 rpm and 660 mm/min advance. A video of all the phases of the FSW process (clamp, plunge, dwell, advance, retract, cool) can be viewed here: https://www.youtube.com/watch?v=pbjZnLV3yXA. The initial conditions for the cooling analysis are shown in Figure 5-60. The material properties are shown in Table 5-9. The convection coefficient, h_{conv} is taken as $10 \, \text{W/m}^2\text{K}$. Time scaling is used for the cooling simulation since

the time to cool the plate entirely is on the order of ~ 30 minutes. Since the mechanical deformation is very minimal during the cooling phase, we have found that we can use a time scaling factor of 1000 without incurring momentum errors (ratio of kinetic energy to internal energy remain small). To accomplish this, the thermal conductivity and the heat loss coefficient are scaled accordingly. A thermal softening perfect plasticity model is used for the aluminum work pieces.
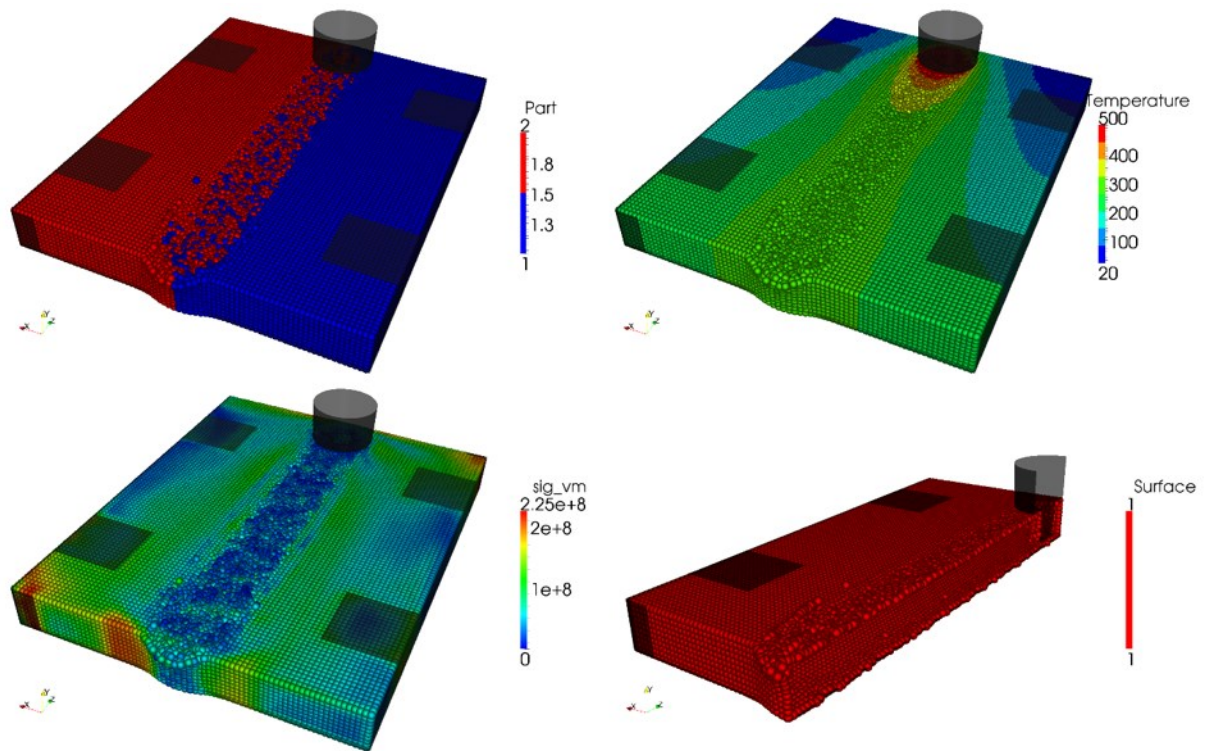
Figure 5-60 – Initial conditions at start of cooling analysis (clockwise from top left): Mixing results in the weld zone, temperature distribution (°C), effective stress (Pa), surface particles

The surface of the work pieces is painted a matte black in order to capture the temperature distribution with a thermal camera during the FSW process. The emissivity of the paint, ε, is ~ 0.95. Because of this, the heat loss due to radiation accounts for a significant portion of the heat loss. The total heat flux at the surface is then a combination of that from convection (q''_{conv}) and radiation (q''_{rad}):

$$q''_{total} = q''_{conv} + q''_{rad} = h_{conv}(T_\infty - T_s) + \varepsilon \sigma_{SB}(T_{surr}^4 - T_s^4) \tag{5-6}$$

σ_{SB} is the Stefan Boltzmann constant (5.67×10^{-8} W/m^{2}K^4) and T_{surr} is the temperature of the surrounding solid envelope (taken as 20°C). Figure 5-61 shows the temperature distribution in the finished weld (a comparison between experiment and simulation was previously shown in Figure 5-11). The temperature distribution predicted by SPHriction-3D (see upper right image in Figure 5-60) is in good agreement with the thermal camera image. Note that the temperature in the weld track cannot be measured by the thermal camera since this zone does not have the matte black paint. The emissivity of the aluminum in the unpainted weld zone is very low (~ 0.1) and leads to an inaccurate reading of the temperature.
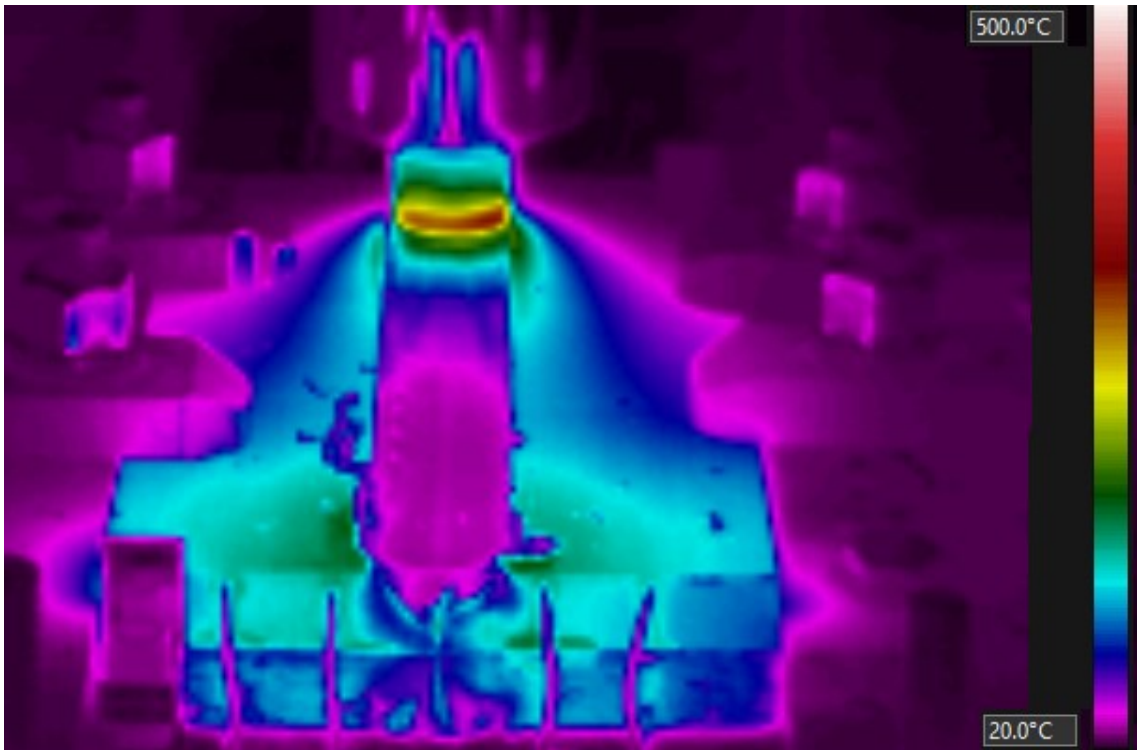
Figure 5-61 – Thermal camera image from FSW experiment at end of advancing phase

The temperature distributions for four different points in time are shown in Figure 5-62. The distributions are strongly influenced by the location of the supports. The last region of the WPs to cool is the weld zone because the supports are made of steel and have a significant thermal mass to absorb the energy from the WPs.
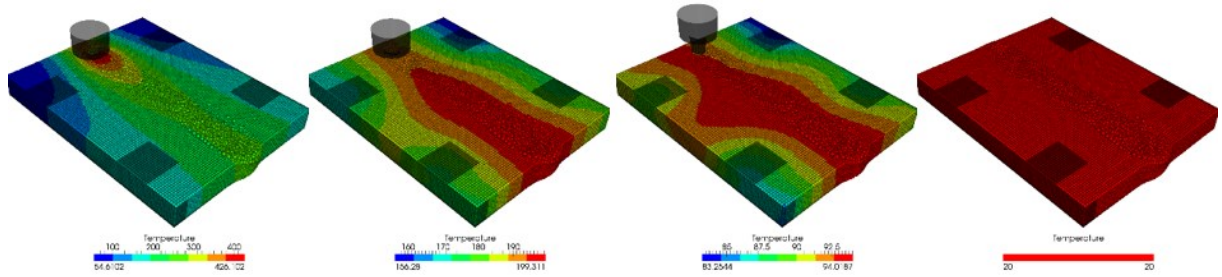
Figure 5-62 – Temperature profiles (left to right): t = 24 s, 32 s, 105 s, and 524 s (final step)

Four "thermocouples" were inserted into the WPs in order to record the temperature history throughout the simulation. The temperature at a so-called "thermocouple" (TC) is determined in the SPH simulation model by interpolating from a set of SPH points in the neighborhood of the TC. In this sense, the temperature at the ith TC is:

$$T_{TC_i} = \frac{\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} T_j W_{ij}}{\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} W_{ij}} \tag{5-7}$$

TC3 and TC4 are inserted on the retreating and advancing side of the weld, 5mm from the weld line. TC1 and TC5 are 23 mm from the weld line (again on the retreating and advancing side). All thermocouples are embedded at a distance of one quarter the length of the WPs (from the end of the weld) and 2.5 mm deep from the surface of the WPs. A schematic of the TC locations was previously provided in Figure 5-2. Figure 5-63 shows the temperature history for the four TCs. From the graph, we are able to infer that the region closest to the weld zone is hottest. Furthermore, according to the simulation results, the temperature on the retreating side is slightly hotter than that on the advancing side. This can be explained by the movement of the material: as the tool rotates, the material on the advancing side is heated and transported to the retreating side.
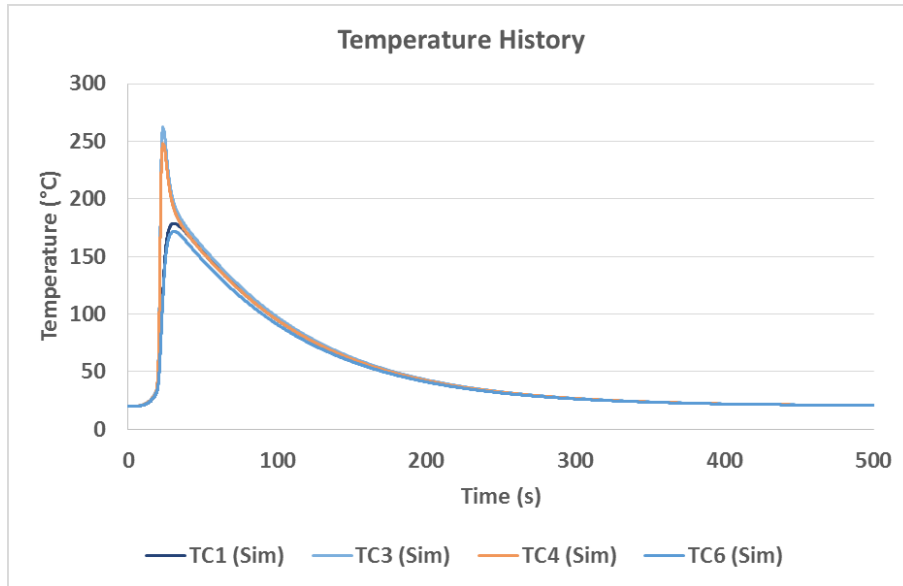
Figure 5-63 – Temperature history for TC1, TC3, TC4, and TC6

5.6.2 DIGITAL IMAGE CORRELATION

A digital image correlation (DIC) (http://www.gom.com/metrology-systems/system-overview/aramis.html) system called ARAMIS, developed by GOM, was used in this work. The system is used to determine the final distortions (and by extension the perceived stresses) in the work pieces. Some example applications of the ARAMIS system are shown in Figure 5-64.
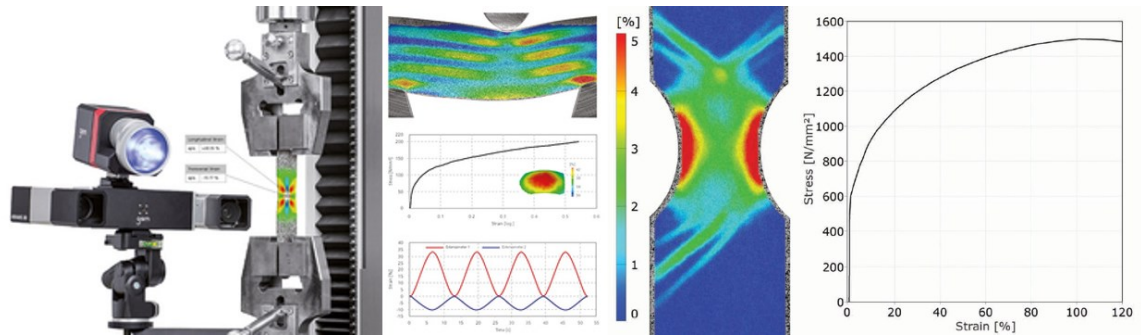
Figure 5-64 – ARAMIS examples (from GOM's website, see above text)

The method works by comparing a series of digital images to a reference image (in the initial un-deformed state). The images are taken with the Aramis digital camera that is connected to the Aramis software. In order for the program to compare the images, a stochastic pattern (as shown on the surface of the work pieces in Figure 5-66) must be painted onto the region that is to be measured. The software

is able to recognize changes in the location of key points within the stochastic pattern and uses these changes to calculate the deformation gradient on the surface of the measured object. The manner in which the deformation calculations are performed by the ARAMIS system are similar to the spatial integration approach used in SPH (collocation).
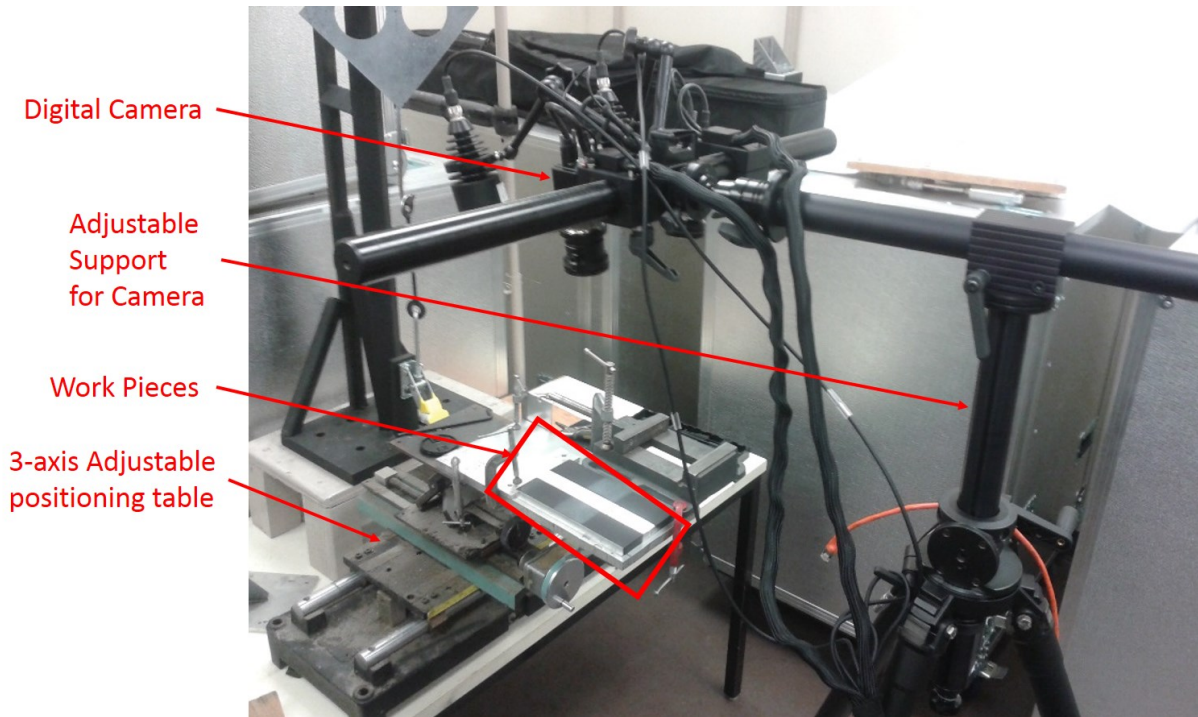
Figure 5-65 – Digital image correlation setup

Our approach to find the residual stresses requires us to take an initial image of the plates prior to welding, then once the weld is completed (and fully cooled), to take a final image. The initial and final images must be taken with the same camera settings and environment. A built in assumption to use this approach for residual stress measurement is that the region where the distortions are measured deforms elastically. This assumption is reasonable since the measurement zone is at least 2 mm from the weld path.

The imaging setup is shown in Figure 5-65. Since we are measuring a flat plate, a 2D measuring approach with one digital camera is appropriate. The imaging equipment is setup in an area of the lab with no traffic to mitigate the possibility that the orientation of the camera could be altered. This would render any obtained results useless because the camera calibration would no longer be correct.

Figure 5-66 – Pre-weld setup for DIC showing painted work pieces, thermocouples and support structure

A summary of the measuring process is provided below:

1. Set up the DIC equipment in a stable area where there is no risk of the setup being disturbed

2. Calibrate the DIC setup using the ARAMIS (provided by GOM) calibration block

3. Paint the region to be measured with a stochastic pattern

4. Take an initial image before the welding process

5. Clamp the work pieces into the welding apparatus

6. Perform the weld

7. Unclamp the plates, let them cool completely

8. Take a final image

9. Use the ARAMIS software to determine the deformation that occurred between the initial and the final images

10. Post process the deformation to obtain the residual stress (assuming linear elastic following Hooke's law)

5.6.3   RESIDUAL STRESS RESULTS

The ARAMIS system was used to determine the residual stresses on the surface of the work piece (on the retreating side). The measured residual stresses (showing effective stress in Pa) from experiment are shown in Figure 5-67.
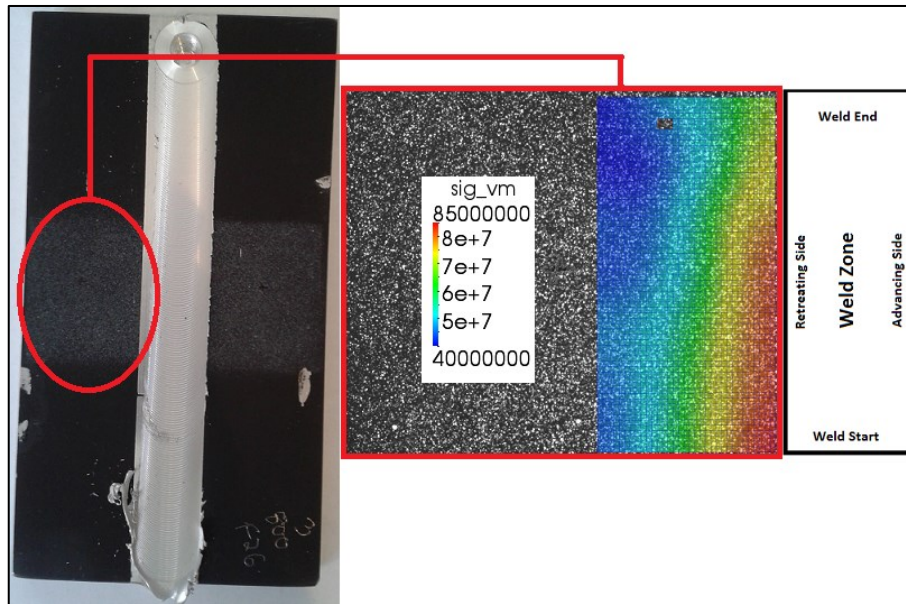
Figure 5-67 – Residual stress results from FSW experiment using DIC
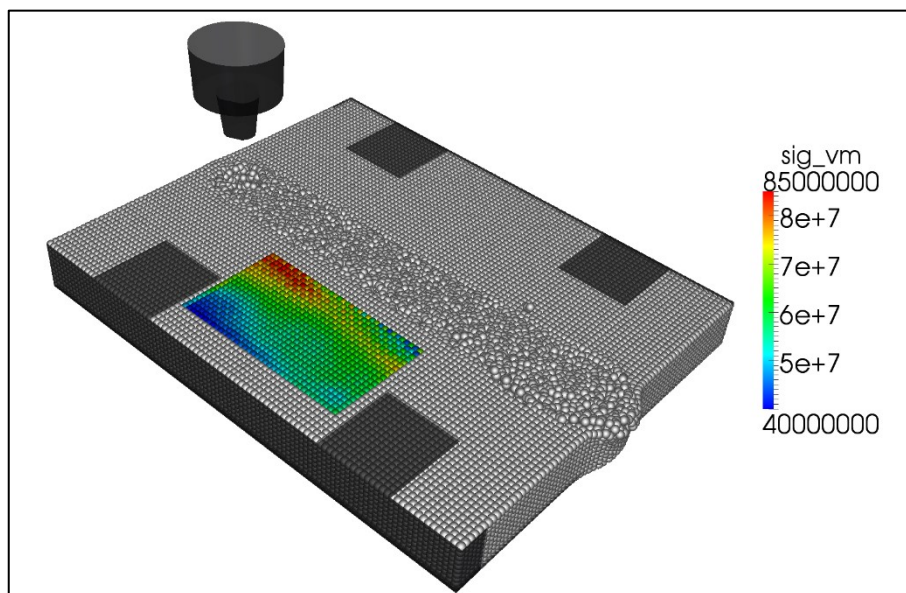
Figure 5-68 – Predicted residual stress from SPH model

The residual stresses cannot be measured in the weld zone because of the large plastic deformation and mixing present in the weld track. The residual stresses determined at the end of the SPH simulation are shown in Figure 5-68. The highest stresses are concentrated close to the weld zone as was found experimentally. The stress contours found experimentally and numerically are in close agreement. Any differences are minute and can be attributed to the complex nature of the heat loss through from the WPs to the supporting structure.

6 Process Parameter Optimization

Given the inherent complexity of the FSW process, the determination of optimal process parameters is an important field of research. Many researchers [151, 160, 293-298] have worked on determining optimal process parameters with varying goals. The general conclusion from these groups is that the rpm and advancing speed play the most important role in the final weld quality. To date, the weld quality and process parameters have not yet been optimized using a large deformation fully coupled thermal-mechanics simulation approach. Such an endeavor would have taken many weeks or even months with the current state of the art simulation methods.

In this chapter the developed meshfree code, SPHriction-3D, is used to optimize the FSW process for a 12.7 mm butt joint weld of AA6061-T6 plates. The model is as described in section 5.1. A total of 15 test cases are simulated. Some of the cases correspond to tests performed in the lab, while others are added to provide enough data sets to form meaningful response surfaces. The cases used in the optimization are summarized in Table 6-1. The weld type is defined as:

- Hot if the weld pitch is less than 0.4

- Medium if the weld pitch is between 0.4 and 0.85

- Cold if the weld pitch is greater than 0.85

Typically, the goal of other research groups has been to find the best possible weld strength and the associated rpm and advancing speed. In this study, a different tactic is employed; the optimization will be based on a series of weld quality metrics that are used to quantify different aspects of the joint. Metrics such as defect, mixing, maximum temperature, temperature variation, and tool wear will be used. In this sense, a company can determine optimal process parameters based on minimizing defects, maximizing the weld quality, or find the fastest possible advancing speed without the inclusion of a certain level of defects.

Table 6-1 – Optimization Study Process Parameters

Case Number	Exp. Number	Weld Pitch (mm/rev)	RPM	Advance (mm/min)	Weld Type
1-1	4-2	0.20	500	101.6	Hot
1-2	3-4	0.61	500	304.8	Medium
1-3	4-5	1.83	500	914.0	Cold
2-1	3-1	0.38	800	304.8	Hot
2-2	3-2	0.64	800	508.0	Medium
2-3	4-4	1.34	800	1069.0	Cold
3-1	4-1	0.18	1100	203.2	Hot
3-2	5-1	0.68	1100	813.0	Medium
3-3	4-3	0.92	1100	1016.0	Cold
4-1	N/A	0.34	1500	508	Hot
4-2	N/A	0.71	1500	1069	Medium
4-3	N/A	1.07	1500	1600	Cold
5-1	N/A	0.28	1800	508	Hot
5-2	N/A	0.59	1800	1069	Medium
5-3	N/A	1.11	1800	2000	Cold

Chapter 6 is organized into the following sections:

Section 6.1 will introduce a set of weld quality metrics that have been developed to evaluate various quality measures during the FSW simulations.

Section 6.2 presents the results from the 15 simulation models with respect to the quality metrics.

Section 6.3 constructs the response surfaces for the data sets. A response surface is created for each quality metric.

Section 6.4 will present and discuss the optimization results from the overall weld-quality response surface.

Section 6.5 will present and discuss the optimization results from the defect response surface.

Section 6.6 concentrates on the maximum possible advancing speed while maintaining an acceptable weld quality (constrained optimization).

6.1 WELD QUALITY METRICS

The metrics have been developed to assess the weld quality based on a number of different diagnostic methods during the advance phase only. The calculation of most of the metrics is inadmissible during the plunge and dwell phase. More importantly, the plunge/dwell region is commonly discarded since this is often a region of decreased strength.

In this sub-section, details on the significance of each of the metrics and the formulation used to assess their values in the models will be presented. Each of the metrics is tailored to highlight a certain diagnostic approach. The metrics are calculated at regular intervals while the simulations run. This provides an understanding of the evolution of the metric throughout the advance phase of the weld. Using this approach ensures an improved understanding on how and why the metric value was attained by examining various other results at the appropriate time step. For example, if the defect metric takes a drastic drop (more defect predicted), the point in time of the change can be investigated in the model to understand the specific reason for the sudden change.

All of the metrics are based on the measuring volume in the weld zone (TMAZ) as shown in blue and green in Figure 6-1. The blue part is on the retreating side and the green is on the advancing side. This approach to define the measuring volume will be critical for quantifying the level of mixing on each side of the center of the weld line.
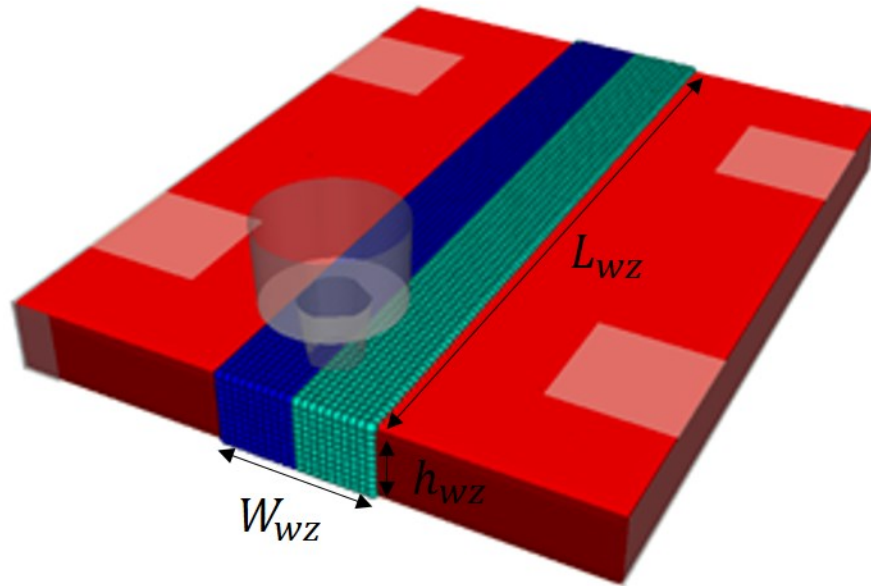
Figure 6-1 – Weld-quality measuring zone used for the metric calculations (shown in blue and green)

6.1.1   DEFECT METRIC

The ultimate goal of the project is to be able to find the probability of a surface or internal defect occurring, given a certain set of operating parameters. For this reason, the defect metric is of the greatest importance for this research work and is thus attributed the greatest weight over of all the other metrics.

The general concept is to calculate the total number of surface elements (found according to section 3.9.3). An example of the free surface elements for experiment number 4-5 is shown in Figure 6-2. Once the free surface elements have been identified, a surface triangulation algorithm (see section 4.6) is used to show the defect region in the model (Figure 6-3). The total number of surface elements is most readily found by looping over all the elements in the domain and incrementing a counter, $N_{surface\,nodes}$, each time a surface element is found. Next, the associated surface area, A_s, of the surface elements is calculated:

$$A_s = \Delta s^2 \, N_{surface\,nodes} \qquad\qquad (6\text{-}1)$$
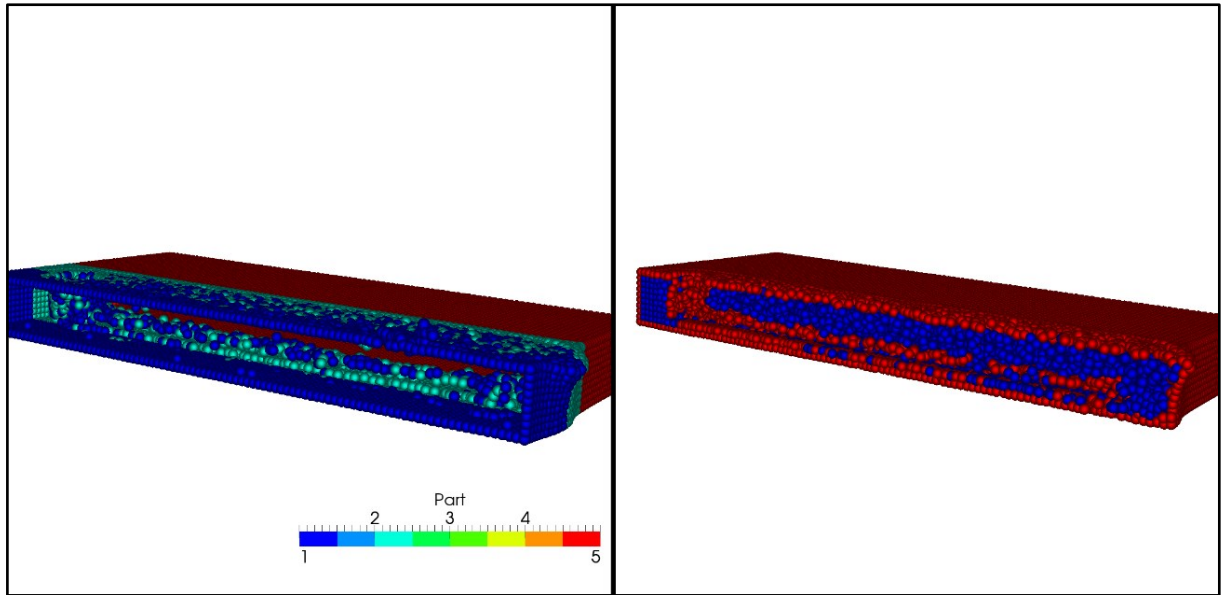
Figure 6-2 – Defect measured from the predicted free-surface elements
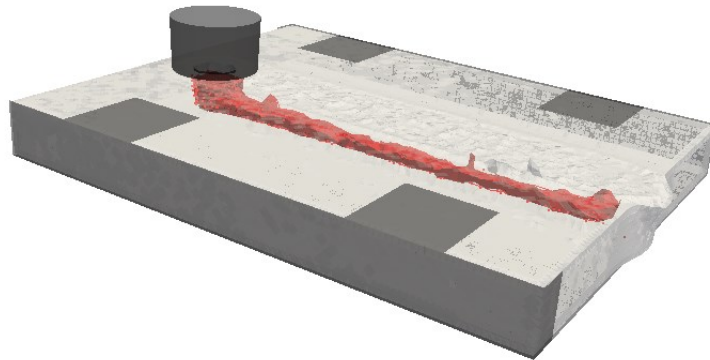
Figure 6-3 – Surface triangulation of the internal defects

Then the defect metric, ψ_{defect}, will be the ratio of the calculated surface area to the ideal, defect free surface area:

$$\psi_{defect} = 1 - \left|1 - A_s/A_{no\,defect}\right|$$ (6-2)

$A_{no\,defect}$ is found from the dimensions of the free surface of the measuring zone shown in Figure 6-1:

$$A_{no\ defect} = 2(L_{wz}\ W_{wz}) + 2(h_{wz}\ W_{wz}) + \pi\ r_{pin}\big(2\ l_{pin} + r_{pin}\big) \tag{6-3}$$

$\pi\ r_{pin}\big(2\ l_{pin} + r_{pin}\big)$ takes into account the surface area of the depression left by the tool pin during welding and upon retraction.

When ψ_{defect} takes on the value of 1, then $A_{surface} = A_{no\ defect}$, leading to the understanding that no defects are present in the finished weld. Practically speaking, this condition cannot be met since the surface area of the weld zone will always increase. This is due to the presence of the semi-circular striations as shown in Figure 6-4. In reality, a very low weld pitch will tend to cause a poor surface finish, and this can be accounted for by the increased bunching of SPH elements in the weld track. One would expect that the value of ψ_{defect} would start out close to 1 at the beginning of the advance phase, and decrease as the end of the advance phase nears. In theory, the defect metric should attain a minimum value at the end of the simulation when there is the possibility of the greatest amount of defects.

Before Weld Completion   **After Weld Completion**

Increased surface area in weld zone
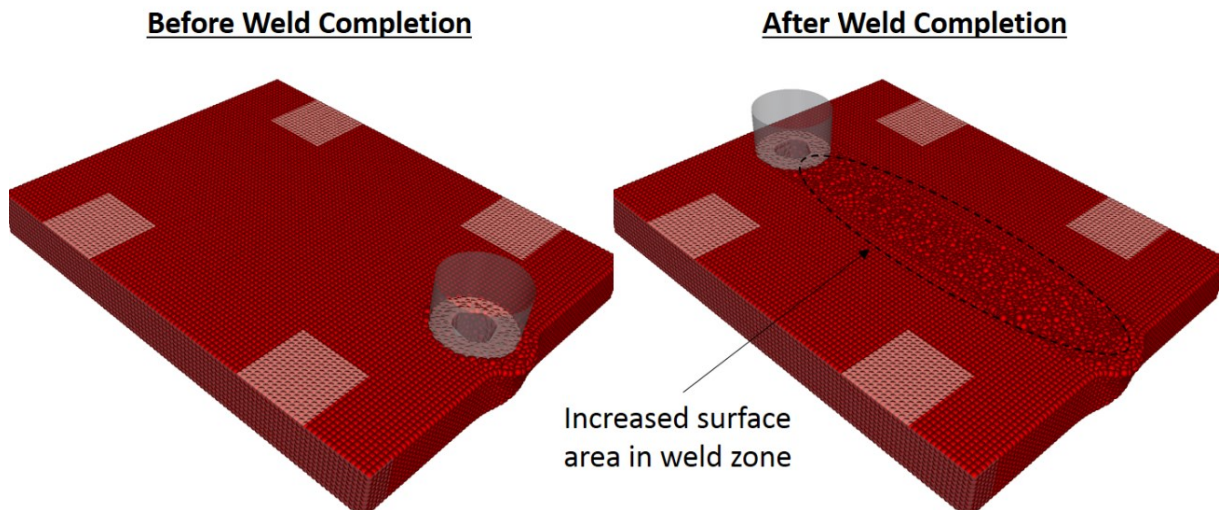
Figure 6-4 – Increased surface area in weld track

Another advantage to comparing the calculated surface area to a theoretical ideal surface area is that the surface area of the flash will also be included in the defect metric. Figure 6-5 shows the free surface result for case 1-1 (see Table 6-1), the image shows the prediction of extensive amounts of flash. The surface area of the flash is then included in the defect metric.
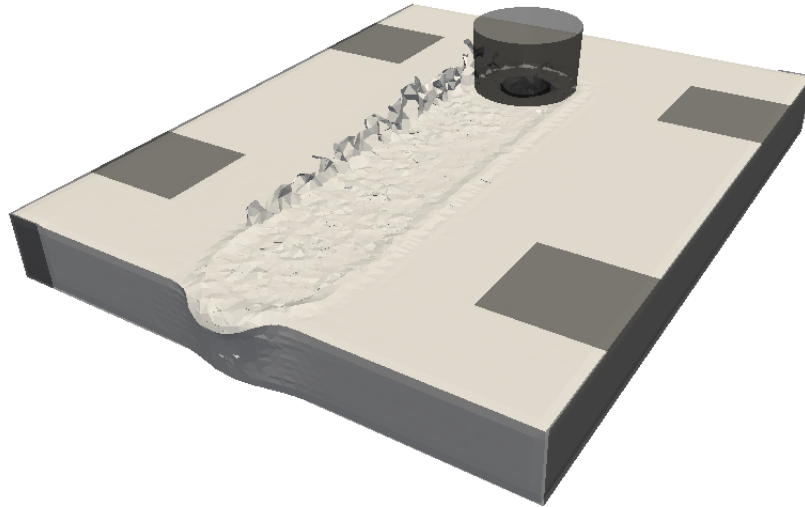
Figure 6-5 – Flash prediction contributing to defect metric

6.1.2 MIXING METRIC

Quantifying the level of mixing in the weld zone is another important diagnostic tool. Because of the Lagrangian nature of the SPH method, the number of elements belonging to each work piece can be calculated on each side of the center of the weld zone. Figure 6-6 shows an image of the level of mixing for experiment number 4-5 before (top) and after (bottom) the weld has been completed. The simulation is started with an even number of elements on each side of the weld centerline (shown as blue and red elements in Figure 6-6).

Typically, the advancing side has the lowest mixing ratio, particularly in the case that a defect exists. For the sake of being conservative, the advancing-side mixing ratio will be used. The mixing metric is found by counting the number of red (from work piece 1, N_{WP1}) and blue (from work piece 2, N_{WP2}) elements on the right side of the centerline.

The mixing metric, ψ_{mixing}, is then given by:

$$\psi_{mixing} = 1 - |1 - N_{WP2}/N_{WP1}| \qquad (6\text{-}4)$$

The metric is evaluated continuously throughout the simulation. The mixing level will understandably increase from zero at the start of the advance phase, to a maximum value at the end of the advance phase.
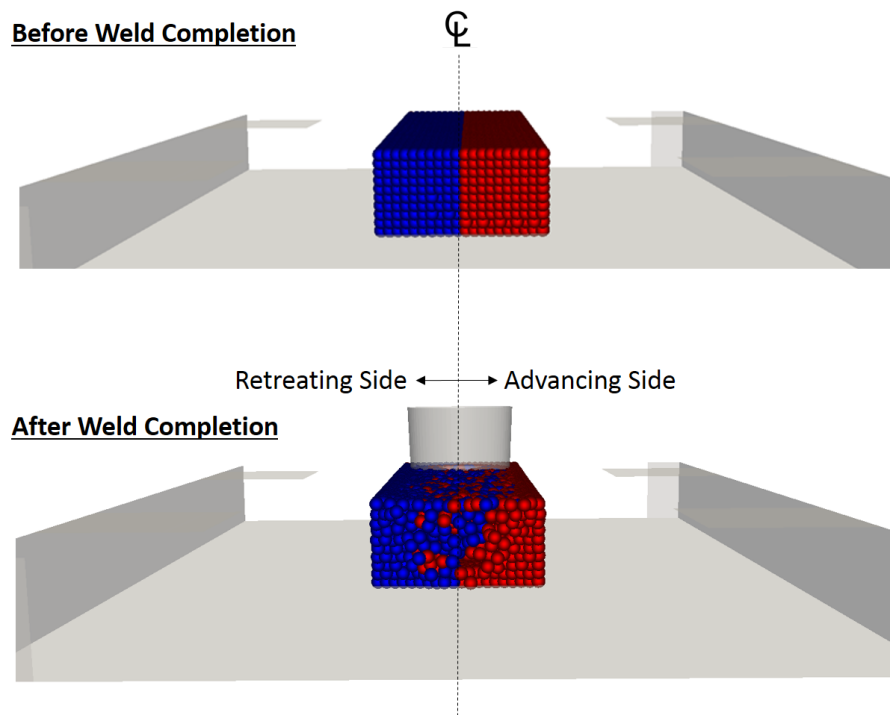
Figure 6-6 – Quantification of work piece mixing

A superior weld quality is attained when the metric approaches a value of one. This would represent an even mix of elements, so that $N_{WP1} = N_{WP2}$ on the advancing side of the weld. One can infer that if more of work piece 1 is found on the advancing side, then the weld will be stronger since there has been more potential for a strong mechanical bond.

Another possibility could be to use the average plastic strain in the weld zone. Figure 6-7 shows a typical cross section of the plastic strain contours throughout the thickness of the welded plates. Such a metric could be beneficial to ensure that a sufficient penetration has been obtained. In this work, a plastic strain metric was not used since the tool used in the experiments (and in the simulation as well) was cut to increase the level of defects. Unquestionably, a weld formed with a partial depth pin will result in a partial penetration weld. For this reason, using such a metric would be misleading and not provide an improved assessment of the weld quality. In the case that a design engineer is looking to find the optimal plunge depth, a plastic strain metric would be beneficial.
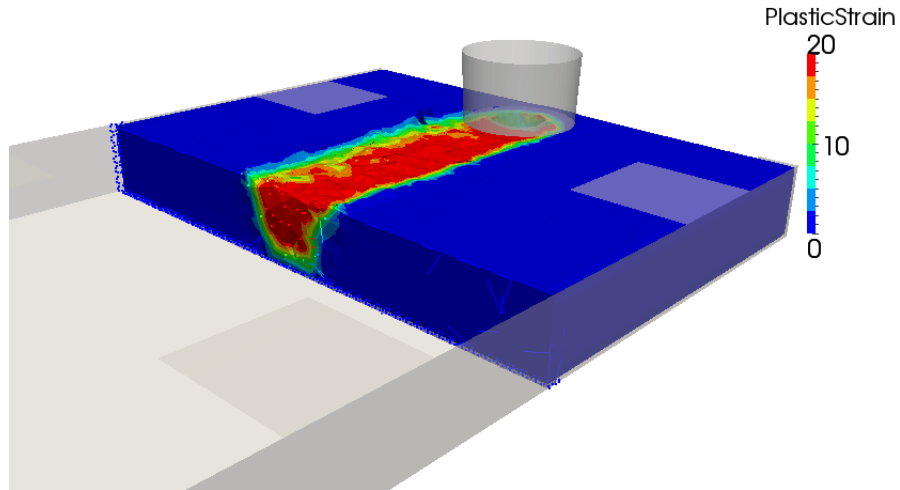
Figure 6-7 – Plastic strain throughout the thickness of the welded plates

6.1.3   MAXIMUM TEMPERATURE METRIC

The maximum temperature is an important criterion for a number of reasons. The size of the HAZ and ultimately the relative quality of the microstructure will be strongly related to the maximum temperature, T_{max}, attained during welding. Higher welding temperature can be associated with grain growth and a course (un-desirable) microstructure. In this work, the maximum temperature in the model is compared to the optimal temperature, T_{opt}, of 443 °C from the analytical model of Qian *et al.* [45].
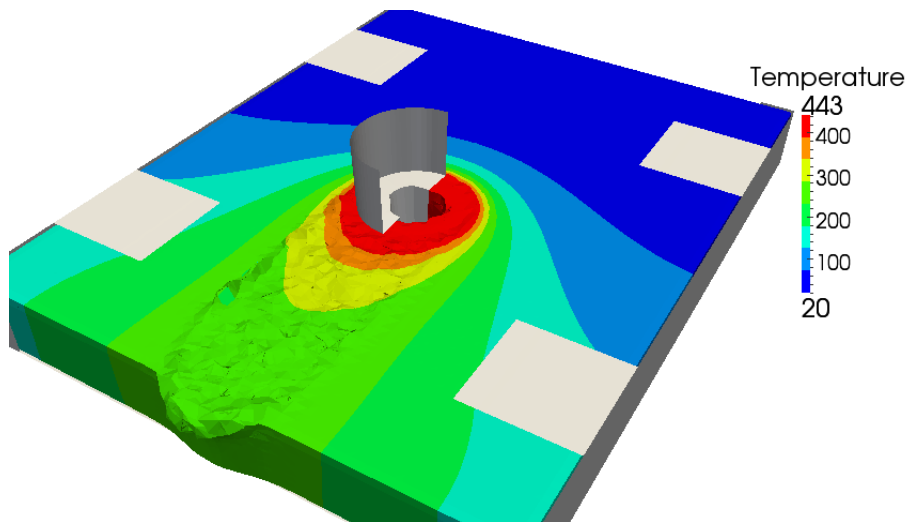
Figure 6-8 – Typical temperature contours

The value chosen for T_{opt} can easily be debated; in fact, in recent years, temperature measurement methods have improved at the tool-work piece interface. Schmale *et al.* [299] have recently shown that the maximum temperature is higher than was previously thought, suggesting that the optimal weld-zone welding temperature predicted by Qian *et al.* [45] is on the low side.

In any case, T_{opt} is a free parameter that the design engineer has the ability to change. The maximum temperature metric, ψ_{Tmax}, is then given by:

$$\psi_{Tmax} = 1 - \left| 1 - \frac{T_{max}}{T_{opt}} \right|, \qquad T_{room} \geq T_{max} \geq T_{melt} \tag{6-5}$$

For the cases considered in this work, the maximum temperature metric is bound between zero and one when T_{max} is between the room and melt temperature. Indeed, in the case of the FSW process, maximum temperatures outside this range are non-physical. The maximum temperature is evaluated at regular intervals by looping through all the work piece elements and keeping track of the highest temperature found. Without exception, the maximum temperature is always found at the interface between the tool and the work pieces as shown in Figure 6-8. Unlike ψ_{defect} and ψ_{mixing}, the value of ψ_{Tmax} will not change monotonically. Instead, it will oscillate throughout the process. When $T_{max} = T_{opt}$, the optimal conditions are met, and ψ_{Tmax} would be equal to 1.0.

6.1.4   MOVING THERMO-COUPLE VARIATION METRIC

The moving thermocouple (MTC1) is placed under the shoulder of the tool as shown in Figure 6-9. As the name implies, the TC follows the tool as it rotates and advances. In practice, this would be accomplished by drilling a hole through the tool on an angle. The TC would then be placed through the hole so that the measuring tip is exposed to the work piece at the shoulder interface. Schmale *et al.* [299] have used this technique with a high data acquisition frequency and were able to show the oscillating nature of the temperature in the weld zone.

A typical temperature history as measured by MTC1 will tend to be oscillatory. As the TC passes from the advancing side to the retreating side, it is expected that a temperature variation will be present. The extent of the variation can give important information regarding the symmetry of the TMAZ and HAZ

zones, microstructure, as well as the overall conditions for performing a sound weld. A common temperature history is shown in Figure 6-10.
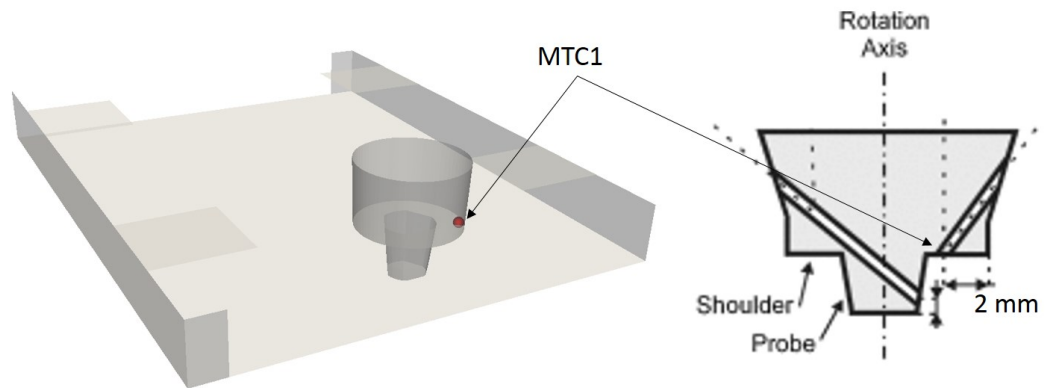
Figure 6-9 – Location of MTC1 and a through hole method for mounting the thermocouple (adapted from [299])
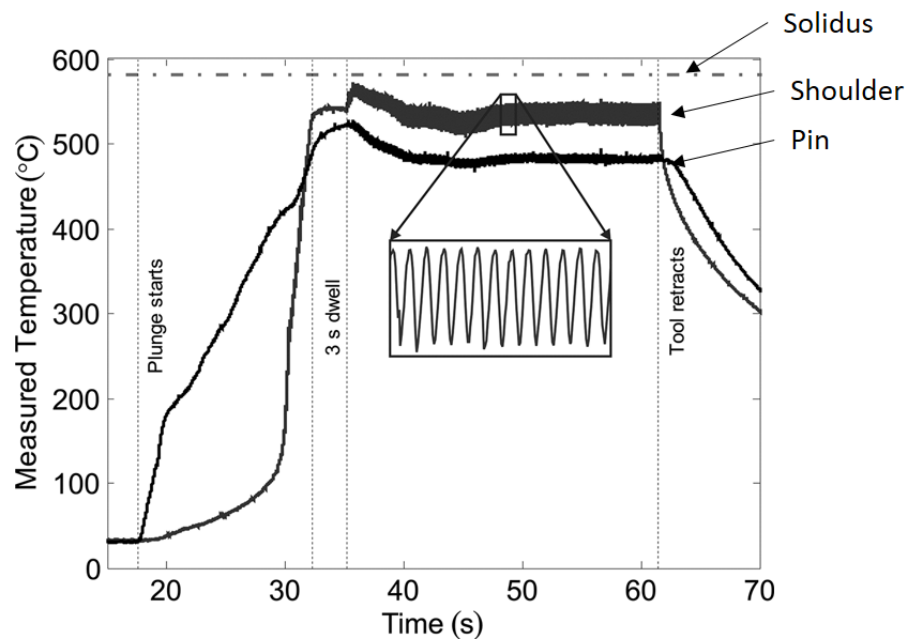
Figure 6-10 – MTC1 temperature history (adapted from [299])

The idea of this metric is to compare the maximum, $T_{var\,max}$, and minimum, $T_{var\,min}$, temperature values measured by MTC1 during the advance phase. The MTC variation metric is then given by:

$$\psi_{T_{var}} = 1 - \left|1 - T_{var\,min}/T_{var\,max}\right| \qquad (6\text{-}6)$$

Usually, the value of $\psi_{T_{var}}$ will start very close to 1.0 ($T_{var\,min} = T_{var\,max}$) at the start of the advance phase. This is because the temperature is commonly very symmetric during the plunge and dwell phase. However, as the tool starts to advance, the ratio $T_{var\,min}/T_{var\,max}$ will start to increase, ultimately causing $\psi_{T_{var}}$ to decrease.

6.1.5 TOOL WEAR METRIC

Although tool wear does not provide a direct measurement of the weld quality, it is an important consideration since the cost of a common FSW tool is significant. If tool wear were not considered, the design engineer could end up picking process parameters that could result in the need to change the tool too often, burning up any perceived gain by optimizing the advance speed. A common wear pattern in a hardened steel tool used to weld aluminum alloys is shown in Figure 6-11 [300].
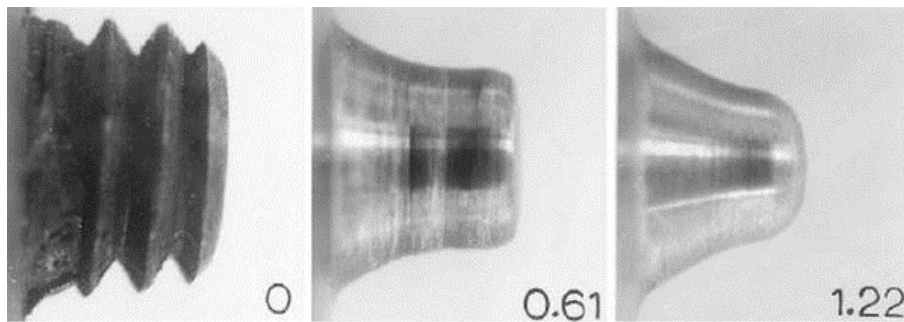
Figure 6-11 – Common tool wear behaviour [300]

The wear metric is based on the developments presented in section 3.14. The wear depth rate, $\dot{d}_{wear}$, is continuously calculated throughout the simulation from equation (3-172), and the wear depth, d_{wear}, is found by explicitly integrateing the wear depth in time. The metric will then compare the worn to the initial tool volume. For the case under consideration in this work, the tool volume is found numerically (from SPHriction-3D) to be V_{Tool} = 5.4x10^{-6} m³. The wear volume is determined by taking the sum of all the elemental wear depths multiplied by the surface area of the respective element:

$$V_{wear} = \sum_{j \in FEM} \frac{1}{2} \left| \bar{x}_{A_j} \times \bar{x}_{B_j} \right| d_{wear_j} \tag{6-7}$$

Then the wear metric, ψ_{wear}, can be stated as:

$$\psi_{wear} = 1 - V_{wear}/V_{Tool} \tag{6-8}$$

Upon inspection of equation (6-8), one can see that the wear metric will monotonically decrease (since d_{wear} and by association V_{wear} are always increasing). In this sense, ψ_{wear} will start close to 1.0 at the beginning of the advance phase and reach a minimum value at the end of the simulation.

6.1.6 OVERALL WELD QUALITY

The overall weld quality will be evaluated based on the various metric weights shown in Table 6-2. Certainly, the value of the weights is expected to be a point of dispute. In general, the weights would be dependent on the design engineer's main goal. It the engineer is more concerned with having a TMAZ that is as symmetric as possible, they would likely put more weight towards the "moving TC variation" ($\xi_{T_{var}}$). In another situation, perhaps the engineer would prefer to minimize the defects and tool wear so that in this case $\xi_{defect} = \xi_{wear} = 0.5$. In any case, the proposed method is parametric and can easily be adjusted to suit the specific company's needs. The weighting values chosen for this work represent a desire to have an all-around good quality weld.

Table 6-2 – Weld Quality Metrics

Quality Metric	Weight
Defects, ξ_{defect}	0.4
Mixing, ξ_{mixing}	0.25
Maximum Temperature, $\xi_{T_{max}}$	0.1
Moving TC variation, $\xi_{T_{var}}$	0.15
Tool wear, ξ_{wear}	0.1
Total	1.0

So that the overall weld quality can be found from:

$$\psi_{weld\ quality} = \xi_{defect}\ \psi_{defect} + \xi_{mixing}\ \psi_{mixing} + \xi_{T\ max}\ \psi_{T\ max} + \xi_{T\ var}\ \psi_{T\ var}$$
$$+ \xi_{wear}\ \psi_{wear}$$

<div align="right">(6-9)</div>

The overall weld quality will figure between 0.0 and 1.0:

$$0.0 \leq \psi_{weld\ quality} \leq 1.0$$

<div align="right">(6-10)</div>

6.2   WELD QUALITY RESULTS

Case numbers 1-1 through to 5-3 (15 cases, see Table 6-1), were run in SPHriction-3D. The results from the different cases for the different metrics will be presented in this section. The three advancing speeds per rpm value were chosen to provide "hot", "warm", and "cold" weld cases. The "hot" cases tend to result in a poor surface finish (increase in the number of SPH elements on the surface in the weld track).

6.2.1   DEFECTS

The defect metric results were compiled following the completion of the 15 process parameter simulations. A comparison of the final defect metric (found at the end of the advancing phase) is shown in Figure 6-12.

Defect Metric Comparison

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
Hot Weld	0.78	0.88	0.83	0.84	0.84
Medium Weld	0.86	0.86	0.88	0.87	0.84
Cold Weld	0.77	0.83	0.87	0.85	0.84
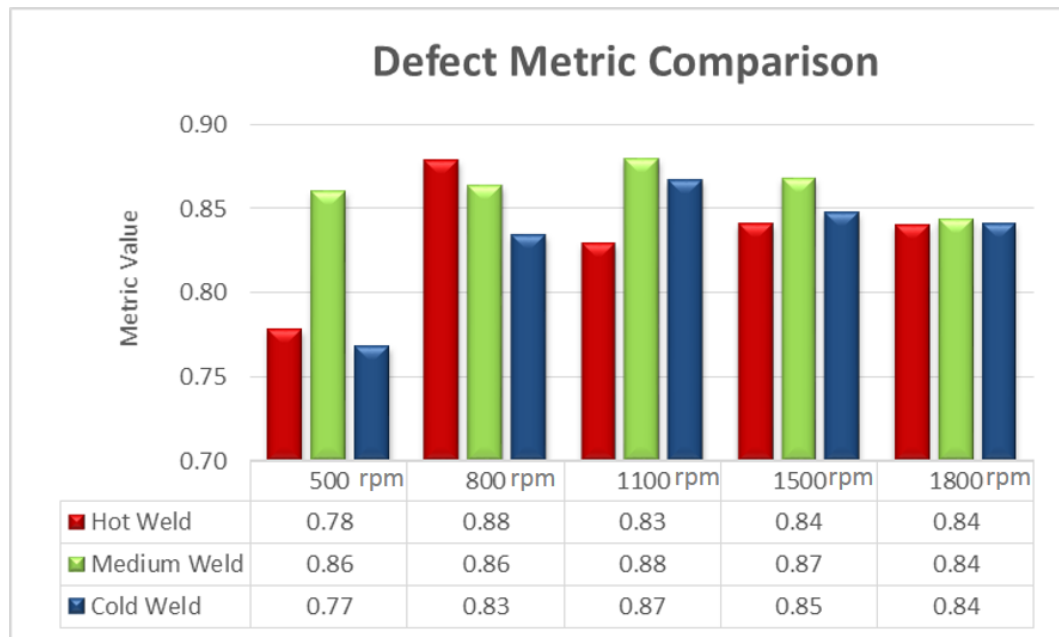
Figure 6-12 – Defect metric comparison at end of advance phase

One of the particularities of using a position controlled CNC machine is that the tool will tend to "dig in" more as the temperature in the weld zone increases. This behaviour was noticed both in the experimental and simulation results. As the tool would dig in further, more material would be expelled past the shoulder of the tool on the retreating side. This leads to an increased level of internal (void) defects since the material is expelled. This effect is compared in Figure 6-13; notice that the predicted flash height is significantly larger in the "hot" weld (500 rpm, 102 mm/min) compared to the "warm" weld (800 rpm, 508 mm/min). This can be explained by the increased amount of thermal expansion associated with the higher weld temperature.
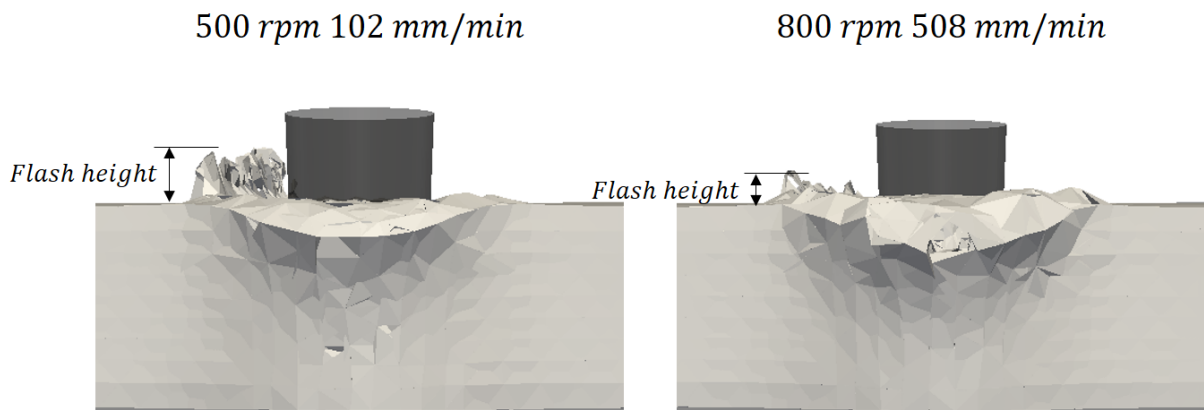
Figure 6-13 – Flash height for 500 rpm with 102 mm/min and 800 rpm with 508 mm/min

The predicted internal defects are shown for the complete set of the 15 tests in Figure 6-14. The results are organized from left to right by increasing advancing speed and from top to bottom by increasing rpm. In general, the "warm" welds are in the center column. These welds typically show the least amount of internal defect as well as the least flash.
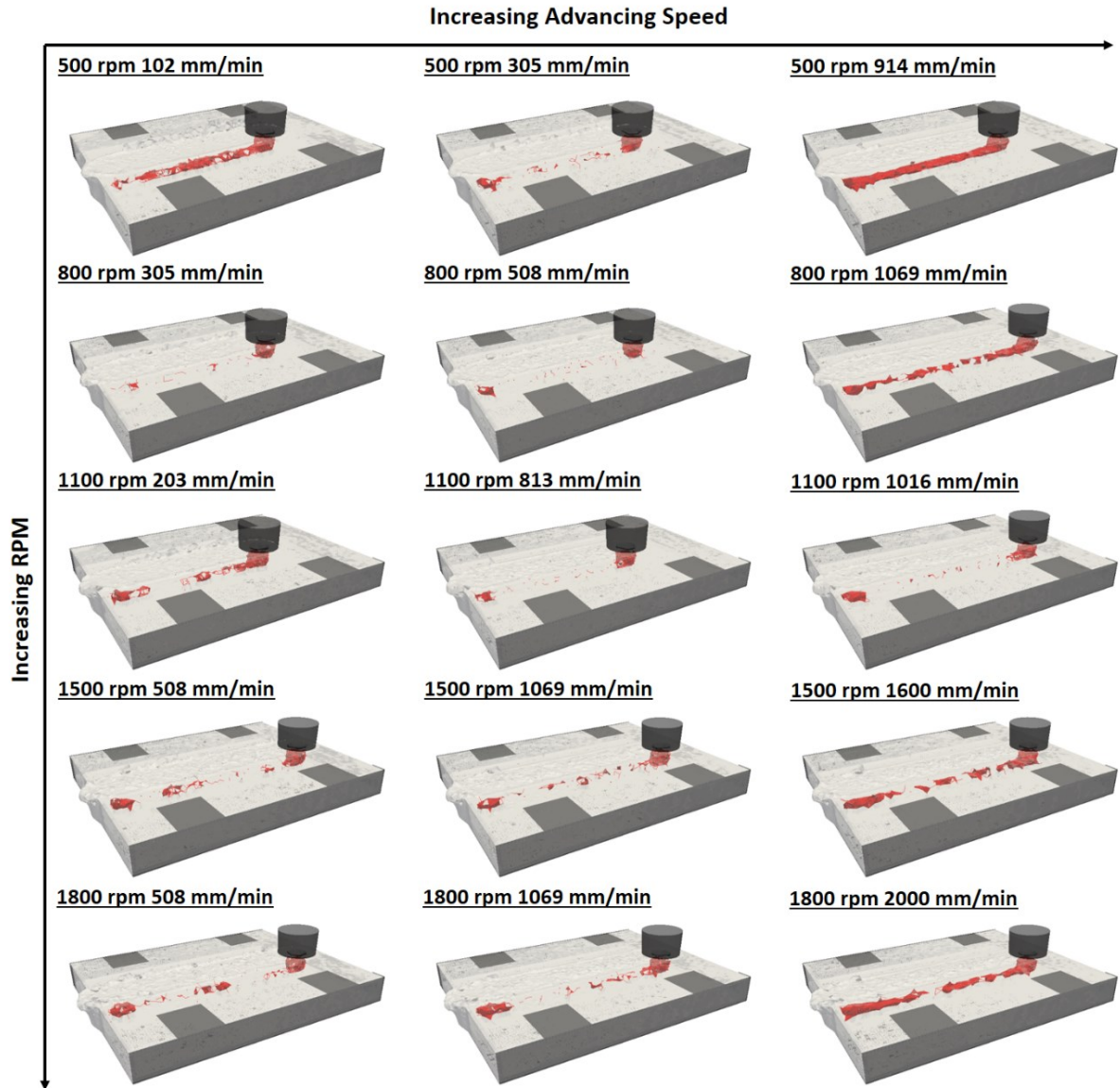
Figure 6-14 - Defect results for the cases considered in the optimization analysis

The simulation results show that a weld pitch of ~0.4 to ~0.6 revolutions/mm leads to the lowest probability of forming internal defects. Values above this range tends to introduce significant internal defects, whereas values below show a prominent flash formation and a reduction in the surface finish. The instantaneous defect metric for the 15 cases is shown in Figure 6-15. The results are intuitive; as the weld pitch increases, so does the defect metric.
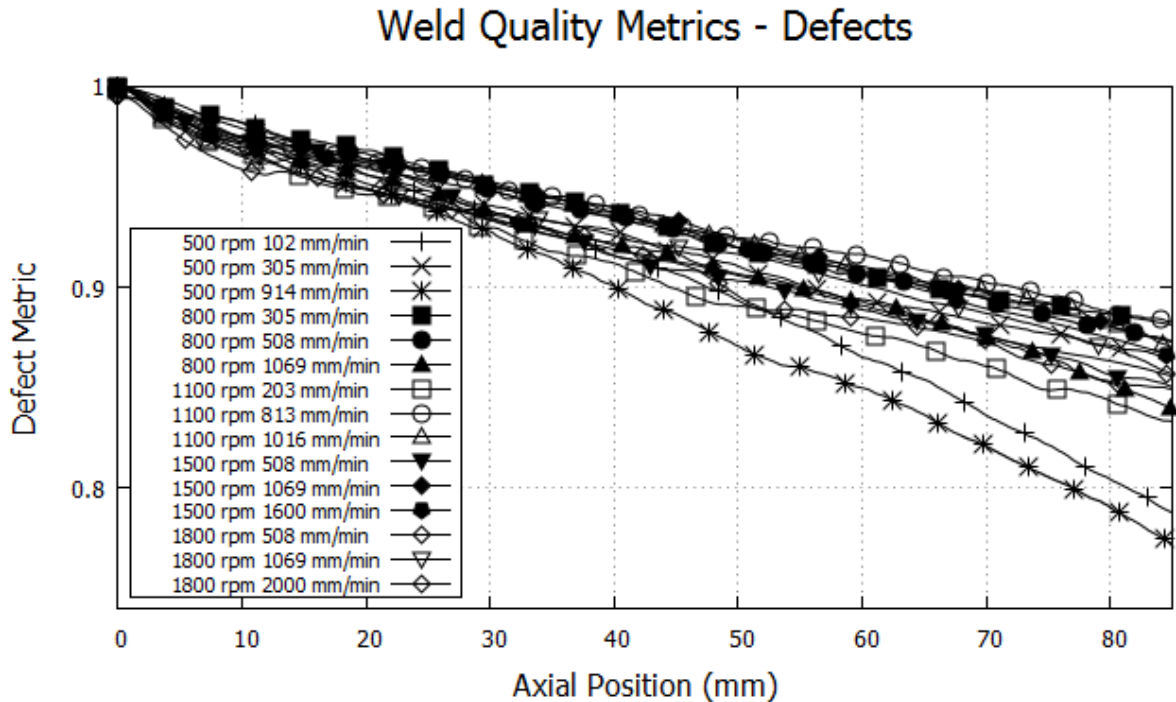
Figure 6-15 – Transient defect metrics comparison for optimization cases

6.2.2   EXPERIMENTAL VALIDATION OF SIMULATED DEFECTS

In order to validate the predicted defects by the simulation code, a series of welds were created in the lab as was described in section 5.1. The defects present in the welds were then evaluated by either cutting the plate in half along the center of the weld line or by using three-dimensional computed tomography (3D x-ray). The 3D x-ray images were performed by the Saguenay branch of CNRC-NRC (Aluminum Technology Center, CTA). A calibration block with internal voids was used by the x-ray technician in order to determine the settings for the x-ray equipment. The parameters used for the x-ray study are listed in the appendices (section 9). The defect results from the experimental welds will be presented in this section and compared to that predicted by the simulation code.

Figure 6-16 shows the defects found within the experimental welds (upper set of images in each row) compared to the simulation results (lower set of images in each row). For the 800 rpm with 305 and 1069 mm/min, the defects are exposed by cutting through the center of the weld. This is accomplished using a band saw with a 1.2 mm thick blade. This approach results in the removal of at least 1.2 mm of material since it is a destructive approach. However, it is suspected that slightly more material is removed, closer

Kirk Fraser - April 2017

to 1.5 mm. The remaining defects are evaluated using the x-ray technology. This approach is non-destructive and provides an excellent understanding of the defects within the weld. One can see that that the simulation code provides an excellent prediction of the location, size, and shape of the internal defects. The surface areas associated with the defects are summarized in Table 6-3.

Table 6-3 – Comparison of defect surface area between experiment and simulation

RPM [rev/min]	Va [mm/min]	Experiment Defect Surface [mm²]	Simulation Defect Surface [mm²]	% Error
500	102	3492	2323	33.3
500	305	1363	1463	7.3
500	914	3729	2424	35.0
800	305	412	736	78.6
800	508	522	915	75.3
800	1069	1186	1249	5.3
1100	203	2963	1784	39.8
1100	813	1764	1261	28.5
1100	1016	2073	1387	33.0

In general, the x-ray results predict a greater surface area than the simulation code. There are a number of reasons for this. First, the defect detection algorithm available for the x-ray images iteratively passes through the sets of x-ray images; when the program finds a region of decreased density, it flags the region as a possible defect. Following that, the program investigates more closely the suspected regions and calculates the probability that the region is a defect and associates an equivalent radius to the defect. For the sake of consistency, a filter was used to exclude defects that had a probability of less than 0.1. This criterion may have been too low, leading to the inclusion of regions that may not have actually been defects (leading to an increase in the calculated surface area). In addition, the fact that the defects are

approximated as an agglomeration of spheres by the x-ray equipment would tend to over predict the defect surface area. This approximation leads to overlapping spheres, which would overestimate the surface area.
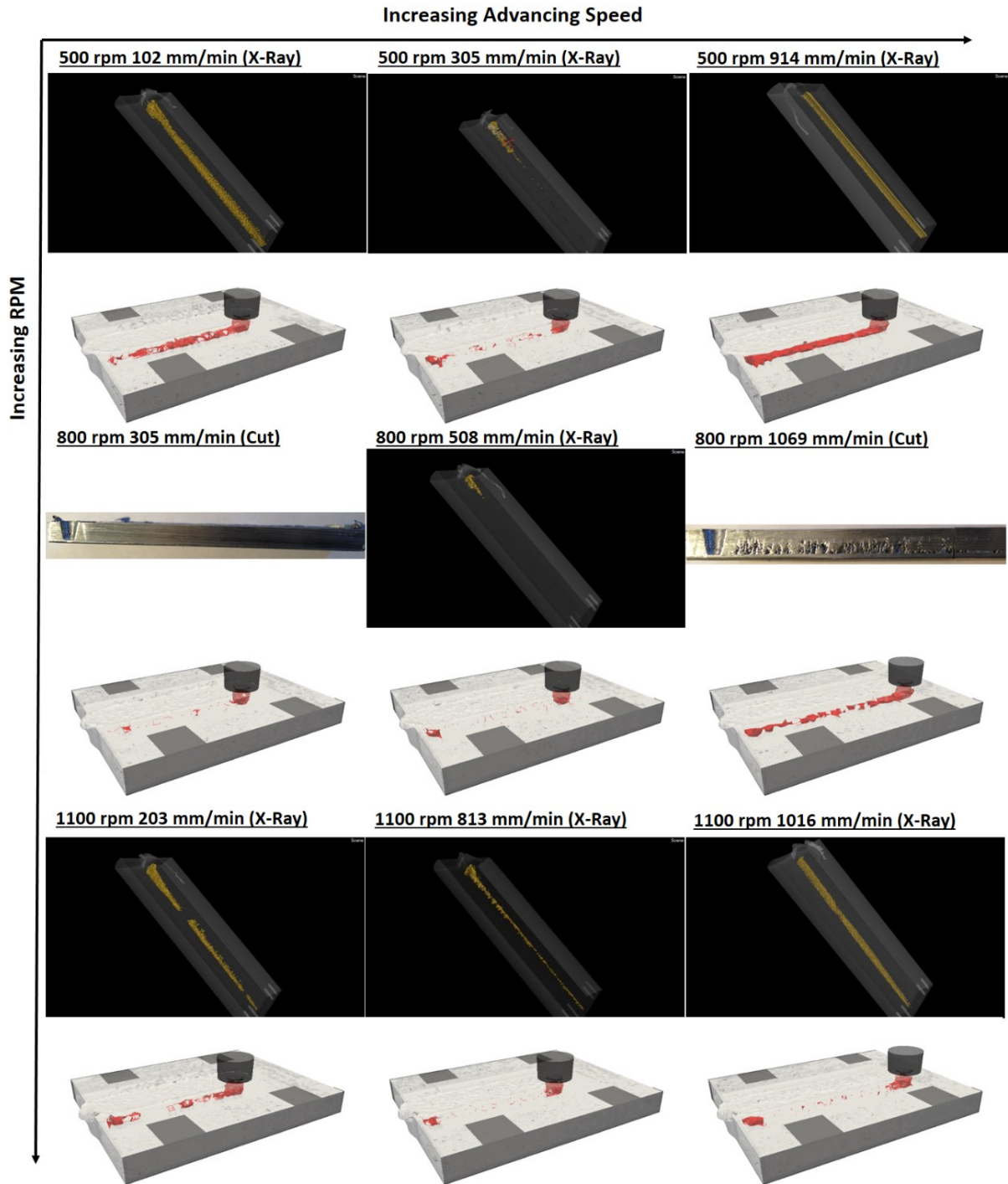
Figure 6-16 – Experimental defect results

The x-ray equipment is capable of detecting defects on the order of a micron, whereas the simulation has a defect prediction precision of ~1 mm (related to the particle grid spacing). For this reason, only the x-ray results showing a detected defect diameter of at least 1.0 mm were included in the surface area calculation (along with the 0.1 probability criteria).

The percent error between the predicted defects and the measured defects is shown in the last column of Table 6-3. Overall, the error is in the range of 5% to 79%. More work is needed to investigate the possibility of wrapping a volume around the x-ray results to get a more precise surface area measurement. An algorithm should be developed to sort through the defects found and ensure that there are no overlapping spheres are present. This could be accomplished using the position of the centroid and the diameter of each sphere. Once a set of exclusive, non-overlapping spheres is found, a volume should be bound that includes the spheres. Then the surface area of this bounding volume should be used to compare to the simulation results. The percent error is expected to decrease overall once such an approach is adopted.

Of particular importance is that the experimental results follow the general trend as predicted by the simulation models. This observation can best be qualified by normalizing the individual defect surface areas with respect to the largest (the 500 rpm 914 mm/min case). By using this approach, the general trend becomes evident, showing that the simulation model provides an excellent general prediction of the internal defects within the weld. Figure 6-17 shows a graph comparing the normalized defect surface area for the experiment (blue) and the simulation (red).
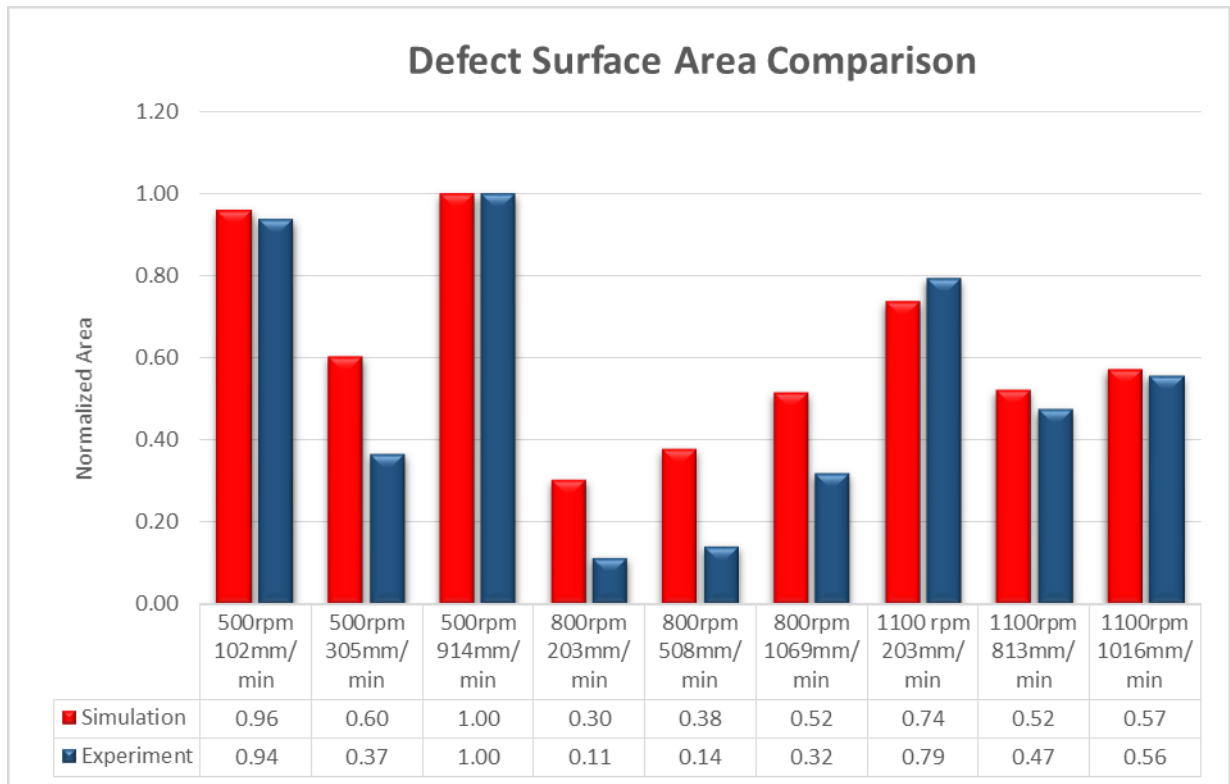
Figure 6-17 – Defect surface area comparison (normalized)

6.2.3 MIXING

The mixing metric has a major impact on the weld quality. Since the invention of the FSW process, many research groups [301-303] have strived to quantify the actual level of mixing in the weld zone. By using a Lagrangian hydrocode, the mixing results can be directly established. A comparison of the results of the mixing metric at the end of the advancing phase for the different cases is provided in Figure 6-18.

Mixing Metric Comparison

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
■ Hot Weld	0.63	0.75	0.70	0.88	0.91
■ Medium Weld	0.71	0.85	0.85	0.84	0.84
■ Cold Weld	0.73	0.77	0.81	0.78	0.80
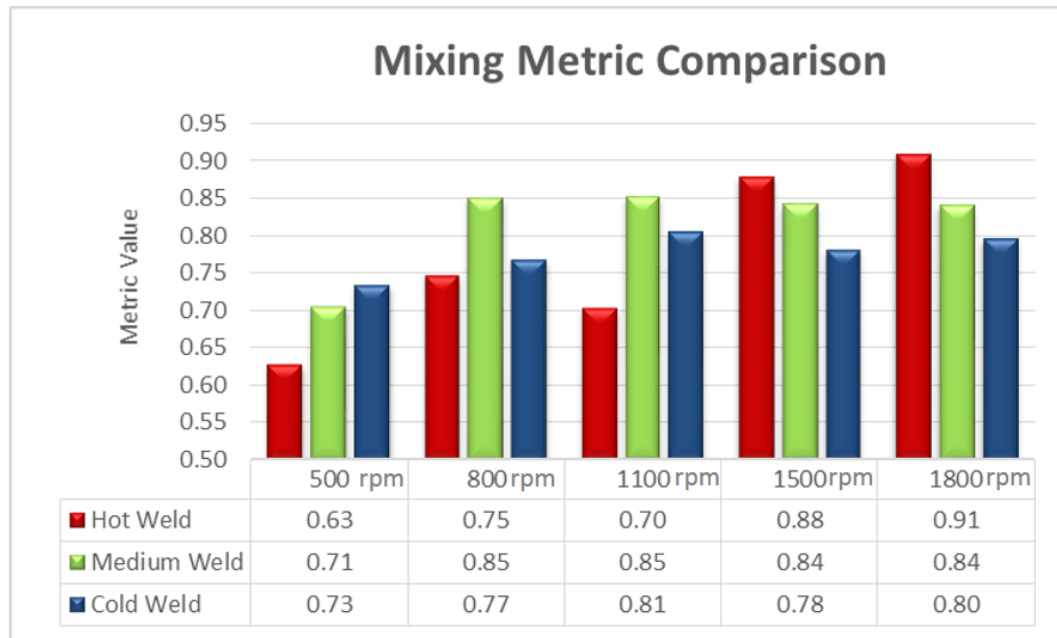
Figure 6-18 – Mixing metric comparison at end of advance phase

The results (Figure 6-19) show that the mixing is improved when the weld pitch is lowest and there is not a significant level of internal defects. Because of the nature of how the mixing metric is evaluated, the result also takes into consideration the presence of voids in the weld zone. This is because the number of particles is counted on the advancing side that were originally associated with the retreating (blue particles) and advancing (green particles) side plates. In the case of a defect being present in the weld zone, the ratio of blue to green particles present on the advancing side will decrease. Furthermore, the void defect will have a tendency to form in the lower region of the advancing side. A comparison of the mixing results for the full set of cases is shown in Figure 6-19. This image helps to clarify the notion that the presence of a defect will decrease the value of the mixing metric.
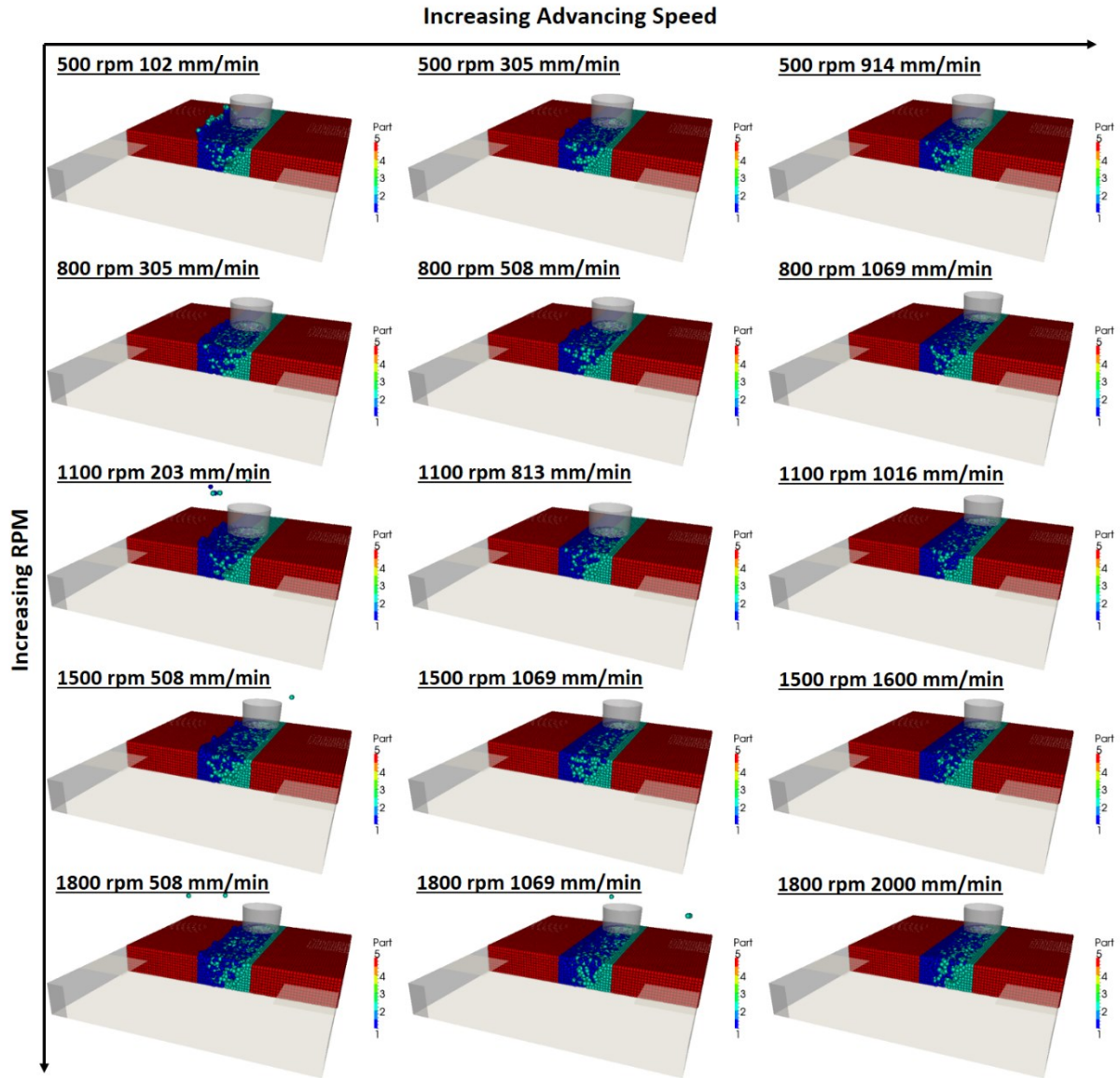
Figure 6-19 – Mixing results for the cases considered in the optimization analysis

Although a plastic strain metric was not used in this work, a comparison of the plastic strain contours through the thickness of the welded plates is shown in Figure 6-21. The comparison is provided to offer the reader a comprehension of how the process parameters affect the plastic strain state. The plastic strain results also deliver great insight into the extent of the TMAZ zone; the greater the extent of the plastic strain, the greater the size of the TMAZ zone (see Figure 6-20). More importantly, the results provide an understanding of the whether the transition from the HAZ to the TMAZ zone is abrupt or gradual. If the plastic strain gradient is large, then the transition is expected to be abrupt.
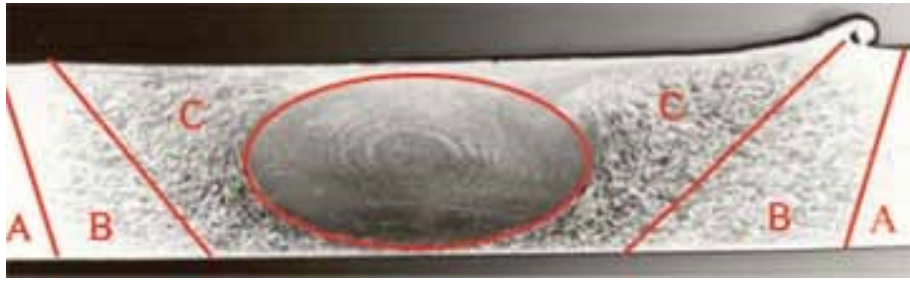
Figure 6-20 –Microstructure zones: A – parent material, B – HAZ, and C – TMAZ (from [304])
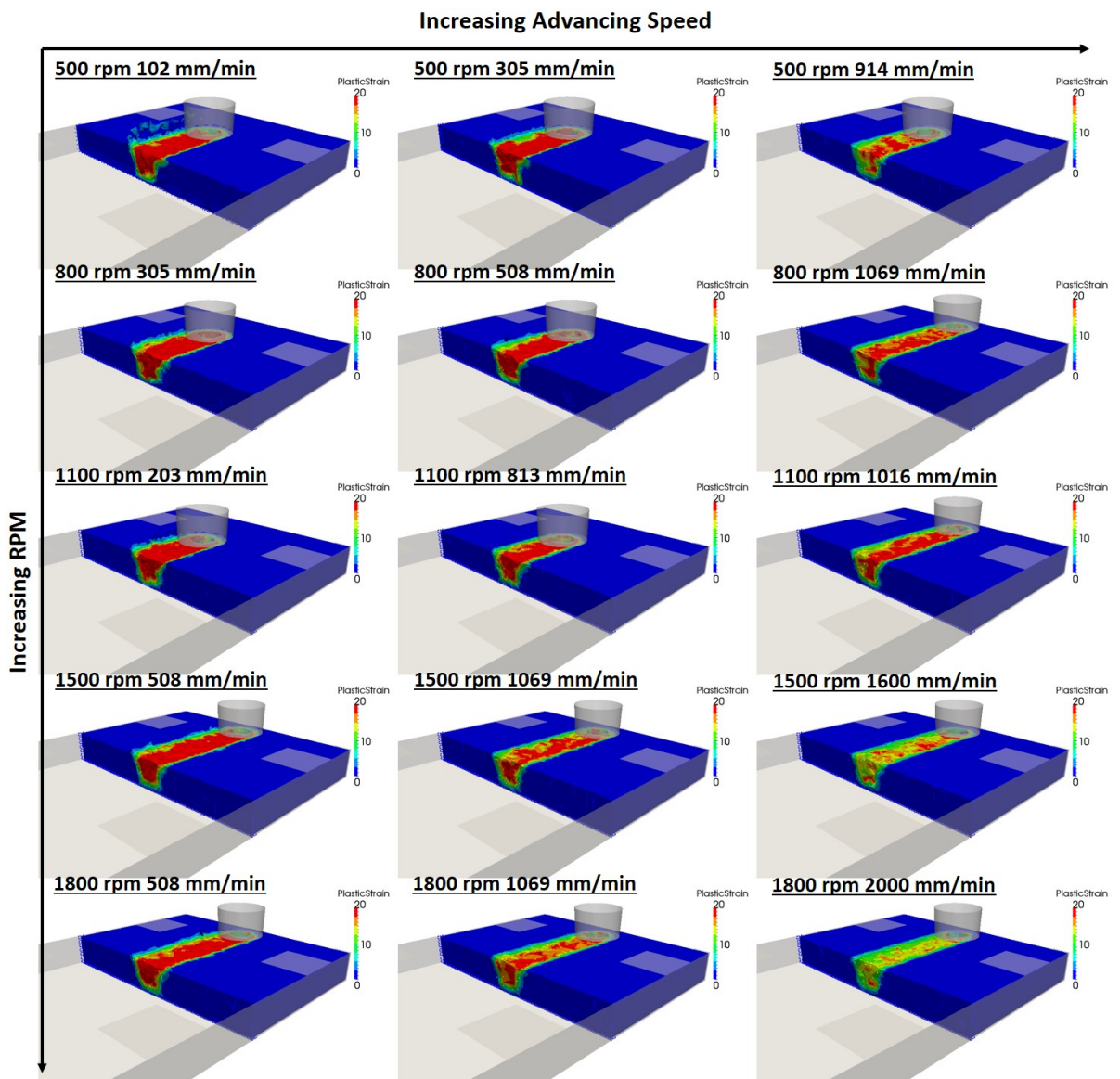
Figure 6-21 – Plastic strain results for the cases considered in the optimization analysis

The transient results for the mixing metric are shown in Figure 6-22. As expected, the mixing metric monotonically increases from a minimum value at the start of the advance phase.
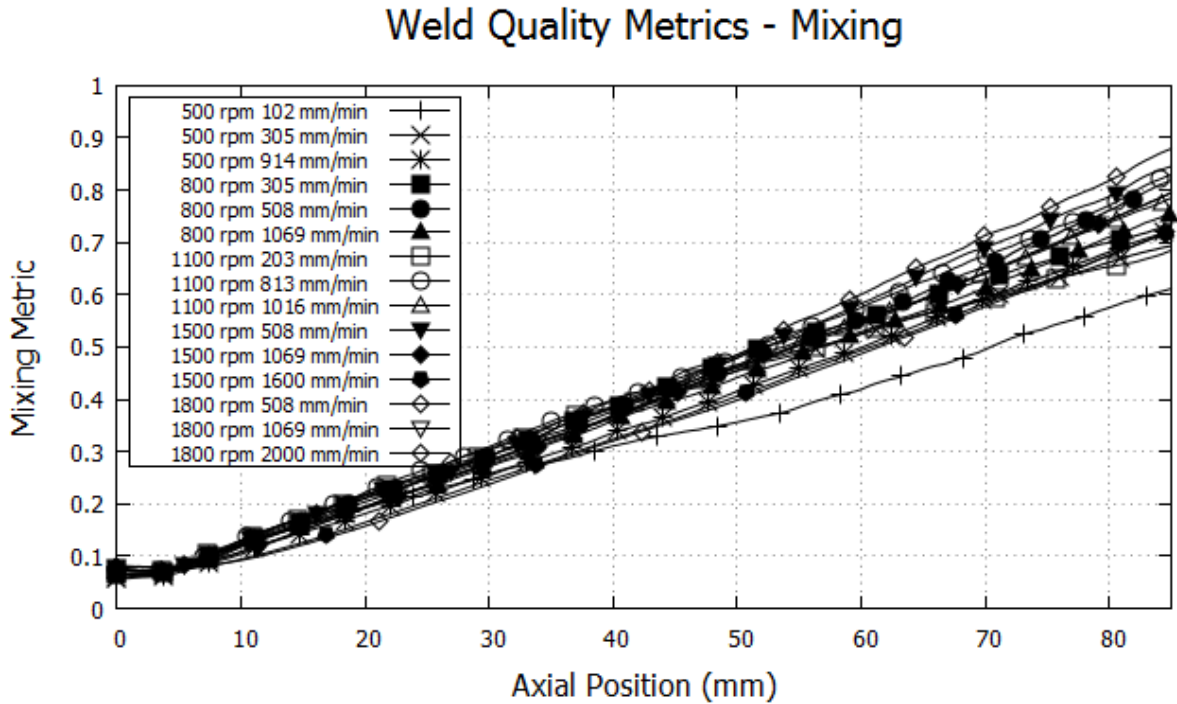
Figure 6-22 - Transient mixing metrics comparison for optimization cases

6.2.4   MAX TEMPERATURE AND MTC1 VARIATION

The maximum temperature and moving thermocouple (MTC1) variation metrics are designed to provide insight into the microstructure transformation in the weld zone. Excessively high weld zone temperatures can lead to unwanted grain growth. In addition, strong temperature variations between the advancing and retreating side can lead to a highly un-symmetrical microstructure. The maximum temperature and MTC1 variation metrics are compiled from the end of the advance phase and compared for the different cases in Figure 6-23 and Figure 6-26 respectively. The higher rpm cases show the lowest maximum temperature metric values. The weld temperature was very close to the solidus point for these cases. The overall length of the work pieces also plays an important role. Because the actual weld length is 4 inches, the high temperature attained during the plunge and dwell phase strongly affects the advancing phase temperature.

Tmax Metric Comparison

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
Hot Weld	0.84	0.75	0.70	0.69	0.69
Medium Weld	0.86	0.77	0.70	0.69	0.69
Cold Weld	0.87	0.79	0.72	0.69	0.69
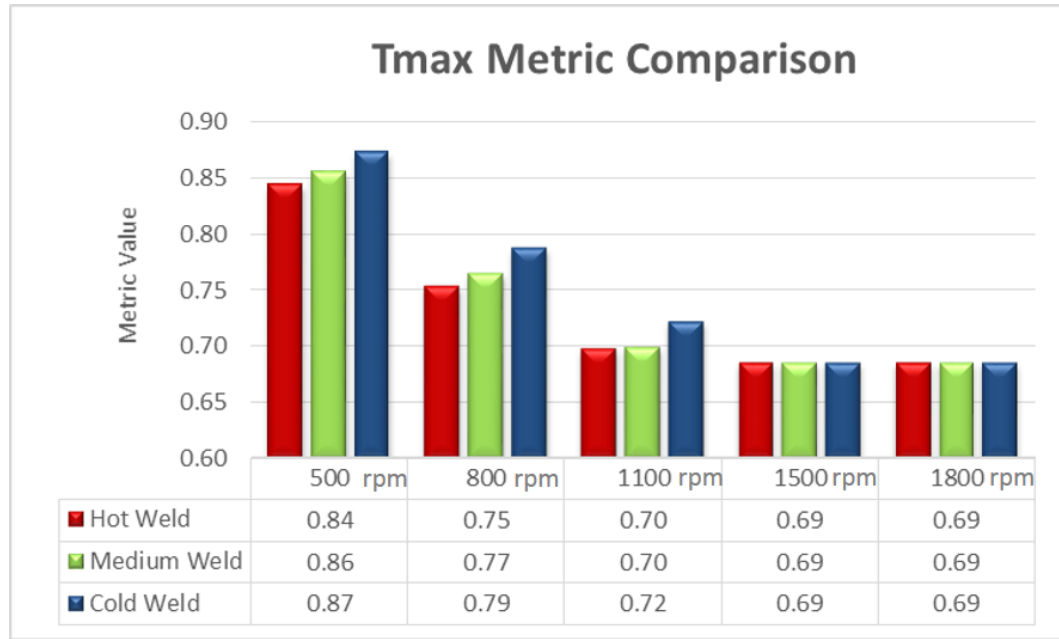
Figure 6-23 – Maximum temperature metric comparison at end of advance phase

A comparison of the surface temperature for the 15 different cases is provided in Figure 6-24. The image clearly shows the effect of the different process parameters. In general (and intuitively), the average temperature throughout the work pieces is higher for the low weld pitch cases ("hot" welds). In fact, by the end of the weld for the 500 rpm and 102 mm/min case, the lowest temperature in the plates is 182 °C, compared to close to room temperature for the 500 rpm and 914 mm/min case.

The instantaneous metric results for the maximum temperature and the MTC1 variation are shown in Figure 6-25 and Figure 6-27 respectively. Notice that the MTC1 results all start at a value of 1, (at the end of the dwell phase/start of advancing phase) meaning that the temperature is initially uniform throughout full tool rotation. As the tool starts to advance, the temperature variation increases. This effect is well explained by the flow of the material predicted in the simulation models. As the tool advances, hot material is transported and deposited on the advancing side, causing the temperature to be higher there.
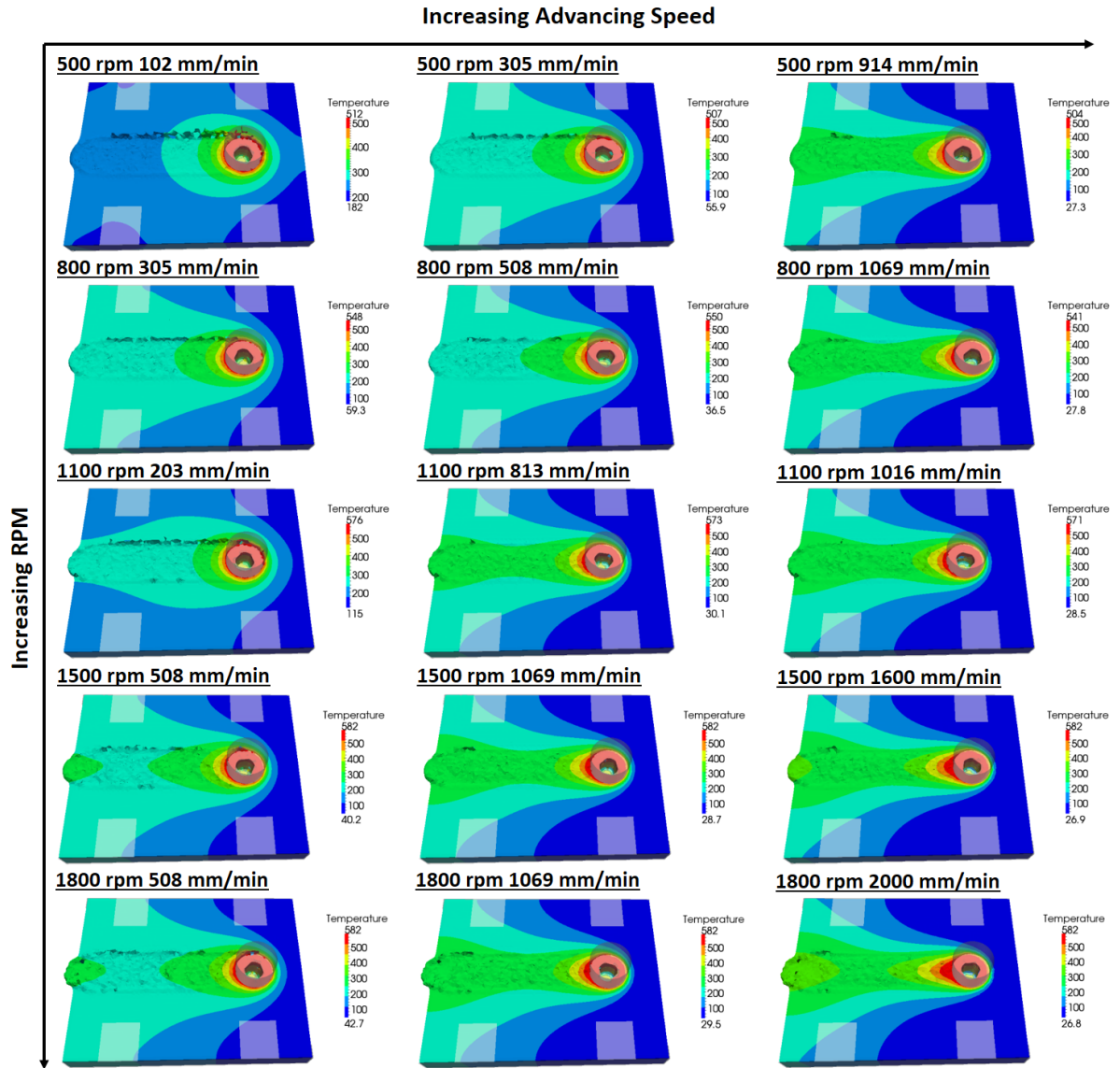
Figure 6-24 – Temperature contours for the cases considered in the optimization analysis
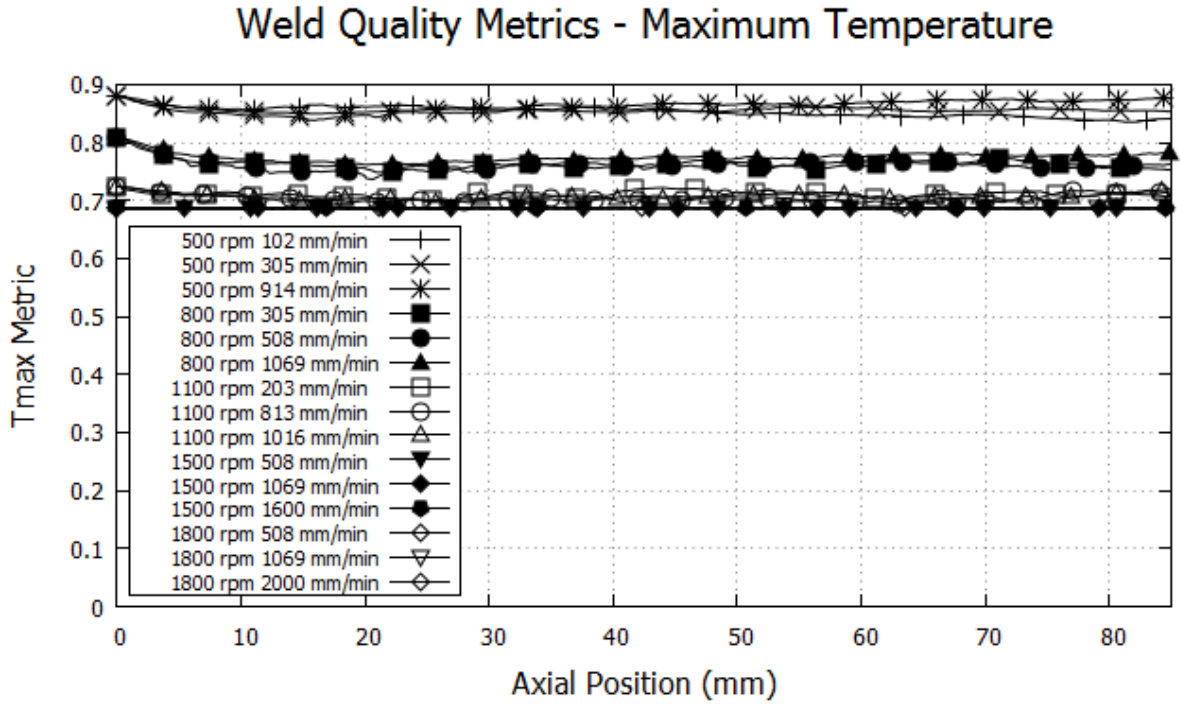
Figure 6-25 – Transient maximum temperature metrics comparison for optimization cases

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
■ Hot Weld	0.92	0.92	0.93	0.91	0.92
■ Medium Weld	0.92	0.90	0.88	0.87	0.87
■ Cold Weld	0.86	0.85	0.86	0.83	0.83
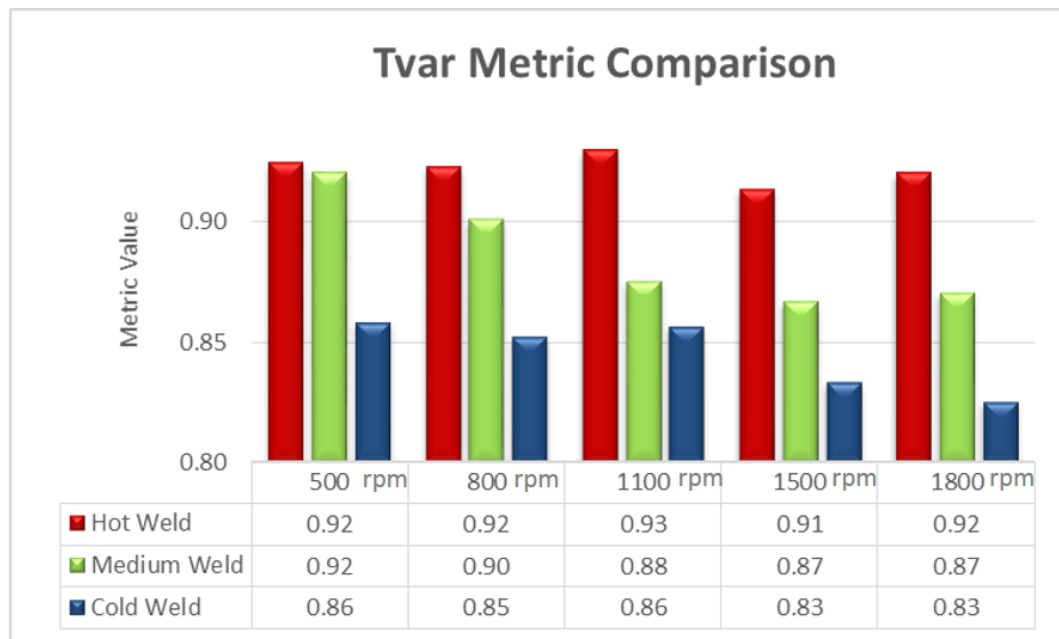
Figure 6-26 – MTC1 variation metric comparison at end of advance phase

Weld Quality Metrics - MTC1 Variation
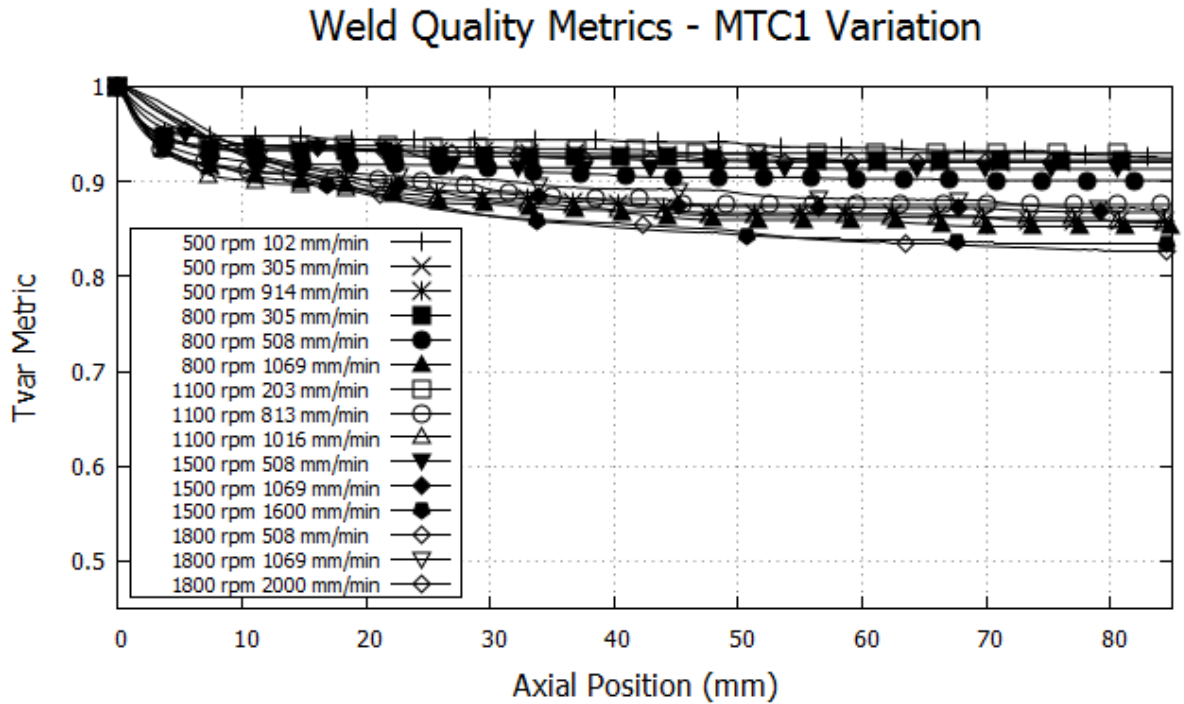
Figure 6-27 – Transient MTC1 variation metrics comparison for optimization cases

6.2.5 TOOL WEAR

A comparison of the tool wear results for the various cases is provided in Figure 6-28. The general trend is that the amount of wear increases (the wear metric decreases) with decreasing weld pitch. This effect is explained by the amount of time that each aluminum "packet" spends in contact with the tool. As the weld pitch decreases, the contact time increases leading to an increase in the tool wear. Furthermore, there is a trend towards increasing tool wear as the rpm increases. This can be explained by the increase in weld temperature, leading to a decrease in the hardness of the tool, ultimately increasing the wear rate.

Wear Metric Comparison

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
Hot Weld	0.97	0.97	0.97	0.97	0.96
Medium Weld	0.98	0.97	0.97	0.97	0.97
Cold Weld	0.98	0.98	0.98	0.97	0.97
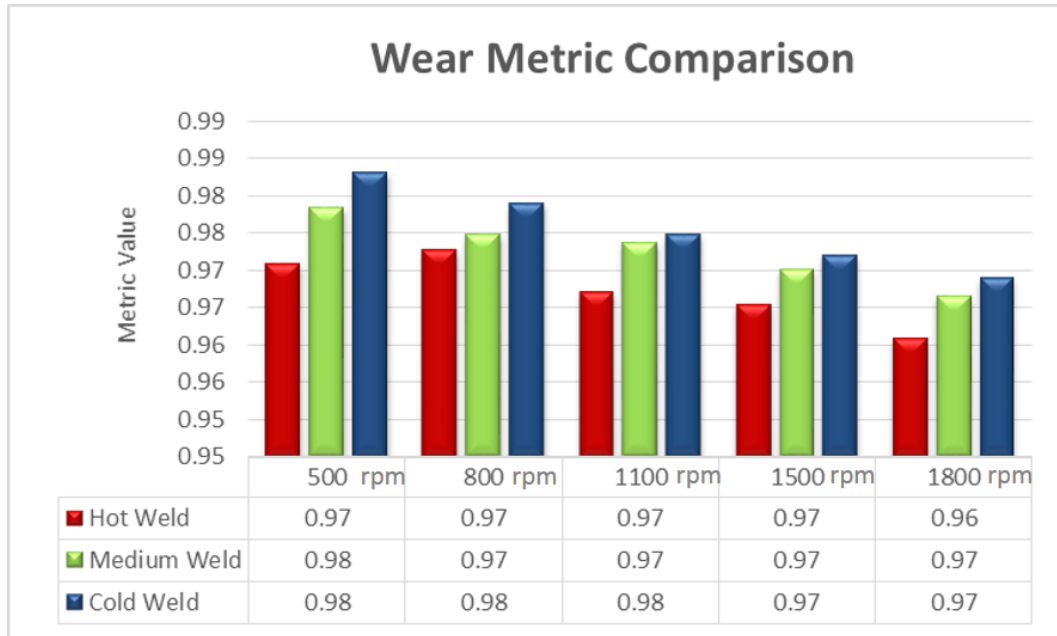
Figure 6-28 – Tool wear metric comparison at end of advance phase

Weld Quality Metrics - Tool Wear

Legend:
- 500 rpm 102 mm/min
- 500 rpm 305 mm/min
- 500 rpm 914 mm/min
- 800 rpm 305 mm/min
- 800 rpm 508 mm/min
- 800 rpm 1069 mm/min
- 1100 rpm 203 mm/min
- 1100 rpm 813 mm/min
- 1100 rpm 1016 mm/min
- 1500 rpm 508 mm/min
- 1500 rpm 1069 mm/min
- 1500 rpm 1600 mm/min
- 1800 rpm 508 mm/min
- 1800 rpm 1069 mm/min
- 1800 rpm 2000 mm/min
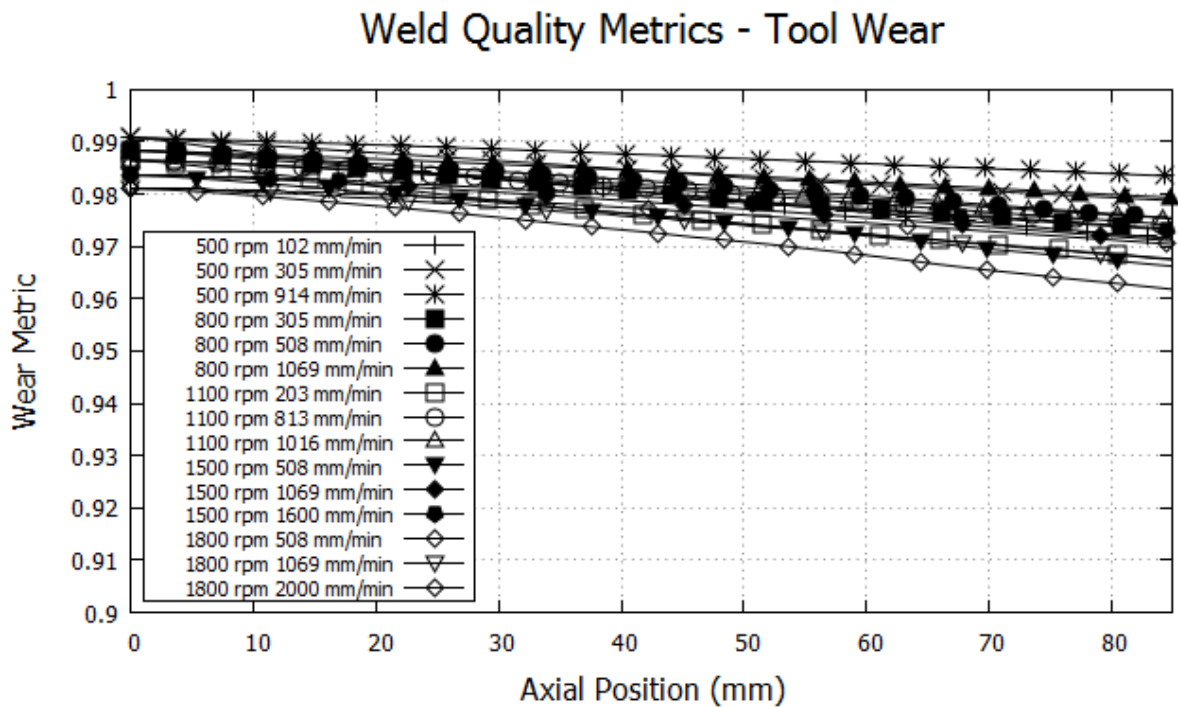
Figure 6-29 – Transient tool wear metrics comparison for optimization cases

A comparison of the tool wear metric throughout the advancing phase is shown in Figure 6-29. As expected, the wear increases (metric decreases) steadily throughout the advancing phase. A

comparison of the wear contours for the different cases is presented in Figure 6-30. These results show the extent of the wear in the pin base, pin side, and the shoulder. The proportion of tool wear on the pin side to the shoulder tends to increase with higher advancing speeds, leading to the understanding that the pressure is higher on the tool side as the weld pitch increases. Obviously, the location and magnitude of the wear contours is dependent on the relative time spent during each of the process phases. Since the weld considered in this work is short, the proportion of tool pin wear, to shoulder wear will be much larger than if a longer weld were performed.
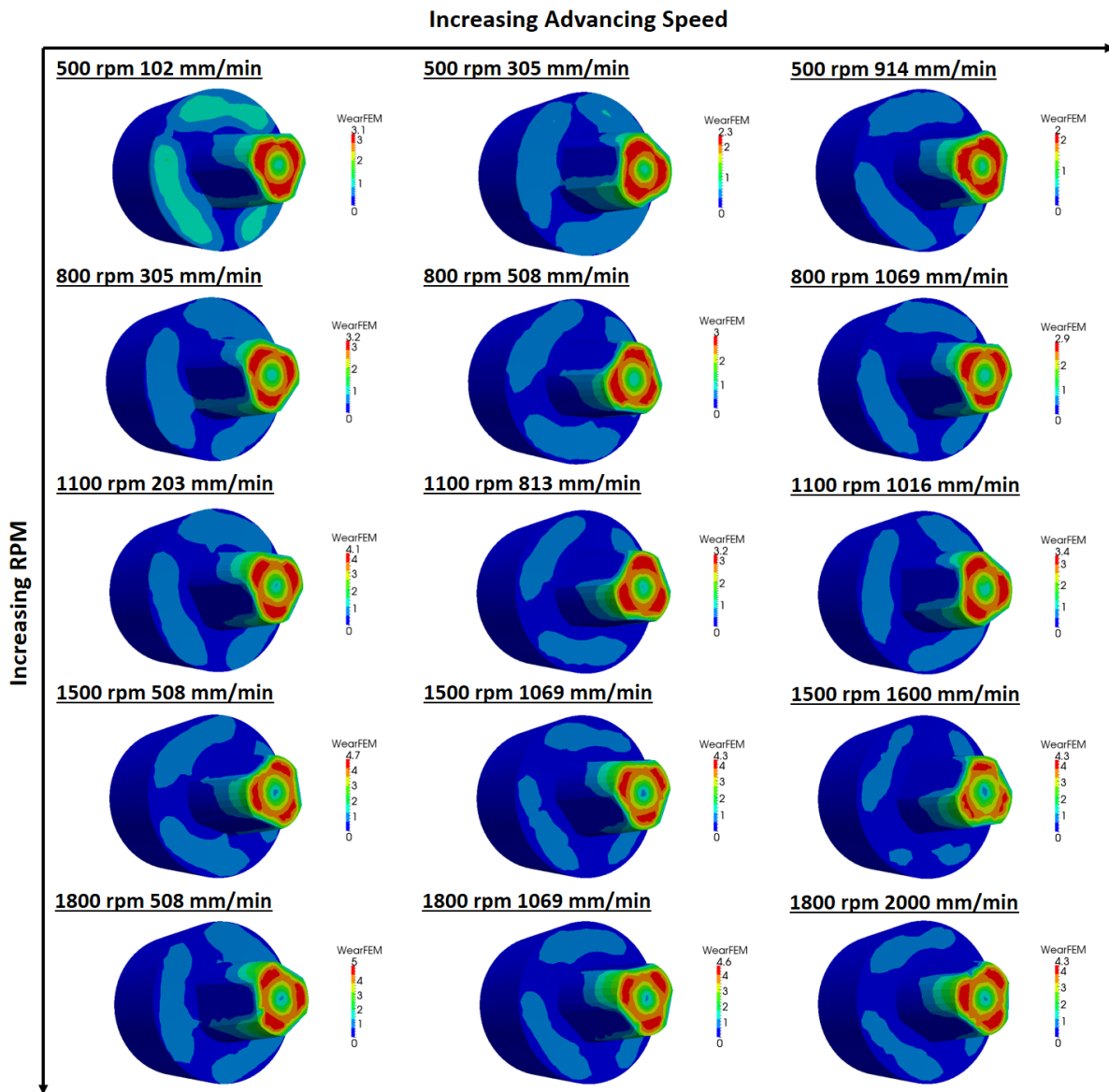
Figure 6-30 – Tool wear contours for the cases considered in the optimization analysis

6.2.6 OVERALL WELD QUALITY

The overall weld quality metric as determined by equation (6-9) for the end of and throughout (instantaneous) the advance phase is shown in Figure 6-31 and Figure 6-32 respectively. The overall weld quality is strongly influenced by the defect metric (since the weighting factor is greater than for the other metrics). The general trend shows improved weld quality for the "warm" welds in the range of 0.4 to 0.6 weld pitch.

	500 rpm	800 rpm	1100 rpm	1500 rpm	1800 rpm
■ Hot Weld	0.79	0.85	0.81	0.86	0.87
■ Medium Weld	0.84	0.87	0.86	0.85	0.84
■ Cold Weld	0.81	0.83	0.85	0.83	0.83
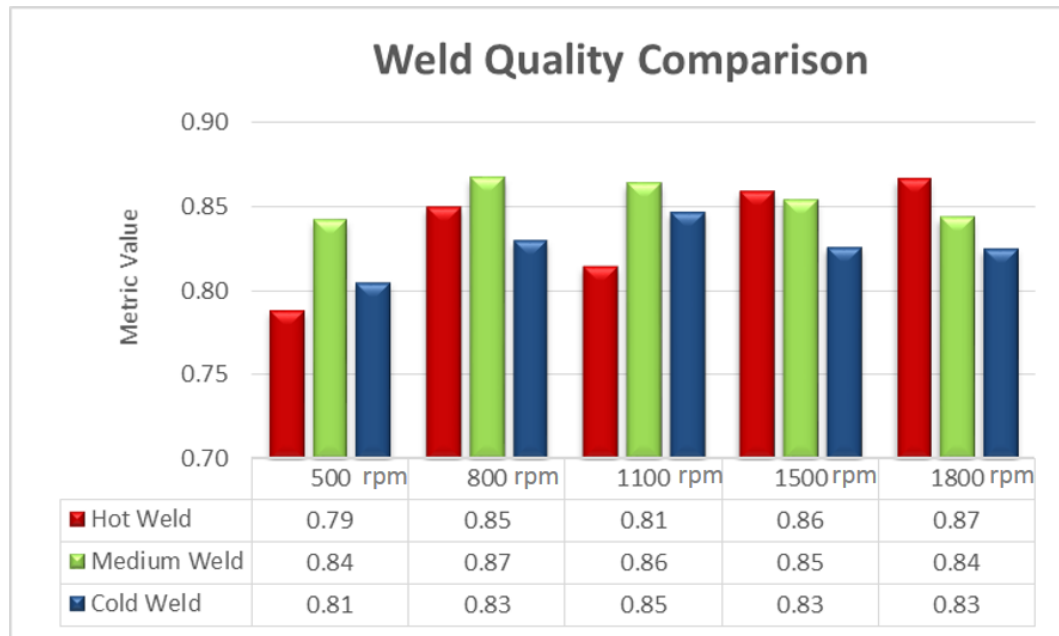
Figure 6-31 – Overall weld quality comparison at end of advance phase

Overall Weld Quality
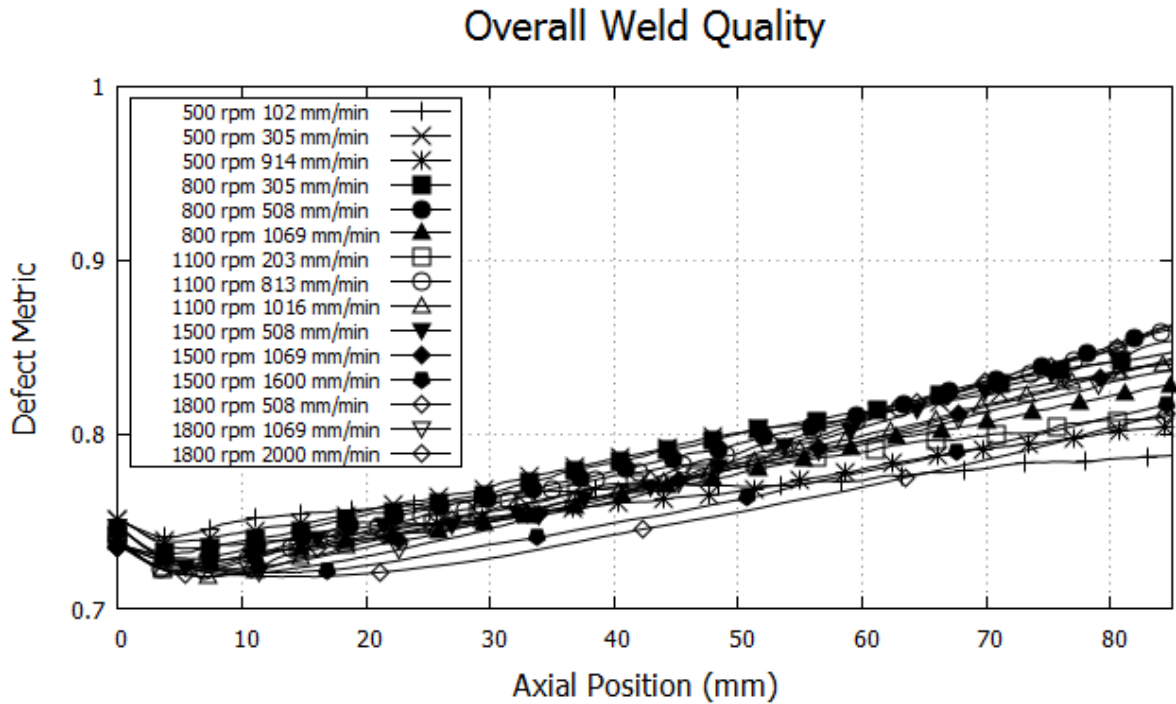
Figure 6-32 - Transient overall weld quality comparison for optimization cases

The transient results show a steady increase in the weld quality as the tool advances. The increase is expected (and necessary) since an incomplete weld, whereby the tool does not advance to the end of the plate, should result in a lower weld quality (no weld is formed, obviously leading to a poor joint). This graphic helps to verify that the proposed weld quality metric formulation is intuitive and meaningful.

6.3 RESPONSE SURFACE CONSTRUCTION

The results from the 15 test cases will now be used to form response surfaces, RS. The surfaces offer invaluable insight into the process, provide a powerful means to determine how a change in process parameters will affect the weld quality, and will afford a means to optimize the process parameters.

To form the RS, a file is created with the value of each metric (z) at the end of the simulation as a function of rpm (ω) and advancing speed (V_a). The data is then used to perform a least squares surface fitting algorithm, called regression. A general polynomial surface of degree three can be constructed of the form:

$$F_{RS}(\omega, V_a) = b_0 + b_1\omega^2 + b_2\omega + c_1 V_a^2 + c_2 V_a + d_1\omega V_a + d_2\omega^2 V_a + d_3\omega V_a^2 \tag{6-11}$$

The form of the proposed polynomial surface is chosen for its ability to provide an excellent fit to a wide variety to data groupings. Although the surface involves eight coefficients that need to be found, the additional work is well worth the reward. In some cases, the higher order terms are very small, ultimately leading to a plane like surface that fits the data.

The least squares approach is, in essence, an optimization problem. The goal is to minimize the sum of the squared residuals. The residual, R_{RS}, is defined as the difference between the actual values (in this case, the metrics) and the proposed function. The ultimate goal is to find the values of the coefficients in equation (6-11). The ith residual is:

$$R_{RS_i} = z_i - F_{RS}(\omega_i, V_{a_i}) \tag{6-12}$$

The objective function to be minimized is the sum of squares of the residual:

$$SS_{res} = \sum_{i=1}^{n-1} R_{RS_i}^2 \tag{6-13}$$

The value of n will be the number of data sets to be used in the regression analysis, in the case at hand, $n = 15$. To minimize SS_{res}, the gradient of the objective function with respect to the undetermined coefficients is set to zero.

$$\frac{\partial}{\partial \beta_\alpha} SS_{res} = 0 \, , \, \beta_\alpha = b_0, b_1, b_2, c_1, c_2, d_1, d_2, d_3 \tag{6-14}$$

This forms a set of linear equations that can be solved by an appropriate method (be it direct or iterative). Once the value of the coefficients are found, the appropriateness of the proposed polynomial can be evaluated based on the coefficient of determination, R^2, value:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6-15}$$

where SS_{tot} is the total sum of squares:

$$SS_{tot} = \sum_{i=1}^{n-1} (z_i - \bar{z})^2 \qquad (6\text{-}16)$$

and $\bar{z}$ is the mean of the data. R^2 will have a value of 1.0, if the polynomial function is a perfect fit to the observed data. In this sense, the closer the R^2 is to 1.0, the better the fit.

6.3.1   RESULTS

The various polynomial coefficients found from the regression analysis are presented for the different cases in Table 6-4. In general, the R^2 values show that the proposed polynomial surface is well suited to represent the simulation results. The defect metric goodness of fit is lower than the other metrics due to the scatter in the data, because of this scatter; the response surface is considered an appropriate fit in this case.

Table 6-4 – Response surface coefficient results

	Defects	Mixing	Tmax	Tvar	Wear
b_0	0.780	0.610	1.070	0.940	0.976
b_1	3.529E-8	2.827E-7	1.917E-7	2.364E-8	3.382E-10
b_2	-3.919E-5	-3.016E-6	-5.523E-4	-2.000E-5	-1.278E-5
c_1	-4.982E-7	-1.040E-6	1.185E-8	-4.785E-8	-1.739E-8
c_2	2.635E-4	9.158E-4	2.251E-5	-5.451E-5	3.147E-5
d_1	2.813E-7	-2.006E-7	1.688E-8	3.230E-9	-8.392E-9
d_2	-2.328E-10	-4.563E-10	-2.024E-11	-2.599E-11	-1.025E-13
d_3	2.754E-10	6.061E-10	-4.674E-12	4.206E-11	7.331E-12
R^2	0.787	0.938	0.994	0.991	0.984

The various response surfaces are shown in Figure 6-33. In general, the models depict a decrease in weld quality as the weld pitch increases (low rpm with high advancing speed). The models for the maximum temperature, MTC1, and tool wear do not contain a local extremum within the region of interest; instead the models predict a continued increase in weld quality (surpassing a metric of 1.0).

Such a result is not physical, requiring the engineer designing the FSW process to take care to understand fully the appropriate design space. Figure 6-34 shows contour plots of the response surfaces. These plots help to visualize the local extremum present in the defect, mixing and weld quality models.
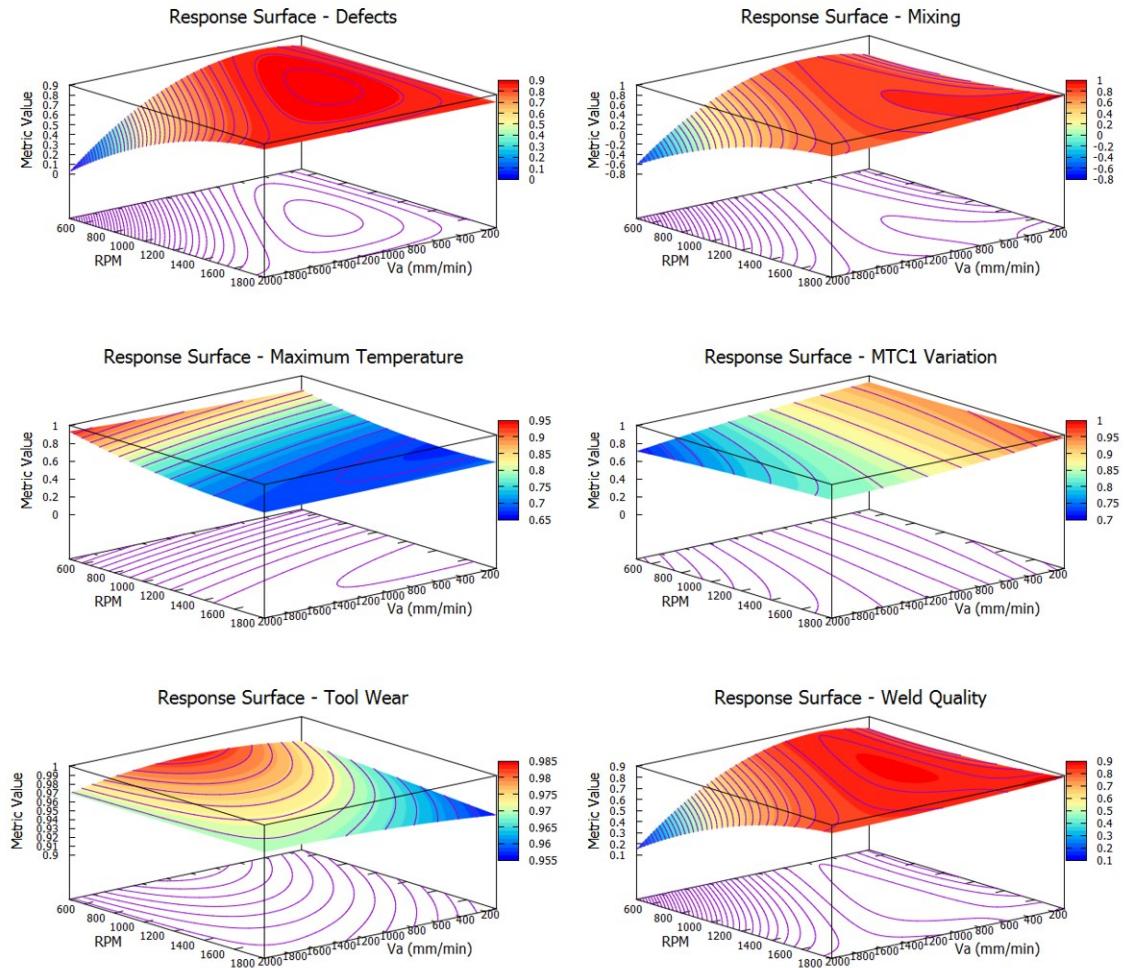
Figure 6-33 – Response surfaces: defects (top left), mixing (top right), maximum temperature (center left), temperature variation (center right), tool wear (bottom left), weld quality (bottom right)
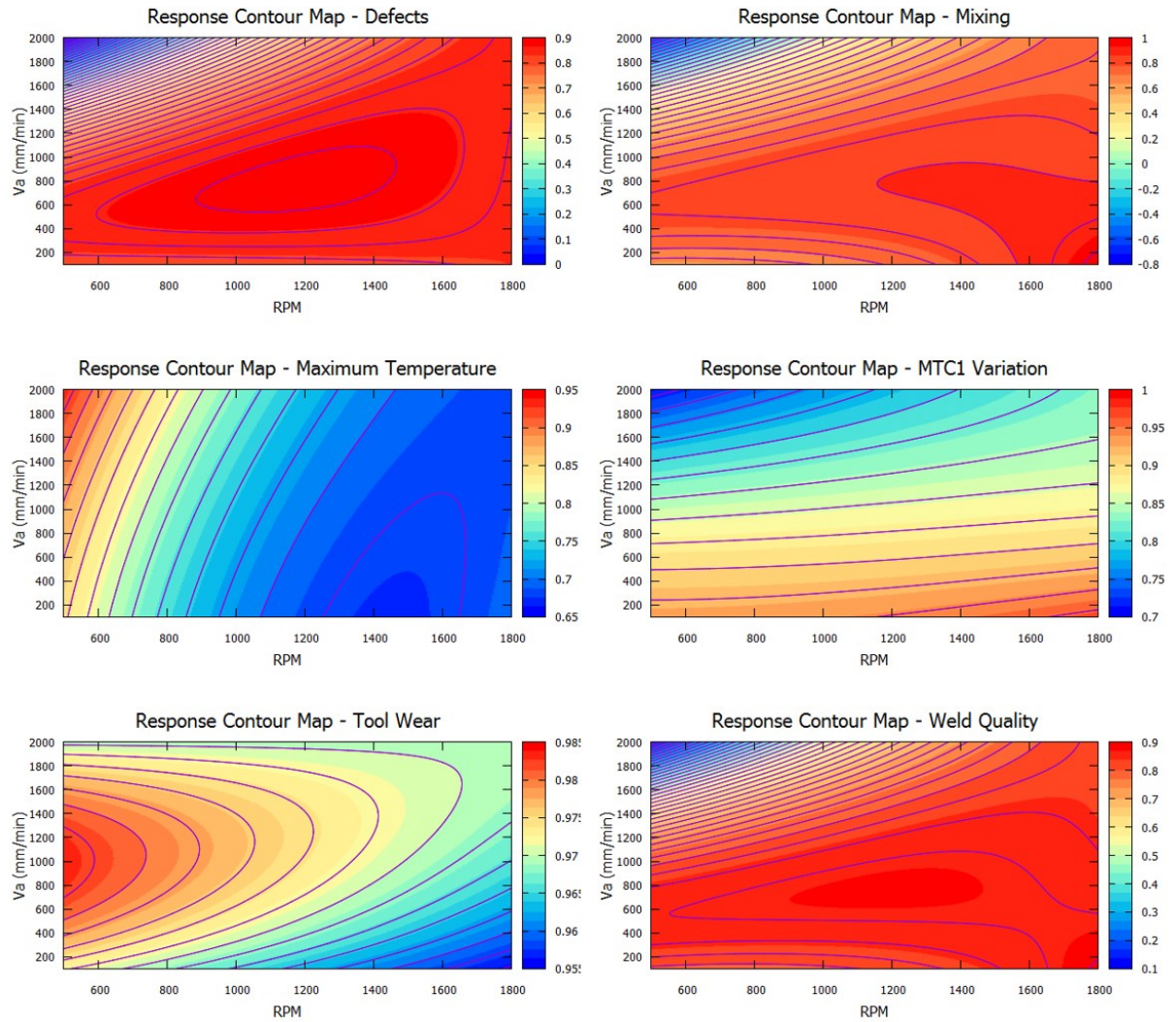
Figure 6-34 – Contour maps of the response surfaces: defects (top left), mixing (top right), maximum temperature (center left), temperature variation (center right), tool wear (bottom left), weld quality (bottom right)

6.4 OPTIMIZATION - MAXIMIZING OVERALL WELD QUALITY

The overall weld quality is an important metric and is a powerful tool that can be used for optimizing the weld quality. The goal of the optimization is to find the best possible weld quality, and the associated rpm (ω^*), and advancing speed ($V_a{}^*$). The optimization problem can be described as:

Maximize: $\qquad\qquad \psi_{weld\ quality}(\omega, V_a)$

Subject to (constraints): $500\ rpm \le \omega \le 1800\ rpm$

$$100\frac{mm}{min} \le V_a \le 2000\ mm/min$$

The weld quality response surface with the constraint bounds is shown in Figure 6-35. From inspection of the surface, a local maximum is clearly present within the desired bounds. The steepest descent approach can be used for this optimization problem since the local minimum will inherently satisfy the constraints. The general idea of this approach is to iteratively find a local minimum by following the gradient (the descent direction) of the function. The algorithm can be described by:

1- Choose a starting point (ω_0, V_{a_0}) within the constraint bounds

2- Choose a descent step size, α_{SD}

3- Calculate the gradient at the current point, $\nabla\psi_{weld\ quality}(\omega_n, V_{a_n})$

4- Find the next iteration rpm from, $\omega_{n+1} = \omega_n - \alpha_{SD}\frac{\partial}{\partial\omega}\psi_{weld\ quality}(\omega_n, V_{a_n})$

5- Find the next iteration advancing speed from, $V_{a_{n+1}} = V_{a_n} - \alpha_{SD}\frac{\partial}{\partial V_a}\psi_{weld\ quality}(\omega_n, V_{a_n})$

6- Evaluate the relative error from the change from the previous iteration $\varepsilon_{rel} = max(|\omega_{n+1} - \omega_n|, |V_{a_{n+1}} - V_{a_n}|)$

7- Evaluate the absolute error from the function value at the current and previous step $\varepsilon_{abs} = |\psi_{weld\ quality}(\omega_{n+1}, V_{a_{n+1}}) - \psi_{weld\ quality}(\omega_n, V_{a_n})|$

8- Return to step 3, continue until ε_{abs} and $\varepsilon_{rel} \le ErrorTOL$ If the desired error tolerance is obtained, set $(\omega^*, V_a{}^*) = (\omega_{n+1}, V_{a_{n+1}})$
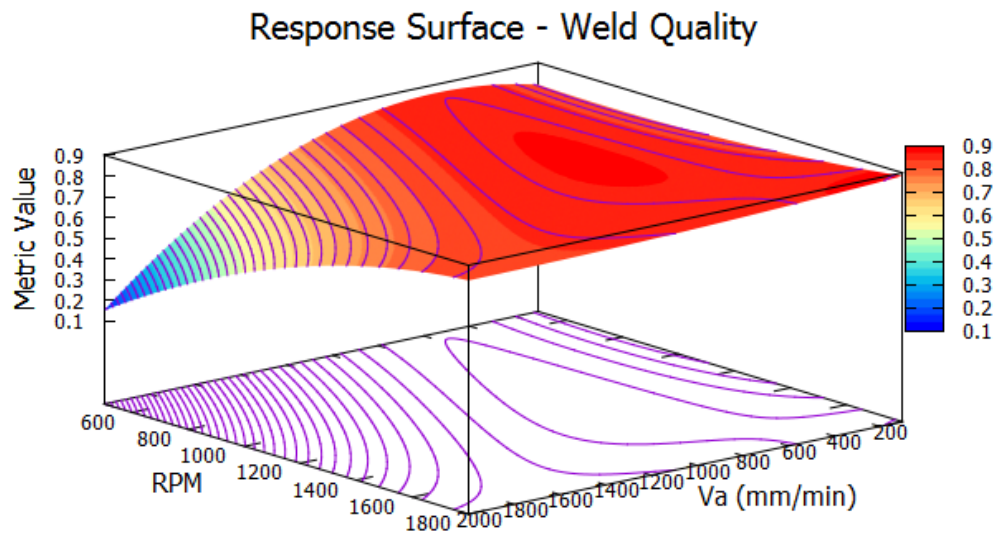
Figure 6-35 – Weld quality response surface

The use of ε_{abs} and ε_{rel} is stringent, but is good practice. Although a fixed descent step size has been used here, alternatively, it can be find by solving the minimization problem $\frac{d}{d\alpha_{SD}}\psi_{weld\ quality}\left(\omega_n - \right.$

$$\alpha_{SD}\frac{\partial}{\partial\omega}\psi_{weld\ quality}\left(\omega_n, V_{a_n}\right), V_{a_{n+1}}, V_{a_n} - \alpha_{SD}\frac{\partial}{\partial V_a}\psi_{weld\ quality}\left(\omega_n, V_{a_n}\right)\right) = 0$$

for α_{SD}. This two-equation system leads to two solutions; the smaller value of the two can be taken as the step size. To use the steepest descent, the negative value of the objective function should be minimized (in order to descend).

The results from the optimization of the weld quality are shown in Table 6-5. The obtained value of 1211 rpm and 765 mm/min results in a weld pitch of 0.63. Figure 6-36 shows the descent path along the response surface towards the local minimum. This result fits well with the findings of Wanjara et al. [46] (found a weld pitch of 0.48 by optimizing the surface finish and internal defects for 3.2 mm thick AA6061-T6 butt joint welds) and Fraser et al. [151] (found a weld pitch of 0.52 by optimizing the maximum weld temperature in 3.2 mm thick AA6061-T6 butt joint welds). Although 0.63 is higher than the aforementioned results. Ultimately, the optimal values will be dependent on many factors such as the dimensions of the plates, the tool design, the support structure, the fabrication (rolled, extruded,

machined, etc.) process for the alloy, the surface preparation of the plates, as well as the specific chemical constituents.

Table 6-5 – Weld quality optimization results

	ω (rpm)	V_a (mm/min)	$\psi_{weld\ quality}$
Starting point, (ω_0, V_{a_0})	600	1800	0.403
Optimal Values, (ω^*, V_a^*)	1211	765	0.870
Weld pitch	0.63		
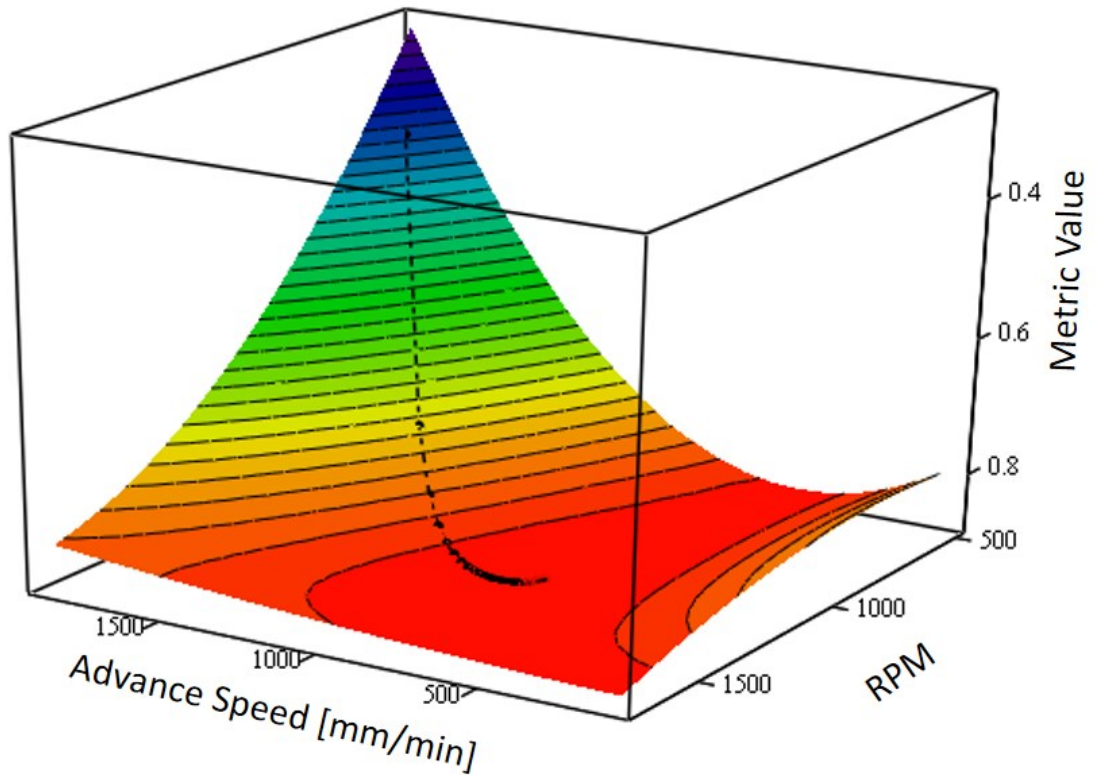
Figure 6-36 – Descent path for weld quality maximization

Unquestionably, simple inspection of the response surface at the determined values shows that indeed a local extremum has been found. To be certain of the result, the Hessian matrix can be checked:

$$\mathcal{H}_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} \psi_{weld\ quality}, \ x_i = \omega, x_j = V_a \tag{6-17}$$

Then, if $\mathcal{H}$ is negative definite, a local maximum has been found. The definiteness of a matrix can be found from:

$$\phi^T \mathcal{H} \phi \tag{6-18}$$

where ϕ is a non-zero vector. For the case at hand, it suffices to determine if the hessian is negative definite at the determined point:

$$(\omega^* \quad V_a{}^*) \mathcal{H} \begin{pmatrix} \omega^* \\ V_a{}^* \end{pmatrix} = -0.218 < 0 \tag{6-19}$$

which proves that the obtained values do indeed satisfy the conditions for a local maximum.

6.5 OPTIMIZATION - MINIMIZING DEFECTS

Finding the optimal process parameters that will minimize the defects can be accomplished using the steepest descent algorithm as well, since the response surface (Figure 6-37) contains a local maximum. The optimization problem can be stated as:

Maximize: $\qquad \psi_{defect}(\omega, V_a)$

Subject to (constraints): $\quad 500\ rpm \leq \omega \leq 1800\ rpm$

$$100\ mm/min \leq V_a \leq 2000\ mm/min$$
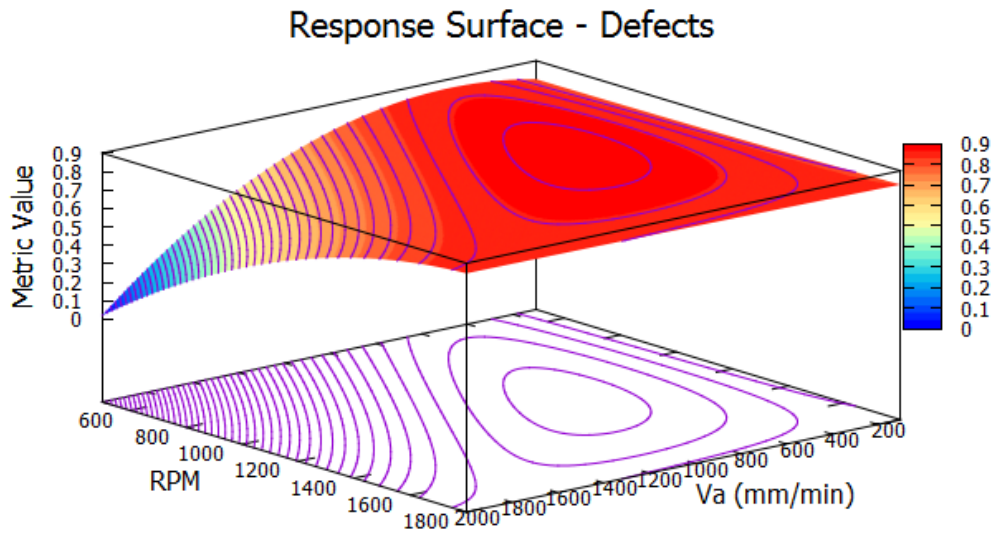
Figure 6-37 - Defects response surface

Using the approach as in section 6.4, the optimal values are found to be 1194 rpm with 790 mm/min. The results are summarized in Table 6-6. The resulting weld pitch of 0.66 is higher than what was found based on overall weld quality. This is an intuitive result since the defect metric will be less stringent compared to the overall weld quality. In optimizing the process based on defects, there is no consideration for the mixing, temperature, or wear. For this reason, the weld pitch should be expected to be slightly higher.

Table 6-6 – Defect optimization results

	ω (rpm)	V_a (mm/min)	$\psi_{weld\ quality}$
Starting point, $(\omega_0, V_{a\,0})$	600	1800	0.318
Optimal Values, (ω^*, V_a^*)	1194	790	0.889
Weld pitch	0.66		

The descent path is shown on the left side of Figure 6-38 for the starting point listed in Table 6-6. The starting point was chosen just within the bounds of the region of interest. Care must be taken to ensure

that the starting point will provide the correct descent direction. If the wrong point is chosen, the descent could occur entirely outside of the region of interest and thus find an invalid set of process parameters. Nevertheless, other starting points can be used. On the right side of Figure 6-38, a starting point of (0,0) is taken. This starting point is admissible, and leads to successfully obtaining the desired extremum.
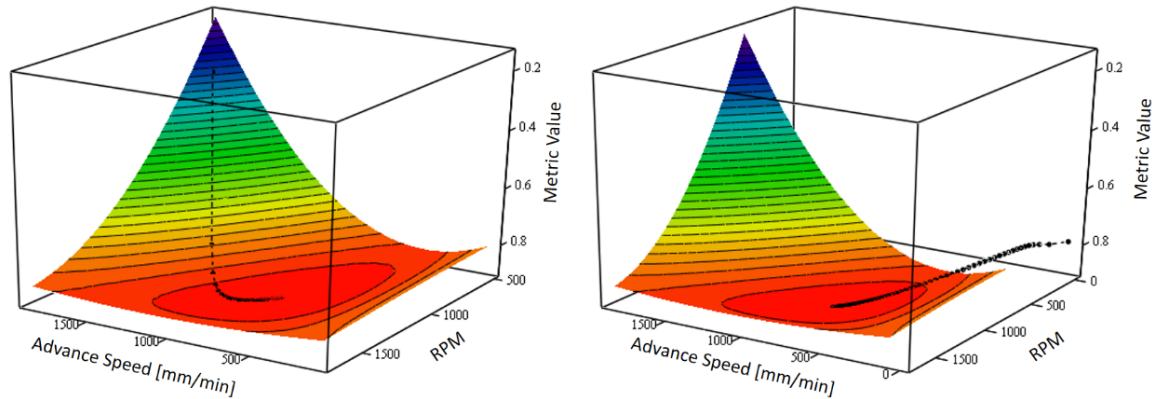
Figure 6-38 - Descent path for defect minimization

Again, for good measure the definiteness of the hessian matrix is checked and verified to result in a negative definite matrix:

$$(\omega^* \quad V_a{}^*)\mathcal{H}\begin{pmatrix} \omega^* \\ V_a{}^* \end{pmatrix} = -0.332 < 0 \tag{6-20}$$

which proves that the obtained values do indeed satisfy the conditions for a local maximum.

6.6 OPTIMIZATION - MAXIMIZING ADVANCE SPEED BASED ON WELD QUALITY

Finding the maximum advancing speed based on weld quality requires a different tactic than was used in section 6.4 and 6.5. The advancing speed response surface is found based on the described procedure in section 6.3, however, the polynomial function is modified as follows:

$$F_{RS}(\omega, \psi) = b_0 + b_1\omega^2 + b_2\omega + c_1\psi^2 + c_2\psi + d_1\omega\psi + d_2\omega^2\psi + d_3\omega\psi^2 \tag{6-21}$$

where ψ is the weld quality metric ($\psi_{weld\ quality}$). Again, the least squares method is used to find the value of the coefficients: $b_0 = -7.006E + 5$, $b_1 = 0.087$, $b_2 = 348.547$, $c_1 = -1.135E + 6$, $c_2 = 1.788E +$

6, $d_1 = -1.057E + 3$, $d_2 = -0.104$, and $d_3 = 764.552$, resulting in $R^2 = 0.787$. The response surface and the associated contour map are shown in Figure 6-39.

Notice that the surface does not have a local maximum in the region of interest. For this reason, the steepest descent method would lead to optimal values outside of the operating window of the CNC machine. In this case, the constraints must be included in the optimization problem solution. The problem is stated as:

Maximize: $\qquad\qquad V_a(\omega, \psi_{weld\ quality})$

Subject to (constraints): $\quad 200\ rpm \leq \omega \leq 1800\ rpm$

$$0.84 \leq \psi_{weld\ quality} \leq 0.88$$

The choice of the range of $\psi_{weld\ quality}$ is chosen to ensure a weld quality of at least 0.84 is obtained. The higher end of 0.88 is just outside the maximum admissible value of $\psi_{weld\ quality}$. Since the constraints will now play a critical role in the solution of the problem, a constrained optimization algorithm will be used.

Response Surface - Advancing Speed

Response Contour Map - Advancing Speed
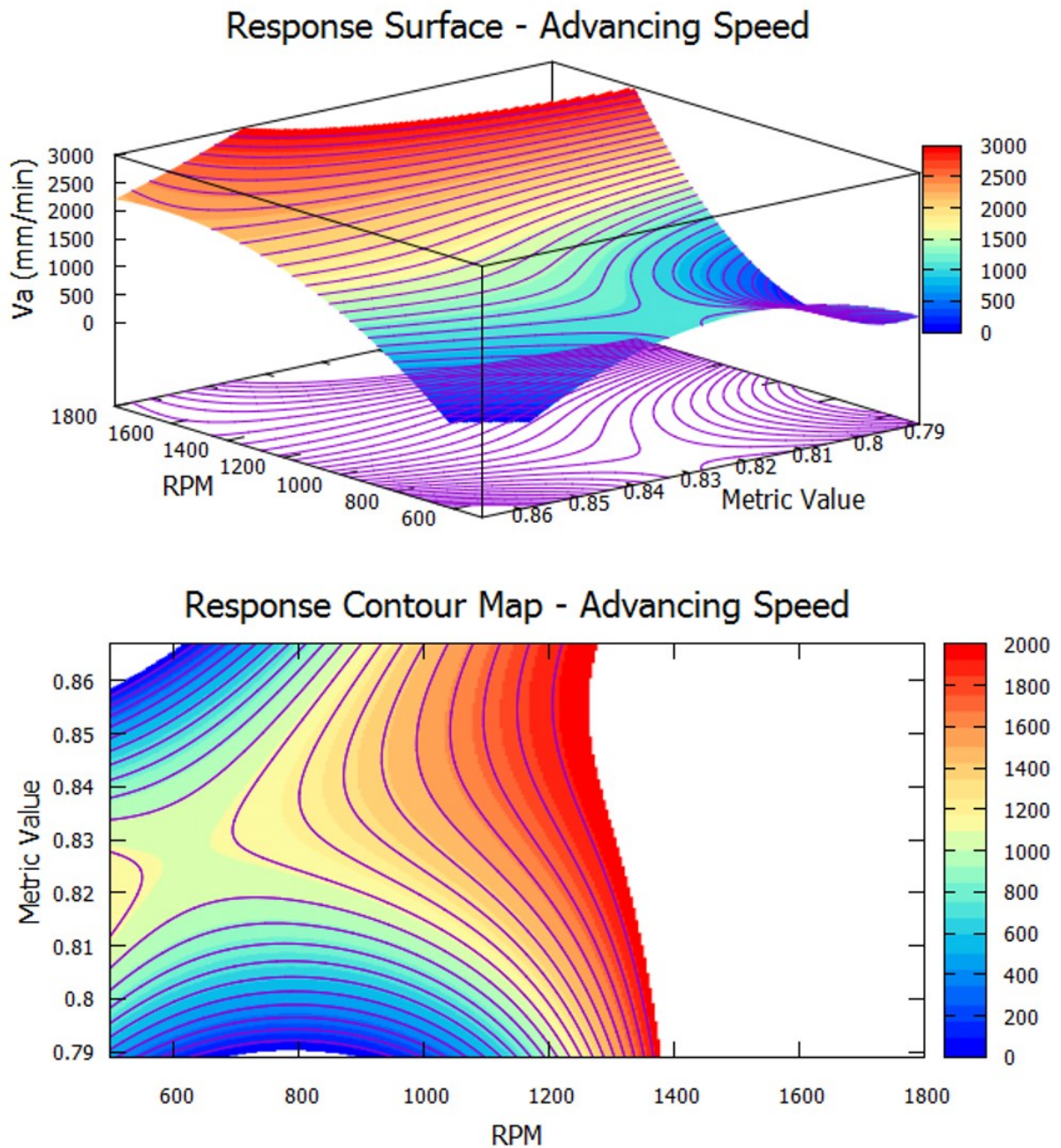
Figure 6-39 – Advancing speed as a function of rpm and weld quality response surface

Although there are many viable algorithm choices, the Lagrange multiplier method will be used here. The approach assembles the objective function and the constraints into the Lagrangian of the problem. Then the Lagrangian is used to find the optimal values. To use the Lagrange multiplier method, the

inequality constraints should be converted to a set of equality constraints. For the problem at hand, a simple and elegant solution is to cast the inequality constraints as an elliptic constraint function:

$$g\left(\omega, \psi_{weld\ quality}\right) = \left(\frac{\omega - 1000}{800}\right)^2 + \left(\frac{\psi_{weld\ quality} - 0.86}{0.02}\right)^2 - 1 \tag{6-22}$$

$g\left(\omega, \psi_{weld\ quality}\right)$ can be seen to loosely satisfy the inequality's constraints. For the problem at hand, this approximation is sufficient. Next, the gradient of the constraint equation must be shown to be linearly independent:

$$\frac{\partial}{\partial \omega} g\left(\omega, \psi_{weld\ quality}\right) = \frac{\omega}{320000} - \frac{1}{320}$$

$$\frac{\partial}{\partial \psi_{weld\ quality}} g\left(\omega, \psi_{weld\ quality}\right) = 5000 \psi_{weld\ quality} - 4300 \tag{6-23}$$

which is linearly independent within the region of interest. Now, the Lagrangian is given by:

$$\mathcal{L}\left(\omega, \psi_{weld\ quality}, \lambda\right) = V_a\left(\omega, \psi_{weld\ quality}\right) + \lambda g\left(\omega, \psi_{weld\ quality}\right) \tag{6-24}$$

where λ is the Lagrange multiplier. To find the optimal parameters, the gradient of the Lagrangian is set equal to zero:

$$\nabla_{\omega, \psi_{weld\ quality}, \lambda} \mathcal{L}\left(\omega, \psi_{weld\ quality}, \lambda\right) = 0 \tag{6-25}$$

The set of equations that result are linear of the form:

$$\bar{\bar{A}}\bar{x} - \bar{b} = 0, \quad \bar{x} \in \omega, \psi_{weld\ quality}, \lambda \tag{6-26}$$

which can be solved by any number of methods. For this particular problem, there are three equations in three unknowns. This can be solved readily by inverting the coefficient matrix, $\bar{\bar{A}}$, to find the value of the solution vector, $\bar{x}$:

$$\bar{x} = \bar{\bar{A}}^{-1}\bar{b} \tag{6-27}$$

Of course, for larger systems an iterative approach is preferred.

Table 6-7 – Advancing speed optimization results

	ω (rpm)	V_a (mm/min)	$\psi_{weld\ quality}$	λ
Optimal Values, $(\omega^*, \psi_{weld\ quality}{}^*)$	1527	1075	0.845	-163.2
Weld pitch		0.70		

Following this approach, the optimal values are found (summarized in Table 6-7) to be $\omega = 1527$ rpm, and $V_a = 1075$ mm/min with a corresponding weld quality of 0.845. To verify that the solution to the optimization problem with the Lagrange multiplier method is valid, the Karush-Kahn-Tucker conditions will be checked. First off, the feasibility condition on the constraint is checked:

$$g(\omega^*, \psi_{weld\ quality}{}^*) = \left(\frac{\omega^* - 1000}{800}\right)^2 + \left(\frac{\psi_{weld\ quality}{}^* - 0.86}{0.02}\right)^2 - 1 \approx 0 \tag{6-28}$$

Next, the optimality condition is checked:

$$\nabla_\omega V_a(\omega, \psi_{weld\ quality}) + \nabla_\omega \lambda g(\omega, \psi_{weld\ quality}) \approx 0 \tag{6-29}$$

$$\nabla_{\psi_{weld\ quality}} V_a(\omega, \psi_{weld\ quality}) + \nabla_{\psi_{weld\ quality}} \lambda g(\omega, \psi_{weld\ quality}) \approx 0 \tag{6-30}$$

Finally, the necessary condition is checked:

$$\nabla_\omega \mathcal{L}(\omega, \psi_{weld\ quality}, \lambda) \approx 0$$

$$\nabla_{\psi_{weld\ quality}} \mathcal{L}(\omega, \psi_{weld\ quality}, \lambda) \approx 0 \tag{6-31}$$

$$\nabla_\lambda \mathcal{L}(\omega, \psi_{weld\ quality}, \lambda) \approx 0$$

Since the aforementioned conditions have been met, the optimization satisfies the necessary and sufficient conditions. The determined weld pitch of 0.7 is certainly pushing the limits of the process window. A weld with such a high rpm and advancing speed for a 12.7 mm plate will undoubtedly result in a certain level of internal defects. However, in many structural applications, where a small amount of defects is acceptable, this result will provide a valuable understanding to the design engineer as to the upper end of the process parameter settings. Most likely, the proposed numerical simulation optimization approach can serve as a starting point for the weld technician to start a round of tests. Without previous

knowledge of the reasonable starting point for the tests, the technician would likely require far more time

to find acceptable (let alone, optimal) operating conditions.

7 CONCLUSIONS AND OUTLOOK

The research project has provided a new approach to study and understand the FSW process. The presented developments will open new avenues to comprehensive simulation studies of the FSW process. Certainly, the idea that the "purpose of computing is insight, not numbers" [305] leads to the recognition that a well-tailored numerical study can provide far more than simply the value of the temperature at a point, or the stress state in the weld zone. More importantly, the model can and should afford valuable insight into the studied process. In a general sense, a model is useful if it provides the engineer or researcher with a new appreciation of why a certain result has occurred. From such a fundamental understanding, the underlying physics are revealed and true innovation can begin.

During the research project, a broad study of the current state of the art was undertaken. The results of the study lead to the understanding that a robust and efficient meshfree simulation code was needed. Furthermore, a need for an improved material model was established. The developed FKS flow stress model is better capable of representing the material behaviour across a wide spectrum of strain rates, temperatures, and strain hardening conditions. The literature consulted also lead to the conclusion that an improved, phenomenological contact interface modeling approach was needed. The developed cumulative damage model more precisely incorporates the underlying physics at the contact interface between the FSW tool and the workpieces. Assessment of the recent work in FSW simulation lead to the development of a robust and efficient thread and scroll modeling approach, wear prediction, as well as novel hybrid thermal-mechanical contact algorithm.

Once the fundamental background was presented for the various algorithms required to simulate precisely the entire FSW process, a number of test cases were presented. These cases highlighted various important aspects of the process; with the focus on temperature histories, tool torque and force, and most importantly, defects. Certainly, the prediction of defects with a numerical simulation model requires the use of advanced numerical methods. Only a meshfree, large plastic deformation code

written to take into account the underlying physics could attain such excellent correlation between simulation and experiment. Furthermore, since SPHriction-3D was developed entirely on the GPU, the aforementioned results were obtained within a reasonable timeframe (less than a day compared to many days or weeks

Finally, the full set of developments was leveraged to perform a comprehensive optimization of the FSW process. To this end, a new and novel series of weld quality metrics was presented that are considered to provide an excellent measure of the quality of the finished weld. These metrics then served as the building blocks for the development of response surfaces to be used to find optimized process parameters. Indeed, the parameters focused on in this work were rpm and advancing speed; however, with little effort, the proposed method could be used to find the optimal tool geometry, the ideal base support material, or even an improved clamping design to reduce residual stresses. Because of the parametric nature of the proposed optimization approach, the possibilities are only limited by the scope of the researcher's imagination. Since the meshfree code was completely developed from ground zero with the goal of being able to easily include new physics, the possibilities are endless.

The project served to develop a fully coupled meshfree solid mechanics and thermal solution entirely on the GPU. Previous fully coupled implementations incurred long run times that would have led to un-reasonably long calculations. This then, would have discouraged the average design engineer in a company wishing to optimize their FSW process from a simulation approach. Since the developed code, SPHriction-3D, is approximately 20 to over 30 times faster than an equivalent meshfree code running on a CPU system, this allows a company to find optimized process parameters within a working day or two (using a multi-GPU system, which would be the fraction of the cost of an equivalent CPU computing cluster). Ultimately, the presented developments will offer a means for companies to understand better the FSW process. One of the main hurdles that FSW has encountered in the past decades is a lack of understanding [3] along with the notion that conventional welding techniques are the best avenue. SPHriction-3D provides a powerful simulation platform for companies that do not wish to or cannot afford to invest large quantities of money into a virtual prototyping simulation approach.
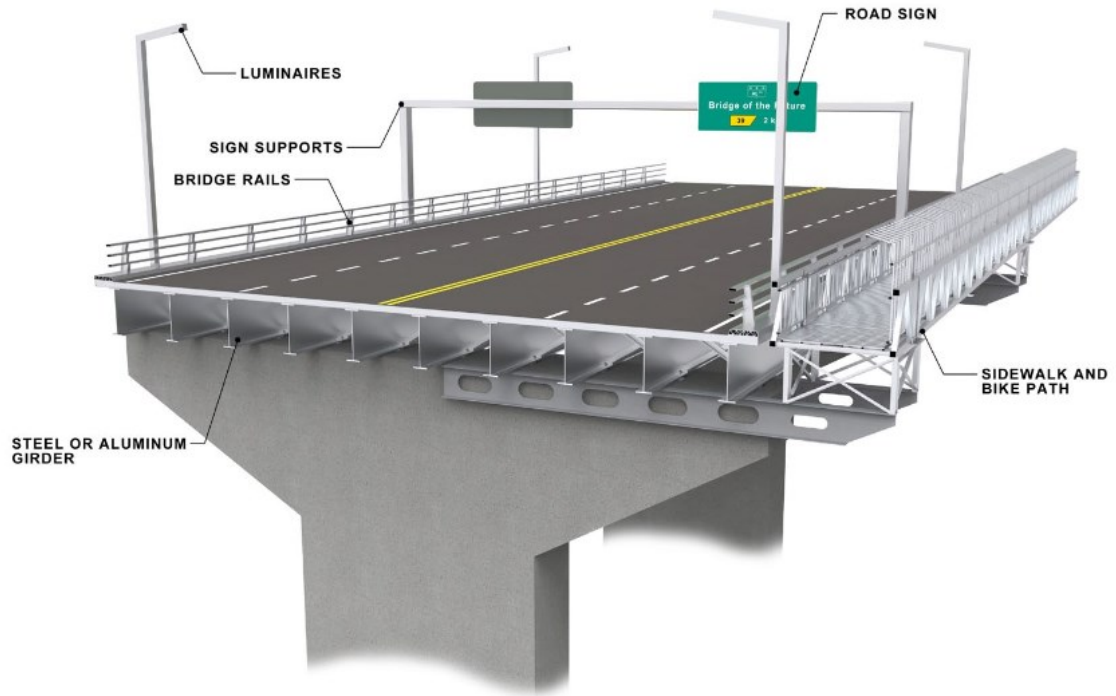
Figure 7-1 – Aluminum bridge construction [306]

Looking forward, a number of the developments presented here warrant further investigation. The FKS flow stress model was developed based on a set of compression tests. There would be a benefit in performing tension and torsion tests as well. This would make the proposed material model more general. Further investigation of the cumulative damage friction law is also warranted. The proposed cumulative damage approach (based on Johnson-Cook damage) could be replaced with a more appropriate damage law that is more specifically tailored to the FSW process. Another path that deserves further investigation and development is the implementation of a fully implicit smoothed particle method (see Appendices for details). The basic formulation has been developed, but at the time that this thesis was written, the method is not yet a viable approach to simulate the FSW process.

Ultimately, the project strives to open up new understanding of the process and to promote the use of FSW. More importantly, in a general sense, helps to promote the use of aluminum alloys in everyday structural applications such as the aluminum bridge picture in Figure 7-1. With the decaying bridge infrastructure in the province of Quebec [307] and elsewhere in Canada and the US [308], there is renewed interest in the use of low corrosion alloys such as aluminum. With a better understanding of the

FSW process will come widespread acceptance, leading to the viability of using typically difficult to join alloys for critical infrastructure projects.

8 REFERENCES

[1] Barham L. From Hand to Handle - The first industrial revolution. NY, USA: Oxford University Press; 2013.

[2] Colwell KC. Two metals enter, one leaves: the mirable of friction stir welding. http://www.caranddriver.com/features/two-metals-enter-one-metal-leaves-the-miracle-of-friction-stir-welding-tech-dept: Car and Driver; 2013.

[3] Thomas WM, Nicholas ED, Watts ER, Staines DG. Friction based welding technology for aluminum. 8th International Conference on Aluminum Alloys. Cambridge UK2002.

[4] Bhamji I, Pruess M, Threadgill PL, Addison AC. Solid state joining of metals by linear friciton welding: a literature review. Material Science and Technology. 2011;27:2-12.

[5] Brannon RM. Geometric Insight into Return Mapping Plasticity Algorithms. New Mexico: University of New Mexico; 2002.

[6] Brannon RM. Functional and Structured Tensor Analysis for Engineers. New Mexico: University of New Mexico; 2013.

[7] Bonet J, Wood RD. Nonlinear Continuum Mechanics for Finite Element Analysis. Cambridge, UK: Cambridge University Press; 1997.

[8] Fung Y-C. A first course in continuum mechanics. 2nd ed ed. Englewood Cliffs, N.J.: Prentice-Hall; 1977.

[9] Spencer AJM. Continuum mechanics. London ; New York: Longman; 1980.

[10] Campanella D, Buffa G, Fratini L, Merklein M. On the solid bonding phenomena in linear friction welding and accumulative roll bonding processes: Numerical simulation insights. Key Engineering Materials2015. 485-91.

[11] Ducato A, Campanella D, Buffa G, Fratini L. Design of numerical simulations of linear friction welding processes: Issues and difficulties. Key Engineering Materials2015. 451-8.

[12] Ji S, Wang Y, Liu J, Meng X, Tao J, Zhang T. Effects of welding parameters on material flow behavior during linear friction welding of Ti6Al4V titanium alloy by numerical investigation. International Journal of Advanced Manufacturing Technology. 2016;82:927-38.

[13] Maio L, Franco F, Squillace A, Lecce L. A simplified approach to numerical simulation of LFW process of Ti6Al4V alloy: investigation on friction and temperature. International Journal of Advanced Manufacturing Technology. 2016:1-12.

[14] Nikiforov R, Medvedev A, Tarasenko E, Vairis A. Numerical simulation of residual stresses in linear friction welded joints. Journal of Engineering Science and Technology Review. 2015;8:49-53.

[15] Schmicker D. A holistic approach on the simulation of rotary firction welding. Magdeburg, Germany: Otto-von-Guericke University of Magdeburg; 2015.

[16] Schmicker D, Naumenko K, Strackeljan J. A robust simulation of Direct Drive Friction Welding with a modified Carreau fluid constitutive model. Computer Methods in Applied Mechanics and Engineering. 2013;265:186-94.

[17] Schmicker D, Persson PO, Strackeljan J. Implicit Geometry Meshing for the simulation of Rotary Friction Welding. Journal of Computational Physics. 2014;270:478-89.

[18] Kren L. Big-bang bonding. http://machinedesign.com/archive/big-bang-bonding: Machine Design; 2000.

[19] Tanaka K. Numerical studies on the explosive welding by smoothed particle hydrodynamics (sph).

[20] Pabst C, Sharafiev S, Groche P, Wagner M. A novel method to investigate the principles of impact welding: Development and enhancement of a test rig, experimental and numerical results. Advanced Materials Research2014. 500-509.

[21] Wang YX, Li XJ, Sun G. Three dimensional simulation of the explosive welding by using of the MPM. Jisuan Lixue Xuebao/Chinese Journal of Computational Mechanics. 2013;30:34-8.

[22] Wang YX, Li XJ, Wang XH, Yan HH, Sun M. Numerical simulation on interfacial wave formation in explosive welding using material point method. Baozha Yu Chongji/Explosion and Shock Waves. 2014;34:716-22.

[23] Xie FH, Wu QH, Li WY, Wan SM, Yang YL. Numerical analysis on dynamics parameters of explosive welding in sharp wedge zone. Zhendong yu Chongji/Journal of Vibration and Shock. 2013;32:96-8+103.

[24] Smith IJ, Lord DDR. FSW Patents - A stirring story. SAE Transactions Journal of Materials and Manufacturing. 2007;116.

[25] ESAB. FSW Tools - Design is critical. http://www.esab.com/automation/en/process/fsw/Tools.cfm2016.

[26] Miranda RM, Gandra J, Vilac P. Surface modification by friciton based processes. Modern Surface Engineering Treatments: InTech; 2013.

[27] Lin YC, Chen JN. Influence of process parameters on friction stir spot welded aluminum joints by various threaded tools. Journal of Materials Processing Technology. 2015;225:347-56.

[28] Bilici MK, Yükler AI. Influence of tool geometry and process parameters on macrostructure and static strength in friction stir spot welded polyethylene sheets. Materials and Design. 2012;33:145-52.

[29] St-Georges L, Kiss LI. Soudage par friction malaxage: principles et applications2016.

[30] Casavola C, Cazzato A, Moramarco V, Pappalettere C. Influence of the clamps configuration on residual stresses field in friction stir welding process. Journal of Strain Analysis for Engineering Design. 2015;50:232-42.

[31] Fratini L, Pasta S. Residual stresses in friction stir welded parts of complex geometry. International Journal of Advanced Manufacturing Technology. 2012;59:547-57.

[32] Ji S, Wen Q, Lü Z, Yang Z. Effects of intense cooling on deformation and residual stresses for FSWed TC4 titanium alloy. Zhongguo Jixie Gongcheng/China Mechanical Engineering. 2016;27:531-6 and 43.

[33] Nateghi E, Hosseinzadeh M. Experimental investigation into effect of cooling of traversed weld nugget on quality of high-density polyethylene joints. International Journal of Advanced Manufacturing Technology. 2016;84:581-94.

[34] Papahn H, Bahemmat P, Haghpanahi M. Effect of cooling media on residual stresses induced by a solid-state welding: underwater FSW. International Journal of Advanced Manufacturing Technology. 2016;83:1003-12.

[35] Richards DG, Prangnell PB, Withers PJ, Williams SW, Nagy T, Morgan S. Efficacy of active cooling for controlling residual stresses in friction stir welds. Science and Technology of Welding and Joining. 2010;15:156-65.

[36] Huang Y, Wan L, Huang T, Lv Z, Zhou L, Feng J. The weld formation of self-support friction stir welds for aluminum hollow extrusion. International Journal of Advanced Manufacturing Technology. 2016:1-9.

[37] Parida B, Vishwakarma SD, Pal S. Design and development of fixture and force measuring system for friction stir welding process using strain gauges. Journal of Mechanical Science and Technology. 2015;29:739-49.

[38] Schmidt H, Hattel JH, Wert J. An analytical model for the heat generation in friction stir welding. Institute of Physics - Modelling and simulation in materials science and engineering. 2004;12:143-57.

[39] Heurtier P, Jones MJ, Desrayaud C, Driver JH, Montheillet F, Allehaux D. Mechanical and thermal modelling of Friction Stir Welding. Journal of Materials Processing Technology. 2006;171:348-57.

[40] Khandkar MZH, Khan JA, Reynolds AP. A thermal model of the friction stir welding process. American Society of Mechanical Engineers, Heat Transfer Division, (Publication) HTD. 5 ed2002. p. 115-24.

[41] Khandkar MZH, Khan JA, Reynolds AP. Prediction of temperature distribution and thermal history during friction stir welding: Input torque based model. Science and Technology of Welding and Joining. 2003;8:165-74.

[42] Arora A, Nandan R, Reynolds AP, DebRoy T. Torque, power requirement and stir zone geometry in friction stir welding through modeling and experiments. Scripta Materialia. 2009;60:13-6.

[43] Durdanović MB, Mijajlović MM, Milčić DS, Stamenković DS. Heat generation during friction stir welding process. Tribology in industry. 2009;31:8-14.

[44] Hussein SA, Tahir ASM, Izamshah R. Generated forces and heat during the critical stages of friction stir welding and processing. Journal of Mechanical Science and Technology. 2015;29:4319-28.

[45] Qian J, Li J, Sun F, Xiong J, Zhang F, Lin X. An analytical model to optimize rotation speed and travel speed of friction stir welding for defect-free joints. Scripta Materialia. 2012.

[46] Wanjara P, Monsarrat B, Larose S. Operational window for friction stir butt 6061-T6. Journal of Materials Processing Technology. 2013.

[47] Awang M, Raza SK, Yahaya MSA. Microstructure and hardness investigation of different welding passes in weld zones. Materialwissenschaft und Werkstofftechnik. 2015;46:492-7.

[48] He J, Ling Z, Li H. Effect of tool rotational speed on residual stress, microstructure, and tensile properties of friction stir welded 6061-T6 aluminum alloy thick plate. International Journal of Advanced Manufacturing Technology. 2015.

[49] Hou JC, Liu HJ, Zhao YQ. Influences of rotation speed on microstructures and mechanical properties of 6061-T6 aluminum alloy joints fabricated by self-reacting friction stir welding tool. International Journal of Advanced Manufacturing Technology. 2014;73:1073-9.

[50] Busu N, Jaffarullah MS, Saiful Bahari Shaari M, Low CY, Armansyah, Jaffar A. System integration of a friction stir welding machine with a customized traverse controlled table. Jurnal Teknologi. 2015;76:83-7.

[51] Chen S, Li H, Lu S, Ni R, Dong J. Temperature measurement and control of bobbin tool friction stir welding. International Journal of Advanced Manufacturing Technology. 2015:1-10.

[52] Shazly M, Sorour S, Alian AR. Effect of process control mode on weld quality of friction stir welded plates. Journal of Mechanical Science and Technology. 2016;30:267-78.

[53] Baratzadeh F, Widener CA, Lankarani HM, Burford DA. Methods to increase the fatigue life of friction stir lap welds in no-load transfer coupons using a retractable pin tool. Journal of ASTM International. 2012;9.

[54] Ding RJ. Retractable Pin-Tool Technology for Friction Stir Welding. ASM Proceedings of the International Conference: Trends in Welding Research1998. p. 585-9.

[55] Threadgill PL, Ahmed MMZ, Matin JP, Perrett JG, Wynee BP. The use of bobbin tools for friction stir welding of aluminum alloys. Thermec 2009. Berlin, Germany2009.

[56] D'Urso G, Giardini C. FEM model for the thermo-mechanical characterization of friction stir spot welded joints. International Journal of Material Forming. 2016;9:149-60.

[57] D'Urso G. Thermo-mechanical characterization of friction stir spot welded AA6060 sheets: Experimental and FEM analysis. Journal of Manufacturing Processes. 2015;17:108-19.

[58] D'Urso G, Giardini C. Thermo-mechanical characterization of Friction Stir Spot Welded sheets: Experimental and FEM comparison between AA6060 and AA7050 alloys. Key Engineering Materials2015. p. 1472-9.

[59] Gao Z, Wang P, Cheng D, Niu J, Sommitsch C. Numerical simulation of material flow in AA6082 during friction stir spot welding. Engineering Review. 2015;35:283-9.

[60] Lakshminarayanan AK, Annamalai VE, Elangovan K. Identification of optimum friction stir spot welding process parameters controlling the properties of low carbon automotive steel joints. Journal of Materials Research and Technology. 2014.

[61] Li Z, Yue Y, Ma L, Ji S, Chai P. Effects of pin geometry on the material flow behavior of friction stir spot welded 2A12 aluminum alloy. China Welding (English Edition). 2015;24:25-9.

[62] Miles MP, Karki U, Lee T, Hovanski Y. Prediction of joint line movement and temperatures in friction stir spot welding of DP 980 steel. TMS Annual Meeting2015. 229-40.

[63] Badarinarayan H, Yang Q, Okamoto K. Effect of weld orientation on static strength and failure mode of friction stir stitch welds in lap-shear specimens of aluminium 6022-T4 sheets. Fatigue and Fracture of Engineering Materials and Structures. 2011;34:908-20.

[64] Gao H, Jiao X. Study for technics of friction stitch welding. 2011 2nd International Conference on Mechanic Automation and Control Engineering, MACE 2011 - Proceedings2011. 5262-5.

[65] Okamoto K, Hunt F, Hirano S. Development of friction stir welding technique and machine for aluminum sheet metal assembly - Friction stir welding of aluminum for automotive applications (2) - Friction s. SAE Technical Papers. 2005.

[66] Rao HM, Jordon JB, Barkey ME, Badarinarayan H. Influence of structural integrity on fatigue behaviour of friction stir spot welding AZ31 Mg alloy. Material Science and Engineering. 2011.

[67] Chen Y, Chen J, Shalchi Amirkhiz B, Worswick MJ, Gerlich AP. Microstructures and properties of Mg alloy/DP600 steel dissimilar refill friction stir spot welds. Science and Technology of Welding and Joining. 2015;20:494-501.

[68] Lacki P, Derlatka A. Experimental and numerical investigation of aluminium lap joints made by RFSSW. Meccanica. 2016;51:455-62.

[69] Lacki P, Więckowski W, Wieczorek P. Assessment of joints using friction stir welding and refill friction stir spot welding methods. Archives of Metallurgy and Materials. 2015;60:2297-306.

[70] Li Z, Gao S, Ji S, Yue Y, Chai P. Effect of Rotational Speed on Microstructure and Mechanical Properties of Refill Friction Stir Spot Welded 2024 Al Alloy. Journal of Materials Engineering and Performance. 2016;25:1673-82.

[71] Li Z, Ji S, Ma Y, Chai P, Yue Y, Gao S. Fracture mechanism of refill friction stir spot-welded 2024-T4 aluminum alloy. International Journal of Advanced Manufacturing Technology. 2016:1-8.

[72] Okada H, Kamimuki K, Yoshikawa S, Fukada S. Refill Friction Spot Joining for Aerospace Application. SAE Technical Papers. 2015;2015-September.

[73] Wang X, Xu Y, Wei W, Luan G. Heat-flow coupling analysis of 6082-T6 with refill friction stir spot welding. Cailiao Kexue yu Gongyi/Material Science and Technology. 2015;23:120-4.

[74] Yue Y, Li Z, Ma Y, Chai P, Xing J. Effects of plunge depth on fracture behaviors of refill friction stir spot welding. Hsi-An Chiao Tung Ta Hsueh/Journal of Xi'an Jiaotong University. 2015;49:122-7.

[75] Waterloo Uo. Friction stir welding equipment - Center for advanced materials joining. https://uwaterloo.ca/centre-advanced-materials-joining/friction-stir-welding-equipment2016.

[76] Ji S, Meng X, Huang Y, Gao S, Lu H, Ma L. Effect of rotational velocity of tool on mechanical properties of stationary shoulder friction stir welding. Hanjie Xuebao/Transactions of the China Welding Institution. 2015;36:51-4.

[77] Ji S, Meng X, Ma L, Lu H, Gao S. Vertical compensation friction stir welding assisted by external stationary shoulder. Materials and Design. 2015;68:72-9.

[78] Ji SD, Meng XC, Li ZW, Ma L, Gao SS. Experimental Study of Stationary Shoulder Friction Stir Welded 7N01-T4 Aluminum Alloy. Journal of Materials Engineering and Performance. 2016;25:1228-36.

[79] Jiang X, Wynne BP, Martin J. Microstructure and texture evolution of stationary shoulder friction stir welded Ti6Al4v alloy. Science and Technology of Welding and Joining. 2015;20:594-600.

[80] Li D, Yang X, Cui L, He F, Zhang X. Investigation of stationary shoulder friction stir welding of aluminum alloy 7075-T651. Journal of Materials Processing Technology. 2015;222:391-8.

[81] Li Z, Yue Y, Ji S, Chai P, Zhou Z. Joint features and mechanical properties of friction stir lap welded alclad 2024 aluminum alloy assisted by external stationary shoulder. Materials and Design. 2016;90:238-47.

[82] Sun Z, Yang X, Li D, Cui L. The local strength and toughness for stationary shoulder friction stir weld on AA6061-T6 alloy. Materials Characterization. 2016;111:114-21.

[83] Thomas WM, Kallee SW, Staines DG, Oakley PJ. Friction stir welding - process variants and developements in the automotive industry. SAE World Congress. Detroit, MI2006.

[84] Li JQ, Liu HJ. Effects of the reversely rotating assisted shoulder on microstructures during the reverse dual-rotation friction stir welding. Journal of Materials Science and Technology. 2015;31:375-83.

[85] Montazerolghaem H, Badrossamay M, Tehrani AF, Rad SZ, Esfahani MS. Dual-rotation speed friction stir welding: Experimentation and modeling. Materials and Manufacturing Processes. 2015;30:1109-14.

[86] Shi L, Wu CS, Liu HJ. Modeling the material flow and heat transfer in reverse dual-rotation friction stir welding. Journal of Materials Engineering and Performance. 2014;23:2918-29.

[87] Shi L, Wu CS, Liu HJ. Analysis of heat transfer and material flow in reverse dual-rotation friction stir welding. Welding in the World. 2015;59:629-38.

[88] Shi L, Wu CS, Liu HJ. The effect of the welding parameters and tool size on the thermal process and tool torque in reverse dual-rotation friction stir welding. International Journal of Machine Tools and Manufacture. 2015;91:1-11.

[89] Norris IM, Thomas WM, Staines DG. Friction stir welding - process variants and recent industrial developments. 10th International Aachen Welding Conference. Aachen2007.

[90] Leonard AJ, Lochyer, S.A. Flaws in friction stir welds. 2003.

[91] Chen G, Shi Q, Feng Z. On the material behavior at tool/workpiece interface during friction stir welding: A CFD based numerical study. TMS Annual Meeting2015. 251-8.

[92] Chen GQ, Shi QY, Li YJ, Sun YJ, Dai QL, Jia JY, et al. Computational fluid dynamics studies on heat generation during friction stir welding of aluminum alloy. Computational Materials Science. 2013;79:540-6.

[93] Colegrove PA, Shercliff HR. 3-Dimensional CFD modelling of flow round a threaded friction stir welding tool profile. Journal of Materials Processing Technology. 2005;169:320-7.

[94] Hasan AF, Bennett CJ, Shipway PH. A numerical comparison of the flow behaviour in Friction Stir Welding (FSW) using unworn and worn tool geometries. Materials & Design. 2015;87:1037-46.

[95] Jacquin D, de Meester B, Simar A, Deloison D, Montheillet F, Desrayaud C. A simple Eulerian thermomechanical modeling of friction stir welding. Journal of Materials Processing Technology. 2011;211:57-65.

[96] Li Z, Yue Y, Ji S, Peng C, Wang L. Optimal design of thread geometry and its performance in friction stir spot welding. Materials & Design. 2016;94:368-76.

[97] Padmanaban R, Ratna Kishore V, Balusamy V. Numerical simulation of temperature distribution and material flow during friction stir welding of dissimilar aluminum alloys. Procedia Engineering2014. 854-63.

[98] Su H, Wu CS, Bachmann M, Rethmeier M. Numerical modeling for the effect of pin profiles on thermal and material flow characteristics in friction stir welding. Materials & Design. 2015;77:114-25.

[99] Zhang J, Shen Y, Li B, Xu H, Yao X, Kuang B, et al. Numerical simulation and experimental investigation on friction stir welding of 6061-T6 aluminum alloy. Materials & Design. 2014;60:94-101.

[100] Bussetta P, Dialami N, Boman R, Chiumenti M, Agelet De Saracibar C, Cervera M, et al. Comparison of a fluid and a solid approach for the numerical simulation of friction stir welding with a non-cylindrical pin. Steel Research International. 2014;85:968-79.

[101] Buffa G, Ducato A, Fratini L. Numerical procedure for residual stresses prediction in friction stir welding. Finite Elements in Analysis and Design. 2011;47:470-6.

[102] Carlone P, Citarella R, Sonne MR, Hattel JH. Multiple Crack Growth Prediction in AA2024-T3 Friction Stir Welded Joints, Including Manufacturing Effects. International Journal of Fatigue.

[103] Hrennikoff A. Solution of problems of elasticity by the framework method. Journal of applied mechanics. 1941;8:169-75.

[104] Feulvarch E, Roux JC, Bergheau JM. A simple and robust moving mesh technique for the finite element simulation of Friction Stir Welding. Journal of Computational and Applied Mathematics. 2013;246:269-77.

[105] Buffa G, Fratini L, Shivpuri R. Finite element studies on friction stir welding processes of tailored blanks. Computers & Structures. 2008;86:181-9.

[106] Buffa G, Hua J, Shivpuri R, Fratini L. A continuum based fem model for friction stir welding—model development. Materials Science and Engineering: A. 2006;419:389-96.

[107] Dialami N, Chiumenti M, Cervera M, Agelet de Saracibar C. An apropos kinematic framework for the numerical modeling of friction stir welding. Computers & Structures. 2013;117:48-57.

[108] Guerdoux S, Fourment L. A 3D numerical simulation of different phases of friction stir welding. Modelling and Simulation in Materials Science and Engineering. 2009;17.

[109] Sadeghi S, Najafabadi MA, Javadi Y, Mohammadisefat M. Using ultrasonic waves and finite element method to evaluate through-thickness residual stresses distribution in the friction stir welding of aluminum plates. Materials & Design. 2013;52:870-80.

[110] Yaduwanshi DK, Bag S, Pal S. Numerical modeling and experimental investigation on plasma-assisted hybrid friction stir welding of dissimilar materials. Materials & Design. 2016;92:166-83.

[111] Grujicic M, He T, Arakere G, Yalavarthy HV, Yen C-F, Cheeseman BA. Fully coupled thermomechanical finite element analysis of material evolution during friction-stir welding of AA5085. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture. 2009.

[112] Belytschko T, Lu YY, Gu L. Element-free Galerkin methods. International Journal for Numerical Methods in Engineering. 1994;37:229-56.

[113] Onate E, Idelsohn SR, Del Pin F, Aubry R. The particle finite element method, an overview. International Journal for Computational Methods. 2004;1:267-307.

[114] Liu GR, Gu YT. A point interpolation method. Prodeedings of 4th Asia-Pacific Conference on Computational Mechanics. Singapore1999. 1009-14.

[115] Traversoni L. Natural neighbor finite elements. International conference on hydraulic engineering software1994. 291-7.

[116] Harlow FH. A machine calculation method for hydrodynamics problems. Los Alamos Scientific Laboratory; 1955.

[117] Sulsky D, Chen Z, Schreyer HL. A particle method for history dependent materials. Computer Methods in Applied Mechanics and Engineering. 1994;118:179-96.

[118] Alfaro I, Bel D, Cueto E, Doblaré M, Chinesta F. Three-dimensional simulation of aluminium extrusion by the α-shape based natural element method. Computer Methods in Applied Mechanics and Engineering. 2006;195:4269-86.

[119] Alfaro Iea. Meshless simulation of friction stir welding.

[120] Wu CT, Hu W, Wang HP. An adaptive meshfree galerkin method for 3D thermo-mechanical flow simulation of FSW. 13th International LS-DYNA Users Conference. Dearborn Michigan: LSTC; 2014.

[121] Fagan T, Lemiale V, Nairn J, Ahuja Y, Ibrahim R, Estrin Y. Detailed thermal and material flow analyses of friction stir forming using a three-dimensional particle based model. Journal of Materials Processing Technology. 2016;231:422-30.

[122] Miyasaka F, Yoshikawa G, Matsuzawa S. Numerical simulation model for FSW employing particle method -effect of tool angle on fluid motion. Advanced Materials Research2014. 1765-9.

[123] Onate E, Idelsohn S, Zienkiewicz OC, Taylor RL. A FINITE POINT METHOD IN COMPUTATIONAL MECHANICS. APPLICATIONS TO CONVECTIVE TRANSPORT AND FLUID FLOW. International Journal for Numerical Methods in Engineering. 1996;39:3839-66.

[124] Altluri SN, Zhu T. A new meshless local petrov-galerkin approach in computational mechanics. Comtut Mech. 1998;22:117-27.

[125] Gingold RA, Monaghan JJ. Smoothed particle hydrodynamics: theory and application to non-spherical stars. Mon Not R Astron Soc. 1977 181 375–89.

[126] Lucy LB. A numerical approach to the testing of the fission hypothesis. Astron. 1977;J. 82 1013–24.

[127] Libersky L, Petschek AG. Smoothed particle hydrodynamics with strength of materials. The Next Free Lagrange Conference. NY: Springer-Verlag; 1991.

[128] Fraser K, St-Georges L, Kiss LI. Prediction of defects in a friction stir welded joint using the Smoothed Particle Hydrodynamics Method. Procedings of the 7th Asia Pacific IIW International Congress on Recent Development in Welding and Joining Methods. Singapore2013.

[129] Fraser K, St-Georges L, Kiss LI. Smoothed Particle Hydrodynamics Numerical Simulation of Bobbin Tool Friction Stir Welding. 10th International Friction Stir Welding Symposium. Beijing China: TWI; 2014.

[130] Vigh LG, St-Georges L, Kiss LI, Fraser K. FSW hegesztett aluminium palyalemez - Technologia, analizis es meretezes. femszerkezetek. 2014;III:31-5.

[131] Bohjwani S. Smoothed Particle Hydrodynamics Modeling of the Friction Stir Welding Process: University of Texas at El Paso; 2007.

[132] Pan W, Li D, Tartakovsky AM, Ahzi S, Khraisheh M, Khaleel M. A new smoothed particle hydrodynamics non-Newtonian model for friction stir welding: Process modeling and simulation of microstructure evolution in a magnesium alloy. International Journal of Plasticity. 2013;48:189-204.

[133] Tartakovsky AM, Grant G, Sun X, Khaleel M. Modeling of friction stir welding process with smooth particle hydrodynamics. SAE International. 2006.

[134] Timesli A, Zahrouni H, Braikat B, Moufki A, Lahmam H. Numerical model based on SPH method to simulate friction stir welding. Revue de Mechanique Appliquee et Theorie. 2011;2:537-46.

[135] Chiumenti M, Cervera M, Agelet de Saracibar C, Dialami N. Numerical modeling of friction stir welding processes. Computer Methods in Applied Mechanics and Engineering.

[136] Gemme F, Verreman Y, Dubourg L, Jahazi M. Numerical analysis of the dwell phase in friction stir welding and comparison with experimental data. Materials Science and Engineering: A. 2010;527:4152-60.

[137] Hilgert J, Schmidt H, dos Santos JF. Bobbin tool FSW - A moving geometry model. Comsol conference 2009. Milan2009.

[138] Chiumenti M, Cervera M, Agelet de Saracibar C, Dialami N. Numerical modeling of friction stir welding processes. Computer Methods in Applied Mechanics and Engineering. 2013;254:353-69.

[139] Assidi M, Fourment L, Guerdoux S, Nelson T. Friction model for friction stir welding process simulation: Calibrations from welding experiments. International Journal of Machine Tools and Manufacture. 2010;50:143-55.

[140] Dorbane A, Ayoub G, Mansoor B, Hamade R, Kridli G, Imad A. Observations of the mechanical response and evolution of damage of AA 6061-T6 under different strain rates and temperatures. Materials Science and Engineering: A. 2015;624:239-49.

[141] Kuykendall K, Nelson T, Sorensen C. On the selection of constitutive laws used in modeling friction stir welding. International Journal of Machine Tools and Manufacture. 2013;74:74-85.

[142] Schwer L. Aluminum plate performation: a comparative case study using lagrange with erosion, multi-material ALE, and SPH. 7th European LS-DYNA Conference. Germany: Dynamore GmbH; 2009.

[143] Kuykendall K. An Evaluation of Constitutive Laws and their Ability to Predict Flow Stress over Large Variations in Temperature, Strain, and Strain Rate Characteristic of Friction Stir Weldin: Brigham Young University - Provo.

[144] MatWeb. AA6061-T6 Material Properties. http://www.matweb.com/search/DataSheet.aspx?MatGUID=1b8c06d0ca7c456694c7777d9e1 0be5b&ckck=12016.

[145] Johnson GR, Cook, W.H. A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures. Proceedings of the 7th international symposium on ballistics. 1983.

[146] Pan W, Bao J, Tartakovsky AM. Smoothed particle hydrodynamics continuous boundary force method for Navier–Stokes equations subject to a Robin boundary condition. Journal of Computational Physics. 2014;259:242-59.

[147] Pan W, Tartakovsky AM, Monaghan JJ. Smoothed particle hydrodynamics non-Newtonian model for ice-sheet and ice-shelf dynamics. Journal of Computational Physics. 2013;242:828-42.

[148] Wang H, Colegrove PA, dos Santos JF. Numerical investigation of the tool contact condition during friction stir welding of aerospace aluminium alloy. Computational Materials Science. 2013;71:101-8.

[149] Pineau F, D'Amours, G. Applicaiton of LS-DYNA SPH formulation to model semi-solid metal casting - Skin layers. 8th European LS-DYNA Users Conference. 2011.

[150] Zhang Z. Comparison of two contact models in the simulation of friction stir welding process. Journal of Materials Science. 2008;43:5867-77.

[151] Fraser K, St-Georges L, Kiss L. Optimization of Friction Stir Welding Tool Advance Speed via Monte-Carlo Simulation of the Friction Stir Welding Process. Materials. 2014;7:3435-52.

[152] Ammouri AH, Kheireddine AH, Kridli GT, Hamade RF. FEM optimization of process parameters and in-process cooling in the friction stir processing of magnesium alloy AZ31B. ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)2013.

[153] Elatharasan G, Kumar VSS. Modelling and Optimization of Friction Stir Welding Parameters for Dissimilar Aluminium Alloys Using RSM. Procedia Engineering. 2012;38:3477-81.

[154] Boulahem K, Salem SB, Bessrour J. Surface roughness model and parametric welding optimization in friction stir welded AA2017 using Taguchi method and response surface methodology. Lecture Notes in Control and Information Sciences2015. 83-93.

[155] Mohamed MA, Manurung YHP, Berhan MN. Model development for mechanical properties and weld quality class of friction stir welding using multi-objective Taguchi method and response surface methodology. Journal of Mechanical Science and Technology. 2015;29:2323-31.

[156] Panneerselvam K, Lenin K. Parameters optimization in FSW of polypropylene based on RSM. Multidiscipline Modeling in Materials and Structures. 2015;11:32-42.

[157] Urbano LO, Gordillo M, Franco F. Optimization of friction stir welding of aluminum alloy AA 6261-T5 through application of response surface methodology. Revista Latinoamericana de Metalurgia y Materiales. 2014;34:55-65.

[158] Vijayan D, Seshagiri Rao V. A parametric optimization of FSW process using RSM based grey relational analysis approach. International Review of Mechanical Engineering. 2014;8:328-37.

[159] Lakshminarayanan AK, Balasubramanian V. Process parameters optimization for friction stir welding of RDE-40 aluminium alloy using Taguchi technique. Transactions of Nonferrous Metals Society of China. 2008;18:548-54.

Kirk Fraser - April 2017

[160] Mohd Hanapi MH, Hussain Z, Almanar IP, Abu Seman A. Optimization processing parameter of 6061-T6 alloy friction stir welded using Taguchi technique. Materials Science Forum2016. 294-8.

[161] Senthilraja R, Sait AN. Optimization of the parameters of friction stir welding for AZ91D magnesium alloy using the Taguchi design. Materials Science. 2015;51:180-7.

[162] Boulahem K, Salem SB, Bessrour J. Surface roughness model and parametric welding optimization in friction stir welded AA2017 using taguchi method and response surface methodology. Lecture Notes in Mechanical Engineering2015. 83-93.

[163] Ghetiya ND, Patel KM, Kavar AJ. Multi-objective Optimization of FSW Process Parameters of Aluminium Alloy Using Taguchi-Based Grey Relational Analysis. Transactions of the Indian Institute of Metals. 2015.

[164] Kumar S, Kumar S. Multi-response optimization of process parameters for friction stir welding of joining dissimilar Al alloys by gray relation analysis and Taguchi method. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2015;37:665-74.

[165] Mustafa FF, Kadhym AH, Yahya HH. Tool geometries optimization for friction stir welding of AA6061-T6 aluminum alloy T-joint using taguchi method to improve the mechanical behavior. Journal of Manufacturing Science and Engineering, Transactions of the ASME. 2015;137.

[166] Parida B, Pal S. Fuzzy assisted grey Taguchi approach for optimisation of multiple weld quality properties in friction stir welding process. Science and Technology of Welding and Joining. 2015;20:35-41.

[167] Shinde VR, Sapre M, Jatti VS. Optimization of friction stir welding process parameters using taguchi's method. Global Journal of Pure and Applied Mathematics. 2015;11:777-88.

[168] Zhang H, Liu H. Mathematical model and optimization for underwater friction stir welding of a heat-treatable aluminum alloy. Materials & Design. 2013;45:206-11.

[169] Heidarzadeh A, Saeid T, Khodaverdizadeh H, Mahmoudi A, Nazari E. Establishing a Mathematical Model to Predict the Tensile Strength of Friction Stir Welded Pure Copper Joints. Metallurgical and Materials Transactions B. 2012;44:175-83.

[170] Cocchetti G, Pagani M, Perego U. Selective mass scaling for distorted solid-shell elements in explicit dynamics: Optimal scaling factor and stable time step estimate. International Journal for Numerical Methods in Engineering. 2015;101:700-31.

[171] Confalonieri F, Ghisi A, Perego U. 8-Node solid-shell elements selective mass scaling for explicit dynamic analysis of layered thin-walled structures. Computational Mechanics. 2015;56:585-99.

[172] De Frías GJ, Aquino W, Pierson KH, Heinstein MW, Spencer BW. A multiscale mass scaling approach for explicit time integration using proper orthogonal decomposition. International Journal for Numerical Methods in Engineering. 2014;97:799-818.

[173] Tkachuk A, Bischoff M. Variational methods for selective mass scaling. Computational Mechanics. 2013;52:563-70.

[174] Tkachuk A, Bischoff M. Local and global strategies for optimal selective mass scaling. Computational Mechanics. 2014;53:1197-207.

[175] Tkachuk A, Bischoff M. Direct and sparse construction of consistent inverse mass matrices: General variational formulation and application to selective mass scaling. International Journal for Numerical Methods in Engineering. 2015;101:435-69.

[176] Limido J, Espinosa C, Salaün M, Lacome JL. SPH method applied to high speed cutting modelling. International Journal of Mechanical Sciences. 2007;49:898-908.

[177] Villumsen MF, Fauerholdt TG. Simulation of metal cutting using smoothed particle hydrodynamics. 7th International LS-DYNA Conference. Bmberg2008.

[178] Monaghan JJ. Smoothed Particle Hydrodynamics - A Review 1977 to 2005. Rep Prog Phys. 2005;68 1703.

[179] Gingold RA, Monaghan JJ. Kernel estimates as a basis for general particle methods in hydrodynamics. Journal of Computational Physics. 1982;46:429-53.

[180] Reddy JN. An Introduction to the Finite Element Method - Third Edition. New York: McGraw Hill; 2006.

[181] Liu GR. Meshfree methods - Moving beyond the finite element method. Boca Raton Fl. USA: CRC Press; 2010.

[182] Liu MB, Liu GR, Lam KY. Constructing smoothing functions in smoothed particle hydrodynamics with applications. Journal of Computational and Applied Mathematics. 2003;155:263-84.

[183] Liu GR, Liu MB. Smoothed particle hydrodynamics : a meshfree particle method. Hackensack, New Jersey: World Scientific; 2003.

[184] Monaghan JJ. An introduction to SPH. Computer Physics Communications. 1988;48:89-96.

[185] Bonet J, Lok L. Variational and momentum preservation aspspects of Smoothed Particle Hydrodynamic formulations. Comput Methods Appl Mech Engrg. 1999;180, 97-115.

[186] Chen JK, Beraun JE. A generalized smoothed particle hydrodynamics method for nonlinear dynamic problems. Computer Methods in Applied Mechanics and Engineering. 2000;190:225-39.

[187] Cleary PW. Extension of SPH to predict feeding, freezing and defect creation in low pressure die casting. Applied Mathematical Modelling. 2010;34:3189-201.

[188] Cleary PW, Prakash M, Das R, Ha J. Modelling of metal forging using SPH. Applied Mathematical Modelling. 2012;36:3836-55.

[189] Das R, Cleary, P.W. Modelling plastic deformation and thermal response in welding using smoothed particle hydrodynamics. 16th Australasian Fluid Mechanics Conference. Australia2007.

[190] Randles PW, Libersky LD. Smoothed Particle Hydrodynamics: Some recent improvements and applications. Computer Methods in Applied Mechanics and Engineering. 1996;139:375-408.

[191] Wendland H. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Advances in computational mechanics. 1995;4(1) 389-396.

[192] Ismail EB, Reddy BD. Smoothed Particle Hydrodynamics for Nonlinear Solid Mechanics. 2009.

[193] Johnson GR, Beissel SR. Normalized smoothing functions for sph impact computations. International Journal for Numerical Methods in Engineering. 1996;39:2725-41.

[194] Yang X, Liu M, Peng S. Smoothed particle hydrodynamics modeling of viscous liquid drop without tensile instability. Computers and Fluids. 2014;92:199-208.

[195] Zhou X, Tamma KK. On the applicatibility of stress update formulations for corotational stress rate hypoelasticity constitutive models. Finite Elements in Analysis and Design. 2003;39:783-816.

[196] Pinsky PM, Ortiz M, Pister KS. Numerical integration of rate constitutive equations in finite deformation analysis. Computer Methods in Applied Mechanics and Engineering. 1983;40:137-58.

[197] Bazant ZP, Vorel J. Objective stress rates in finite strain of inelastic solid and their energy consistency. Evanston, Illinois, USA: Northwestern University; 2012.

[198] Anandarajah A, SpringerLink (Service en ligne). Computational Methods in Elasticity and Plasticity [ressource électronique] : Solids and Porous Media. First. ed. New York, NY: Springer Science+Business Media, LLC; 2010.

[199] Banerjee B. An evaluation of plastic flow stress models for the simulation of high-temperature and high strain-rate deformation of metals. Acta Materialia. 2006;58:6810-27.

[200] Berstad T, Hopperstad, O.S., Langseth, M. Elasto-Viscoplastic constitutive models in the explicit finite element code LS-DYNA3D. Norwegian Institute of Technology.

[201] Mitsoulis E. Flows of viscoplastic materials: models and computations. In: Binding DM, Hudson NE, Keunings R, editors. Reology Reviews. London, UK: Brit. Soc. Rheol.; 2007. 135-78.

[202] Zaera R, Fernandez-Saez J. An implicit consistent algorithm for the integration of thermoviscoplastic consitutive equations in adiabatic conditions and finite deformations. International Journal of Solids and Structures. 2006;43:1594-612.

[203] Price D. Smoothed Particle Hydrodynamics and Magnetohydrodynamics. 2004.

[204] Von Neumann J, Richtmyer RD. A method for the numerical calculation of hydrodynamic shocks. Journal of Applied Physics. 1950;21:232-7.

[205] Monaghan JJ, Gingold RA. Shock simulation by the particle method SPH. Journal of Computational Physics. 1983;52:374-89.

[206] Monaghan JJ. On the problem of penetration in particle methods. Journal of Computational Physics. 1989;82:1-15.

[207] Xu J. Smoothed particle hydrodynamics in LS-DYNA. Livermore, California: LSTC; 2011.

[208] Libersky LD, Petschek AG, Carney TC, Hipp JR, Allahdadi FA. High Strain Lagrangian Hydrodynamics: A Three-Dimensional SPH Code for Dynamic Material Response. Journal of Computational Physics. 1993;109:67-75.

[209] Libersky LD, Randles PW, Carney TC, Dickinson DL. Recent improvements in SPH modeling of hypervelocity impact. International Journal of Impact Engineering. 1997;20:525-32.

[210] Bonet J, Kulasegaram S. A simplified approach to enhance the performance of smooth particle hydrodynamics methods. Applied Mathematics and Computation. 2002;126:133-55.

[211] Monaghan JJ. SPH without a Tensile Instability. Journal of Computational Physics. 2000;159:290-311.

[212] Swegle JW, Hicks DL, Attaway SW. Smoothed Particle Hydrodynamics Stability Analysis. Journal of Computational Physics. 1995;116:123-34.

[213] Vignjevic R, Reveles JR. SPH in a total Lagrangian formalism. Computer Modelling in Engineering and Science. 2006;14:181-91.

[214] J.P. Gray JJM, R.P. Swift. SPH elastic dynamics. Computer Methods in Applied Mechanics and Engineering. 2000.

[215] Jubelgas M. Cosmological Hydrodynamics: Thermal Conduction and Cosmic Rays. München: Ludwig-Maximilians Universität München; 2007.

[216] Rook R, Yildiz M, Dost S. Modeling Transient Heat Transfer Using SPH and Implicit Time Integration. Numerical Heat Transfer, Part B: Fundamentals. 2007;51:1-23.

[217] Xu J. Heat transfer with explicit sph method in LS-DYNA. 12th International LS-DYNA Users Conference. 2012;Computing Technologies.

[218] Cleary PW, Monaghan JJ. Conduction Modelling Using Smoothed Particle Hydrodynamics. Journal of Computational Physics. 1998;148.

[219] Incropera FP. Fundamentals of heat and mass transfer. 6th ed. Hoboken, N.J.: J. Wiley; 2007.

[220] Marrone S, Colagrossi A, Le Touzé D, Graziani G. Fast free-surface detection and level-set function definition in SPH solvers. Journal of Computational Physics. 2010;229:3652-63.

[221] Ericson C. Real Time Collision Detection San Fransisco California: Morgan Kauffmann; 2005.

[222] A. V. Khomenko IAL. A stochastic model of stick-slip boundary friction with account for the deformation effect of the shear modulus of the lubricant. Journal of Friction and Wear. 2010.

[223] Sakamoto T. Normal displacement and dynamic friction characteristics in a stick-slip process. Tribology International. 1987;20:25-31.

[224] Schneider J, Beshears R, Nunes AC. Interfacial sticking and slipping in the friction stir welding process. Materials Science and Engineering: A. 2006;435-436:297-304.

[225] Vissel Y. Frictional stick-slip oscillation as a first-passage problem. EPL Journal. 2011.

[226] Durdanovic MBea. Heat generation during friction stir welding process. Tribology in industry. 2009;31.

[227] LSTC. LS-DYNA Materials User Manual Volume 2. 2012.

[228] Majzoobi GH, Dehgolan FR. Determination of the constants of damage models. Engineering Procedia. 2011:764-73.

[229] Schmidt H, Hattel JH. Thermal and material flow modelling of friction stir welding using Comsol. Comsol Conference 2008. Hannover2008.

[230] Wang J, Su J, Mishra RS, Xu R, Baumann JA. Tool wear mechanisms in friction stir welding of Ti–6Al–4V alloy. Wear. 2014;321:25-32.

[231] Hartman T, Miles MP, Hong ST, Steel R, Kelly S. Effect of PCBN tool grade on joint strength and tool life in friction stir spot welded DP 980 steel. Wear. 2015;328-329:531-6.

[232] Luo J, Wang H, Chen W, Li L. Study on anti-wear property of 3D printed-tools in friction stir welding by numerical and physical experiments. International Journal of Advanced Manufacturing Technology. 2015;77:1781-91.

[233] Yasavol N, Ramalho A. Wear properties of friction stir processed AISI D2 tool steel. Tribology International. 2015;91:177-83.

[234] Zmitrowitz A. Wear patternal and laws of wear - a review. Journal of theorectical and applied mechanics. 2006.

[235] Cleary PW, Ha J, Prakash M, Nguyen T. 3D SPH flow predictions and validation for high pressure die casting of automotive components. Applied Mathematical Modelling. 2006;30:1406-27.

[236] Das R, Cleary PW. Evaluation of Accuracy and Stability of the Classical SPH Method Under Uniaxial Compression. Journal of Scientific Computing. 2014:1-40.

[237] Prakash M, Cleary PW. Three dimensional modelling of lava flow using Smoothed Particle Hydrodynamics. Applied Mathematical Modelling. 2011;35:3021-35.

[238] Chapman B, Jost G, van der Pas R. Using OpenMP - Portable shared memory parallel programmin. Cambridge, Massachusetts, USA: The MIT Press; 2008.

[239] Gropp W, Lusk E, Skjellum A. Using MPI - Portable parallel programming with the message-passing interface second edition. Cambridge, Massachusetts, USA: The MIT Press; 1999.

[240] Anderson JA, Lorenz CD, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. Journal of Computational Physics. 2008;227:5342-59.

[241] Balfour J. Introduction to CUDA. CME343: NVIDIA; 2011.

[242] Cercos-Pita JL. AQUAgpusph, a new free 3D SPH solver accelerated with OpenCL. Computer Physics Communications. 2015;192:295-312.

[243] Delbosc N, Summers JL, Khan AI, Kapur N, Noakes CJ. Optimized implementation of the Lattice Boltzmann Method on a graphics processing unit towards real-time fluid simulation. Computers & Mathematics with Applications. 2014;67:462-75.

[244] Govender N, Wilke DN, Kok S. Collision detection of convex polyhedra on the NVIDIA GPU architecture for the discrete element method. Applied Mathematics and Computation.

[245] Group TP. CUDA Fortran Programming Guide and Reference - Release 2014. OR, USA: NVIDIA; 2014.

[246] Harada T, Koshizuka S, Kawaguchi Y. Smoothed particle hydrodynamics on GPUs. . Proc Comput Graph Intl2007. p. 63–70.

[247] Hérault A. SPH on GPU with CUDA. Journal of Hydraulic Research. 2009;48:000.

[248] Hérault A, Bilotta G, Vicari A, Rustico E, Negro CD. Numerical simulation of lava flow using a GPU SPH model. Annals of Geophysics. 2011;54.

[249] NVIDIA. Introduction to CUDA Fortran. GTC 2103: NVIDIA; 2013.

[250] Ruetsch G, Fatica M. CUDA Fortran for Scientists and Engineers. Waltham, MA, USA: Elsevier Inc.; 2014.

[251] Rustico E, Bilotta G, Herault A, Del Negro C, Gallo G. Advances in multi-GPU smoothed particle hydrodynamics simulations. IEEE Transactions on Parallel and Distributed Systems. 2014;25:43-52.

[252] Valdez-Balderas D, Domínguez JM, Rogers BD, Crespo AJC. Towards accelerating smoothed particle hydrodynamics simulations for free-surface flows on multi-GPU clusters. Journal of Parallel and Distributed Computing. 2013;73:1483-93.

[253] Wang Y, Dou Y, Guo S, Lei Y, Zou D. CPU-GPU hybrid parallel strategy for cosmological simulations. Concurrency Computation Practice and Experience. 2014;26:748-65.

[254] Wilt N. The CUDA Handbook - A Comprehensive Guide to GPU Programming. Indiana, USA: Addison-Wessley; 2013.

[255] Xia X, Liang Q. A GPU-accelerated smoothed particle hydrodynamics (SPH) model for the shallow water equations. Environmental Modelling and Software. 2016;75:28-43.

[256] Crespo AJC, Domínguez JM, Gesteira MG, Rogers BD, Longshaw S, Canelas R, et al. User guide for DualSPHysics code: University of Manchester; 2013.

[257] Borvik T, Olovsson L, Hanssen AG. Impetus AFEA Solver - command manual. Norway: Impetus; 2016.

[258] Sutmann G, Stegailov V. Optimization of neighbor list techniques in liquid matter simulations. Journal of Molecular Liquids. 2006;125:197-203.

[259] Everaers R, Kremer K. A fast grid search algorithm for molecular dynamics simulations with short-range interactions. Computer Physics Communications. 1994;81, 19-55.

[260] Du S, Kanai T. GPU-based adaptive surface reconstruction for real-time SPH fluids. 22nd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, WSCG 2014, Full Papers Proceedings - in co-operation with EUROGRAPHICS Association2014. 141-50.

[261] Jin S, Zheng X, Duan W. Viscosity flow simulation using improved SPH method based on GPU parallel calculation. Harbin Gongcheng Daxue Xuebao/Journal of Harbin Engineering University. 2015;36:1011-8.

[262] Joselli M, Junior JRDS, Clua EW, Montenegro A, Lage M, Pagliosa P. Neighborhood grid: A novel data structure for fluids animation with GPU computing. Journal of Parallel and Distributed Computing. 2015;75:20-8.

[263] Mokos A, Rogers BD, Stansby PK, Domínguez JM. Multi-phase SPH modelling of violent hydrodynamics on GPUs. Computer Physics Communications. 2015;196:304-16.

[264] Nie X, Chen L, Xiang T. Real-time incompressible fluid simulation on the GPU. International Journal of Computer Games Technology. 2015;2015.

[265] Nie X, Chen L, Xiang T. A gpu-based method for weakly compressible fluids. Proceedings - 17th IEEE International Conference on Computational Science and Engineering, CSE 2014, Jointly with 13th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2014, 13th International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2014 and 8th International Conference on Frontier of Computer Science and Technology, FCST 20142015. p. 121-6.

[266] Nishiura D, Furuichi M, Sakaguchi H. Computational performance of a smoothed particle hydrodynamics simulation for shared-memory parallel computing. Computer Physics Communications. 2015;194:18-32.

[267] Sun H, Tian Y, Zhang Y, Wu J, Wang S, Yang Q, et al. A Special Sorting Method for Neighbor Search Procedure in Smoothed Particle Hydrodynamics on GPUs. Proceedings of the International Conference on Parallel Processing Workshops2015. 81-5.

[268] Awile O, Büyükkeçeci F, Reboux S, Sbalzarini IF. Fast neighbor lists for adaptive-resolution particle simulations. Computer Physics Communications. 2012;183:1073-81.

[269] Domínguez JM, Crespo AJC, Valdez-Balderas D, Rogers BD, Gómez-Gesteira M. New multi-GPU implementation for smoothed particle hydrodynamics on heterogeneous clusters. Computer Physics Communications. 2013;184:1848-60.

[270] Hoetzlein RC. Fast fixed-radius nearest neighbors: Interactive million-particle fluids. NVIDIA GPU Technology Conference. San Jose, Ca.2014.

[271] Pelfrey B, House D. Adaptive neighbor pairing for smoothed particle hydrodynamics. 2010. p. 192-201.

[272] Fraser K. Adaptive smoothed particle hydrodynamics neighbor search algorithm for large plastic deformation computational solid mechanics. 13th International LS-DYNA Users Conference. Dearborn Michigan: LSTC; 2014.

[273] Ferrer JCS, Chevére DG, Manian V. Photorealistic image synthesis and camera validation from 2D images. Proceedings of SPIE - The International Society for Optical Engineering2014.

[274] Santamaria-Ibirika A, Cantero X, Huerta S, Santos I, Bringas PG. Procedural playable cave systems based on voronoi diagram and delaunay triangulation. Proceedings - 2014 International Conference on Cyberworlds, CW 20142014. 15-22.

[275] Gomit G, Chatellier L, Calluaud D, David L, Fréchou D, Boucheron R, et al. Large-scale free surface measurement for the analysis of ship waves in a towing tank. Experiments in Fluids. 2015;56.

[276] Chen J, Zheng G, Luo Z. A method of automatic iso-surface segmentation for NC machining. Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics. 2015;27:924-9.

[277] Gonçalves G, Duro N, Sousa E, Figueiredo I. Automatic extraction of tide-coordinated shoreline using open source software and Landsat imagery. International Archives of the

Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. 7W3 ed2015. 953-7.

[278] Adhyapak SM, Menon PG, Rao Parachuri V. Restoration of optimal ellipsoid left ventricular geometry: Lessons learnt from in silico surgical modelling. Interactive Cardiovascular and Thoracic Surgery. 2014;18:153-8.

[279] De Goes F, Liu B, Budninskiy M, Tong Y, Desbrun M. Discrete 2-tensor fields on triangulations. Computer Graphics Forum. 2014;33:13-24.

[280] Dupuis J, Kuhlmann H. High-precision surface inspection: Uncertainty evaluation within an accuracy range of 15μm with triangulation-based laser line scanners. Journal of Applied Geodesy. 2014;8:109-18.

[281] Ibraheem F, Hussain MZ, Bhatti AA. C<sup>1</sup> positive surface over positive scattered data sites. PLoS ONE. 2015;10.

[282] Kwintiana B, Roller D, Hasni R, Snasel V. Free-form surface reconstruction from 3D multi-resolution data on spherical space. World Automation Congress Proceedings2014. p. 923-9.

[283] Sharma G, Sharma A. Determining the angle and depth of puncture for fluoroscopy-guided percutaneous renal access in the prone position. Indian Journal of Urology. 2015;31:38-41.

[284] Mills KC. Recommended values of thermophysical proterties for commercial alloys. Cambridge, UK: Woodhead Publishing Ltd.; 2002.

[285] BSI. Eurocode 9 - Design of aluminum structures. 2007.

[286] Kallee SW, Davenport J. Trends in the design and fabrication of rolling stock. European Railway Review. 2007;13.

[287] Sprovieri J. Friction stir spot welding. Assembly magazine. 2016.

[288] Fehrenbacher A, Duffie NA, Ferrier NJ, Pfefferkorn FE, Zinn MR. Toward automation of friction stir welding through temperature measurement and closed-loop control. Journal of Manufacturing Science and Engineering, Transactions of the ASME. 2011;133.

[289] Hilgert J, Schmidt HNB, dos Santos JF, Huber N. Thermal models for bobbin tool friction stir welding. Journal of Materials Processing Technology. 2011;211:197-204.

[290] Sued MK, Pons D, Lavroff J, Wong EH. Design features for bobbin friction stir welding tools: Development of a conceptual model linking the underlying physics to the production process. Materials and Design. 2014:632-43.

[291] De Vuyst T, Vignjevic R. Total Lagrangian SPH modelling of necking and fracture in electromagnetically driven rings. International Journal of Fracture. 2013;180:53-70.

[292] Das R, Cleary PW. Effect of rock shapes on brittle fracture using Smoothed Particle Hydrodynamics. Theoretical and Applied Fracture Mechanics. 2010;53:47-60.

[293] Ahmadnia M, Shahraki S, Kamarposhti MA. Experimental studies on optimized mechanical properties while dissimilar joining AA6061 and AA5010 in a friction stir welding process. International Journal of Advanced Manufacturing Technology. 2016:1-16.

[294] Deepandurai K, Parameshwaran R. Multiresponse Optimization of FSW Parameters for Cast AA7075/SiCp Composite. Materials and Manufacturing Processes. 2016;31:1333-41.

[295] Elfar OMR, Rashad RM, Megahed H. Process parameters optimization for friction stir welding of pure aluminium to brass (CuZn30) using taguchi technique. MATEC Web of Conferences2016.

[296] Ghetiya ND, Patel KM, Kavar AJ. Multi-objective Optimization of FSW Process Parameters of Aluminium Alloy Using Taguchi-Based Grey Relational Analysis. Transactions of the Indian Institute of Metals. 2016;69:917-23.

[297] Lakshminarayanan AK. Enhancing the properties of friction stir welded stainless steel joints via multi-criteria optimization. Archives of Civil and Mechanical Engineering. 2016;16:605-17.

[298] Malopheyev S, Vysotskiy I, Kulitskiy V, Mironov S, Kaibyshev R. Optimization of processing-microstructure-properties relationship in friction-stir welded 6061-T6 aluminum alloy. Materials Science and Engineering A. 2016;662:136-43.

[299] Schmale J, Fehrenbacher A, Shrivastava A, Pfefferkorn FE. Calibration of Dynamic Tool-Workpiece Interface Temperature Measurement during Friction Stir Welding. Measurement.88:331-42.

[300] Wang D, Xiao BL, Ni DR, Ma ZY. Friction stir welding of discontinuously reinforced aluminum matrix composites: a review. Acta Metallurgica Sinica. 2014;27:816-24.

[301] Gerlich A, Su P, Yamamoto M, North TH. Material flow and intermixing during dissimilar friction stir welding. Science and Technology of Welding and Joining. 2008;13:254-64.

[302] Grujicic M, Arakere G, Pandurangan B, Ochterbeck JM, Yen CF, Cheeseman BA, et al. Computational analysis of material flow during friction stir welding of AA5059 aluminum alloys. Journal of Materials Engineering and Performance. 2012;21:1824-40.

[303] Wójcicka A, Mroczka K, Kurtyka P, Binkowski M, Wróbel Z. X-ray microtomography analysis of the aluminum alloy composite reinforced by SiC after friction stir processing. Journal of Materials Engineering and Performance. 2014;23:3215-21.

[304] Threadgill PL, Leonard AJ, Shercliff HR, Withers PJ. Friction stir welding of aluminum alloys. International Materials Review. 2009;54.

[305] Hamming RW. Numerical Methods for Scientists and Engineers: McGraw Hill; 1962.

[306] Walbridge S, Chevrotiere A. Opportunities for the use of aluminum in vehicular bridge construction. Montreal Qc.: MAADI Group; 2012.

[307] Peirce J. Canada's decaying infrastructure - the lessons of Laval. http://www.pipsc.ca/portal/page/portal/website/news/magazine/winter07/Canada%E2%80%99 s%20Decaying%20Infrastructure?_template=/website/t_printerfriendly: The Professional Institute of the Public Service of Canada; 2007.

[308] Fleming S. US infrastructure decay forecast to cost trillions. http://www.ft.com/cms/s/0/6aa759f8-16c0-11e6-b197-a4af20d5575e.html#axzz4BBlA5jKU: US Politics and Policy; 2016.

[309] Kreyszig E, Herbert K, Norminton EJ. Advanced engineering mathematics. 8th ed. New York: J. Wiley; 1999.

[310] Violeau D. Fluid Mechanics and the SPH Method: Theory and Applications. Oxford, UK: Oxford University Press; 2012.

[311] Young WC, Budynas RG. Roark's formulas for stress and strain - Seventh Edition. New York, NY: McGraw-Hill; 2002.

[312] Liu GR. Meshfree Methods - Moving Beyond the Finite Element Method 2nd Edition. Boca Raton, FL.: CRC Press; 2009.

[313] Smith IM, Griffiths DV, Margetts L. Programming the Finite Element Method - Fifth Edition. West Sussex, UK2014.

[314] Saad Y. Iterative methods for sparse linear systems: Society for Industrial and Applied Mechanics; 2003.

9 APPENDICES

USEFUL FORMULAE

The following section very briefly outlines some important formulae that are often required when deriving the SPH equations presented in this document. More rigorous treatment is available from Kreyszig [309] or Violeau [310].

Nabla or Del Operator

The nabla or del operator is very important in vector calculus; it is written in Cartesian coordinates *(x,y,z)* as:

$$\nabla = \frac{\partial}{\partial x}i + \frac{\partial}{\partial y}j + \frac{\partial}{\partial z}k \tag{9-1}$$

Gradient Operator

The gradient provides a convenient means to determine the direction of greatest increase of a function. The gradient of a scalar function, $f(x, y, z)$, is a vector; it can be written as:

$$grad f = \nabla f = \frac{\partial f}{\partial x}i + \frac{\partial f}{\partial y}j + \frac{\partial f}{\partial z}k \tag{9-2}$$

Divergence Operator

The divergence is an important operator in continuum mechanics; it can be thought of as quantifying the compressibility of a vector field. If the divergence is zero, then the field is incompressible. The divergence of a vector, $\bar{v}$, can be written as:

$$div\bar{v} = \nabla \cdot \bar{v} = \left(\frac{\partial}{\partial x}i + \frac{\partial}{\partial y}j + \frac{\partial}{\partial z}k\right) \cdot \left(v_x i + v_y j + v_z k\right) \tag{9-3}$$
$$= \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}\right)$$

If the divergence is positive, we say that a source exists; if it is negative, then a sink exists. The divergence of a tensor of order n will in general result in a tensor of order $n - 1$.

Material Derivative

The material derivative is used to describe the rate of change of a scalar or vector quantity that is dependent on time and spatial location. This is an important notion to switch from a Lagrangian formulation to an Eulerian one (or vice-verse). The material derivative of a scalar quantity, f, is:

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + v \cdot \nabla f = \frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i}$$

(9-4)

The material derivative of a vector, $\bar{v}$, is:

$$\frac{D\bar{v}}{Dt} = \frac{\partial \bar{v}}{\partial t} + \bar{v} \cdot \nabla \bar{v} = \frac{\partial v_i}{\partial t} + v_i \frac{\partial v_i}{\partial x_j}$$

(9-5)

Liebniz's Theorem

Given a vector function, $\bar{F}$, S is a moving surface bound by ∂S, $\bar{v}$ is the velocity of the moving surface, dA is a vector element of the surface S, and ds is a vector element of the curve ∂S:

$$\frac{d}{dt} \iint_V \bar{F} dA = \int_V \frac{\partial \bar{F}}{\partial t} dA + \oint_{\partial S} \bar{F} \bar{v} \cdot n ds$$

(9-6)

Gauss's Divergence Theorem

Gauss's theorem is a very common operation used in the development of the conservation equations. It is used to transform a volume, V, integral into a surface integral, S. $\bar{F}$ is a vector function, and $\bar{n}$ is a normal vector to the surface:

(9-7)

$$\iiint_V div\bar{F} dV = \iint_S \bar{F} \cdot \bar{n} dA$$

This can be written in components as:

(9-8)

$$\iiint_V \left(\frac{\partial F^1}{\partial x} + \frac{\partial F^2}{\partial y} + \frac{\partial F^3}{\partial z} \right) dxdydz = \iint_S (F^1 dydz + F^2 dxdz + F^3 dxdy)$$

Dyadic product (Tensor Product)

The tensor product is an important operator in vector calculus. As the name suggests, the tensor product of two vectors is a tensor; for example, the tensor product of del and velocity is:

$$\nabla \otimes \bar{v} = \frac{\partial v_j}{\partial x_i}, \quad (\nabla \otimes \bar{v})^T = \bar{v} \otimes \nabla = \frac{\partial v_i}{\partial x_j} \tag{9-9}$$

VERIFICATION CASE - HEAT TRANSFER IN A SOLID BLOCK

In order to verify the heat conduction with convection boundary conditions, a simulation model is built in LS-DYNA® with Finite elements and in SPHriction-3D with SPH elements. The block is initially at 20°C, and the left end is heated with a convection boundary condition. The same parameters are used in both models. The coefficient of convection is set to 100.0 W/m²K and the external temperature is set at 500°C. The specific heat capacity and thermal conductivity of the solid are set to 1.0 J/kgK and 3000.0 W/mK. The simulation runs for 10.0 seconds. A graphical comparison of the results is provided in Figure 9-1. We can see that the temperature contours and the relative magnitude of the temperature is essentially the same in LS-DYNA® and SPHriction-3D.

Let's take a closer look at how the temperature varies at three discrete points in the solid throughout the simulation. We will investigate at $x = 0$, $x = 1/2L$ and $x = L$. The results are shown in Figure 9-2, we can see that the values obtained from SPHriction-3D agree excellently with the values obtained from LS-DYNA®.
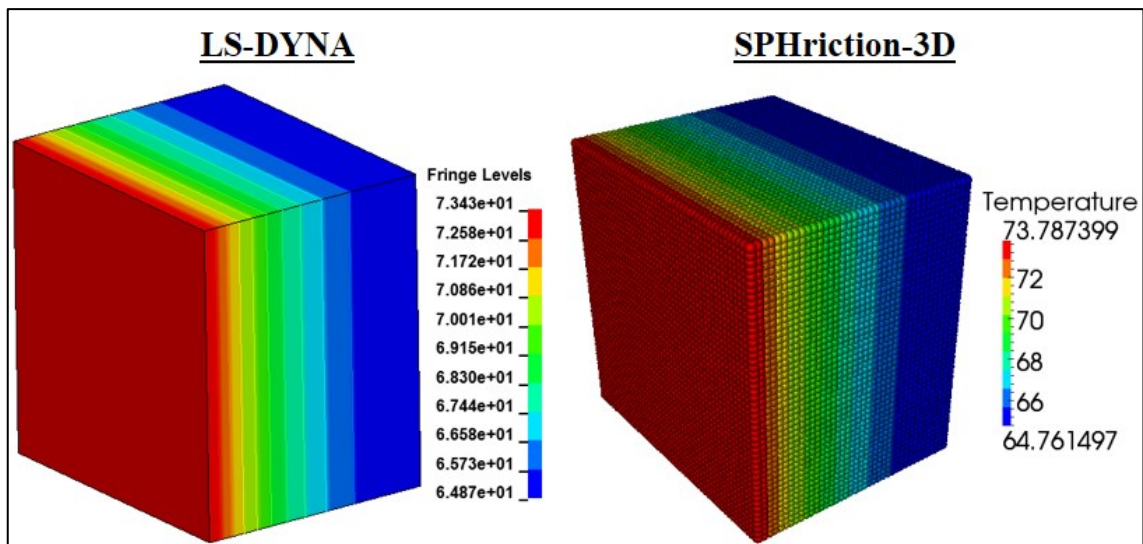

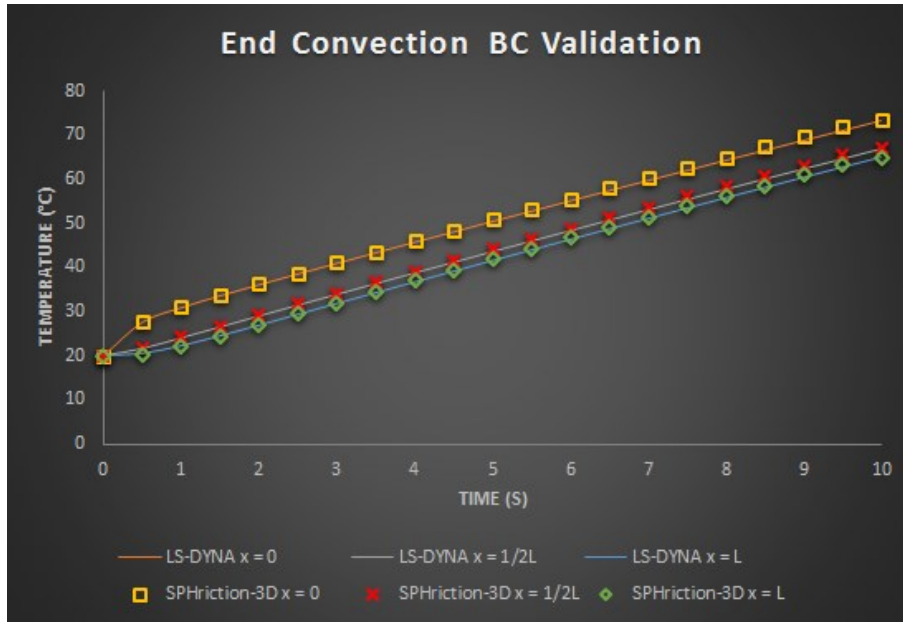
Figure 9-1 – Convection BC on one surface of block

Figure 9-2 – Results comparison for end convection validation case

VERIFICATION CASE - BEAM WITH FIXED ENDS

A model of a beam with fixed ends will be used to test the convergence behaviour of the SPH formulation used in this work. The beam (shown in Figure 9-3) has a square cross section with depth of $D = 1.0$ m, a length of $L = 10.0$, and a uniform load of $W = 50$ kN/m.
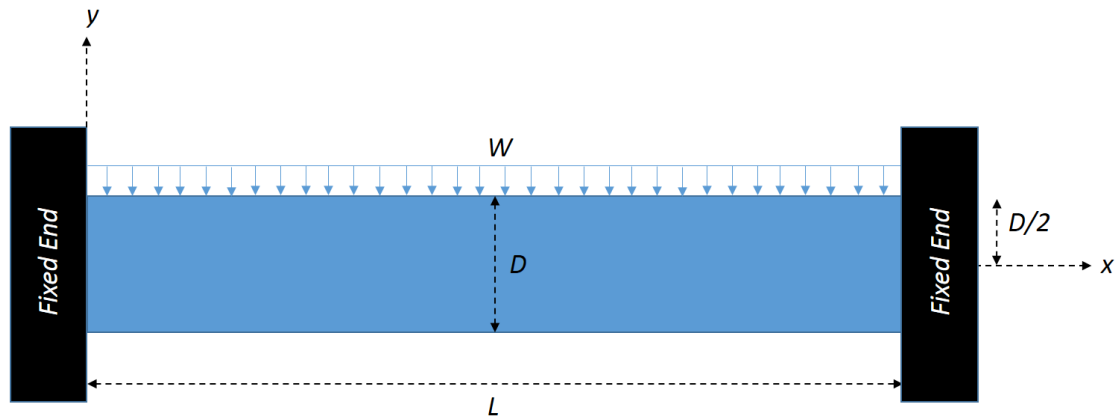

Figure 9-3 – Beam with fixed ends

The analytical solution for the vertical deflection is [311]:

$$u_y(x, y) = -\frac{Wx^2}{24EI}(L^2 - 2Lx + x^2)$$ (9-10)

the horizontal deflection:

$$u_x(x, y) = -\frac{Wxy}{12EI}(L^2 - 3Lx + 2x^2)$$ (9-11)

and the maximum bending stress:

$$\sigma_{xx}(x, y) = \frac{Wy}{12I}(L^2 - 6Lx + x^2)$$ (9-12)

where E and I are the modulus of elasticity and the inertia of the beam respectively. Using the theory calculations, the maximum deflection is mid span and is $u_y(L/2, 0) = 7.5$ mm. The maximum bending stress is $\sigma_{xx}(0 \text{ or } L, \pm D/2) = \pm 25.0$ MPa. The convergence behaviour will be shown using the displacement error norm given by:

$$d_{error} = \sqrt{\frac{\sum_{i=1}^{N_p}\left(u_i|_{theory} - u_i|_{sim}\right)^2}{\sum_{i=1}^{N_p}\left(u_i|_{theory}\right)^2}}$$

(9-13)

where N_p is the number of calculations points in the model, $u_i|_{sim}$ is the displacement at the ith SPH calculation point, and $u_i|_{theory}$ is the displacement calculated from theory sampled at the same locations in the simulation model. Three different particle spacing's are used in this study:

- 6 particles through the thickness, $\Delta s_p = 0.167$ m

- 9 particles through the thickness, $\Delta s_p = 0.111$ m

- 11 particles through the thickness, $\Delta s_p = 0.091$ m

The respective models are shown in Figure 9-4. The blue particles represent the boundary nodes (in each case three layers of boundary nodes are used), the red and green particles discretize the beam and the uniformly distributed load is applied to the red particles. The SPH results will be compared to the analytical results. The same boundary conditions, loading, and geometry is used in all cases. The SPH model uses the standard explicit time integration scheme with the stability criterion:

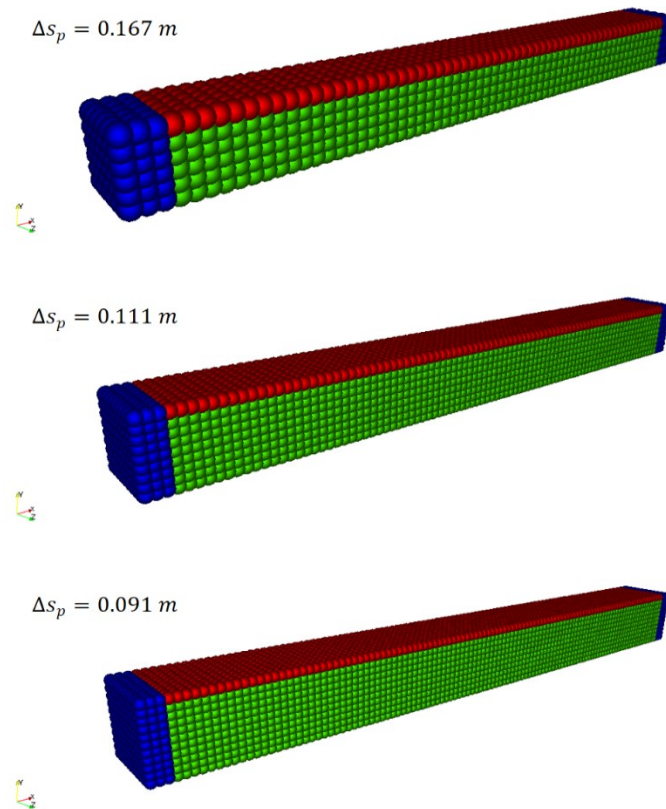$$\Delta t_{mech} = CFL \frac{h}{c + |\bar{v}|}$$

(9-14)

$\Delta s_p = 0.167 \; m$

$\Delta s_p = 0.111 \; m$

$\Delta s_p = 0.091 \; m$

Figure 9-4 – Beam models

For this test case $\mathrm{CFL} = 1.0$. The load is ramped up from zero to full force over 0.5 seconds; then, the load is held steady from 0.5 to 1.0 seconds (end time for the explicit simulation). This procedure is required to ensure that the explicit solution reaches a steady state by the end of the calculation. The uniformly distributed load is converted into a nodal point load based on:

$$F_{NP} = \frac{W \, L}{N_{LP}} \tag{9-15}$$

where N_{LP} is the number of nodes used to discretize the point load. We have found that at least 6 calculation points through the thickness of the part are required to provide a minimum level of precision in the numerical solution. Certainly, more particles through the thickness will improve the resolution and precision of the solution, however, the solution time is expectantly higher.

A comparison of the bending stress results from theory and the explicit SPH are shown in Figure 9-5. A close inspection of the SPH stress contours show that the solution is in good agreement with the analytical solution.
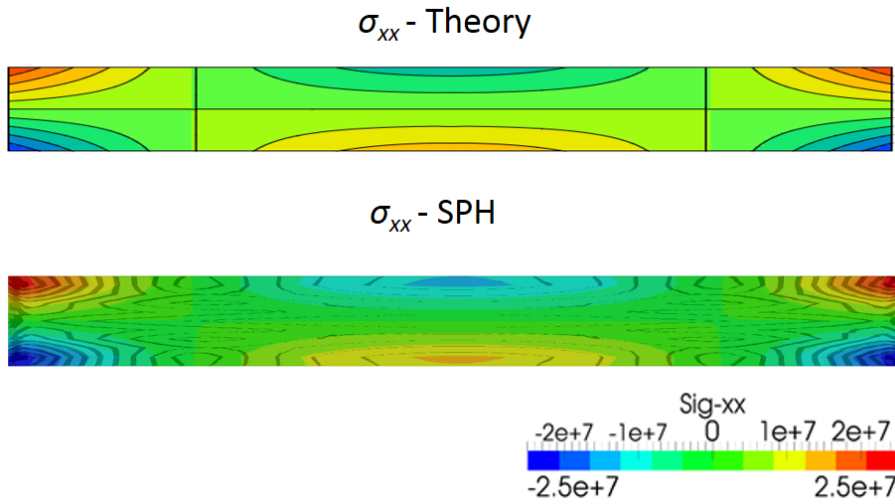

Figure 9-5 – Bending stress (σ_{xx}) Comparison between theory and SPH

The displacement error-norm results are shown in Figure 9-6. The error as a function of particle spacing is provided on the left. For this fixed-end beam problem, standard (explicit) SPH has a convergence rate of 0.93. On the right side of Figure 9-6 is the error as a function of the smoothing length scale factor; $h_{scale} = 1.1$ appears to be best suited for the SPH approach for this case.
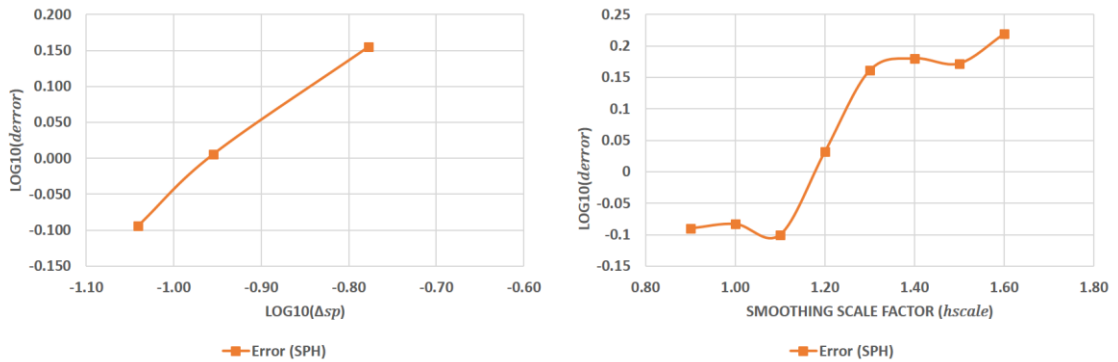

Figure 9-6 – Displacement error norm comparison

Verification Case - Elastic Vibration of an Aluminum Cantilever Beam

Elastic deformation is an important part of the simulation code that we have developed. We prefer to model the FSW process using a solid-mechanics approach, and for this reason, we must be certain that our implementation of elastic stress and strain is correct. Moreover, we should also ensure that the code conserves momentum exactly without the presence of dissipation. A simple elastic vibration model of an aluminum cantilever beam has been set up in SPHriction-3D using SPH and in LS-DYNA® using FEM. The left most end of the beam is fixed and the rightmost end is free. An initial deformation of 10% of the beam thickness is imposed on the tip of the beam. The simulation is run for a total of 0.02 s to capture the transient deformation of the elastic body. The beam is made of AA6061-T6, with an elastic modulus of 70.0 GPa, density of 2700 kg/m^3, and a Poisson ratio of 0.33. The beam has a rectangular cross section, 24 mm thick and 12 mm wide. The total length of the beam is 100 mm. The particle and finite element grid is 2 mm. The deflection is applied by pressing a rigid plate down onto the end of the beam. The plate presses down 2.4 mm over the course of 0.005 s, from t = 0.005 to t = 0.01 s the plate is held steady. At t = 0.01 s, contact between the plate and the beam is disallowed and the beam begins to vibrate.

Figure 9-7 shows a comparison of the effective stress at peak deflection. This point in the simulation corresponds to the moment just before the contact is disallowed (t = 0.01 s). We can see that the contours are almost identical for the two models reaching a maximum value close to the fixed end of 500 MPa.
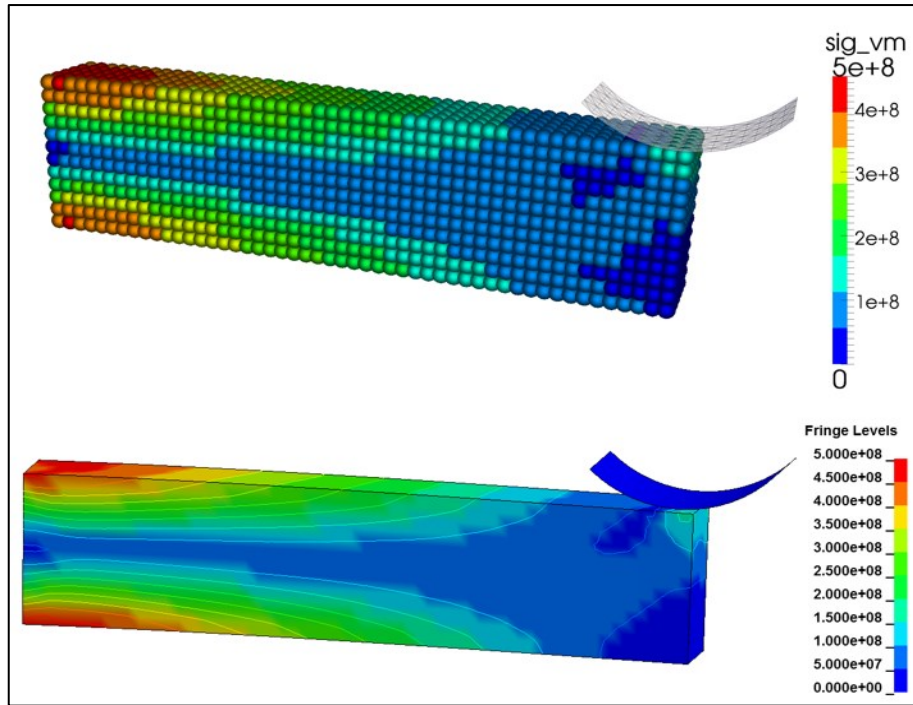
Figure 9-7 – Effective stress from SPHriction-3D (top) and LS-DYNA® (bottom) comparison for the vibrating beam
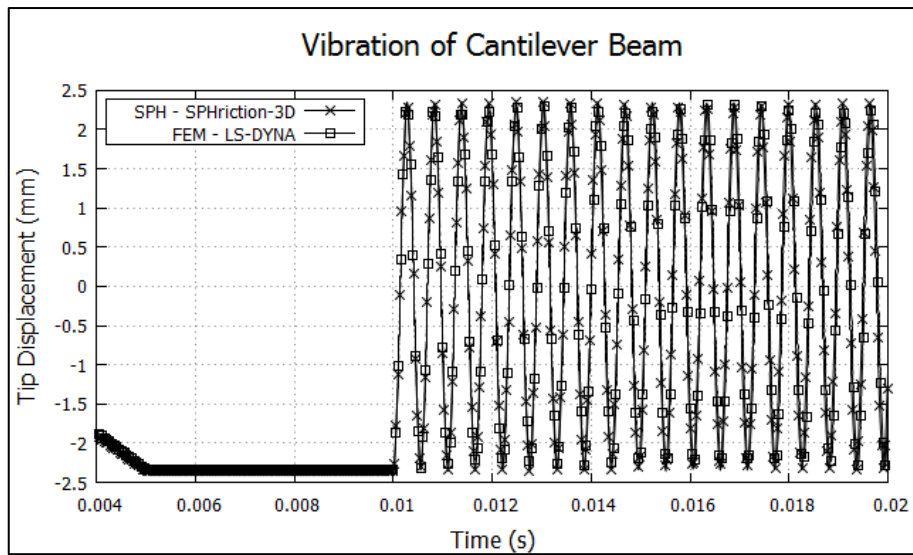

Figure 9-8 – Dynamics response of the vibrating beam for SPHriction-3D and LS-DYNA®

Once the beam starts to vibrate, we expect to see that the period and magnitude remain the same. If the magnitude were to decrease, then the code would not conserve momentum exactly. Figure 9-8 shows the dynamic response of the SPH and FEM beams. The natural frequency of the SPH beam is found to

be 1920 Hz, and that of the FEM beam is 1950 Hz (found by taking an FFT of the time domain signal). Most importantly, we can see that there is no numerical damping present, as the magnitude does not decrease. We have captured ~20 cycles from t = 0.01 to t = 0.02 s. When the simulation is run for a longer period, the undamped simple harmonic motion continues indefinitely.

VERIFICATION CASE - ELASTIC-PLASTIC TENSILE TEST

A tensile test of a cylindrical specimen will be used to evaluate the elastic-plastic algorithms in the SPH code. A baseline case is setup and run in LS-DYNA® using finite elements, the geometry and properties of the steel cylinder are shown in Figure 9-9.
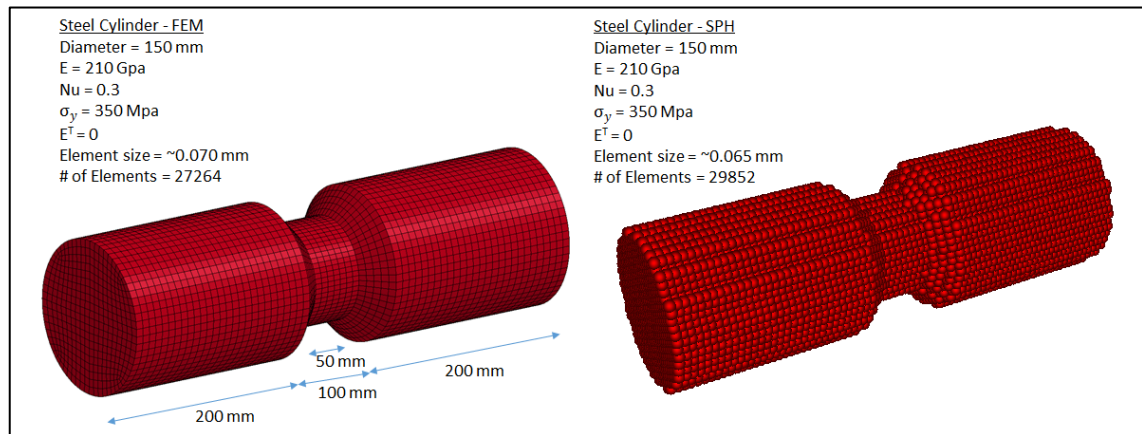

Steel Cylinder - FEM
Diameter = 150 mm
E = 210 Gpa
Nu = 0.3
σ_y = 350 Mpa
E^T = 0
Element size = ~0.070 mm
# of Elements = 27264

Steel Cylinder - SPH
Diameter = 150 mm
E = 210 Gpa
Nu = 0.3
σ_y = 350 Mpa
E^T = 0
Element size = ~0.065 mm
# of Elements = 29852

50 mm
200 mm
100 mm
200 mm

Figure 9-9 – FEM and SPH Steel Cylinder Dimensions and Properties

One end of the cylinder is held fixed, while the other end of the cylinder is given a prescribed velocity that increases from zero to 1.0 m/s at t = 0.005s; the velocity is held constant at 1 m/s from 0.005s to 0.01s (simulation end time). This test is designed to show that SPHriction-3D has a robust and precise implementation of the radial return algorithm for a solid body that can be approximated with the perfect plastic assumption (tangent modulus, E^T, equal to zero).
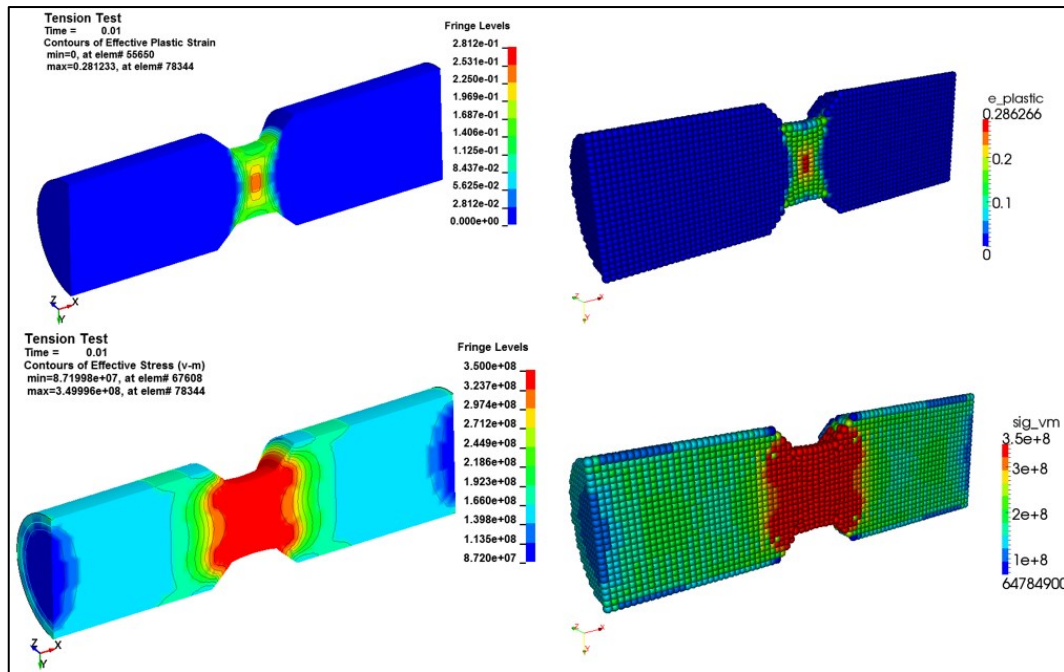
Figure 9-10 – Effective stress (Pa) and plastic strain comparison for LS-DYNA® (left) and

SPHriction-3D (right)
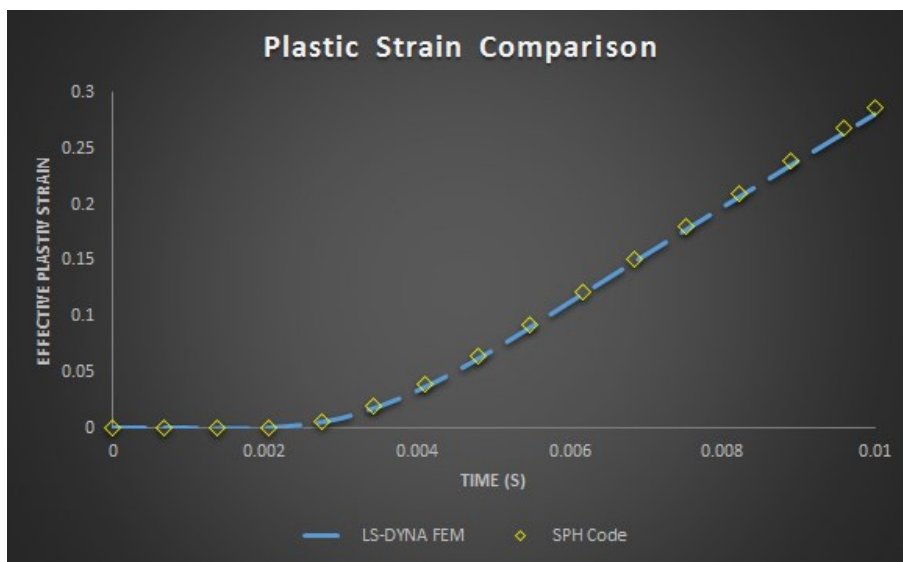

Figure 9-11 – Comparison of effective plastic strain at center of specimen

Figure 9-10 shows a comparison of the effective stress and plastic strain in the FEM and SPH model. We can see that the cylinder yields at the reduced area in both models. The maximum effective plastic strain is 0.281 and 0.286 in the FEM and SPH model respectively. The shape and location of the contours

for the stress and strain in the SPHrictoin-3D code line up with those of the LS-DYNA model with little discrepancy.

Figure 9-11 gives a comparison of the effective plastic strain for an element located at the center of the steel bar (location of max plastic strain). The graph shows that the result obtained with SPHriction-3D is almost identical to that of LS-DYNA®. This then verifies that the radial return algorithm has been appropriately implemented, as there is a very good correlation between the two models.

VERIFICATION CASE - ELASTIC-PLASTIC COMPRESSION TEST WITH HEAT GENERATION

The compression test is setup as shown in Figure 9-12. The loading is provided as a function of time, t, by a velocity, $v_{compress}$, boundary condition:

$$v_{compress} = \begin{cases} \dfrac{v_{max}}{\tau_1} t & 0 \leq t \leq \tau_1 \\ v_{max} & t > \tau_1 \end{cases} \tag{9-16}$$

This type of ramp loading function is recommended for SPH models since it avoids numerical instabilities associated with large accelerations. We typically apply the loading function over a sufficiently long time to ensure dynamic effects are minimal (we are focused on quasi-static response). For the compression test, $v_{max} = 10.0$ and $\tau_1 = 0.005$. The base of the cylinder is fixed in all three translation degrees of freedom. The cylinder is 560 mm long, a diameter of 300 mm and is made of AA6061-T6. A total simulation time of 0.01 s is used with a time step size of 2.5E-06 s ($CFL = 1.0$) for a total of 4000 time steps. The SPH grid is set up with an inter particle spacing of 10mm. The largest element size in the FEM model is 10mm.
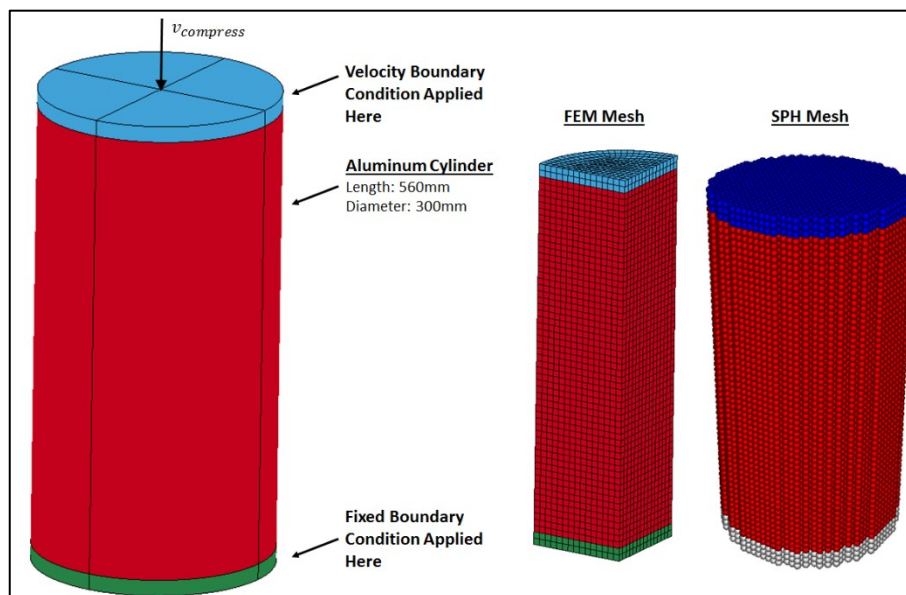

Figure 9-12 – Compression Test Model

The FEM model is composed of a total of 12960 finite elements. The SPH model uses 42956 particles. Once the aluminum cylinder starts to deform plastically, heat is generated, which increases the temperature of the aluminum. The increased temperature causes thermal softening because of our choice of material model. A comparison of the results of the simulations is shown in Figure 9-14. The maximum temperature in the aluminum cylinder is found to be at the geometric center of the specimen. The maximum temperature for the FEM model is 46.5 °C. The standard and adaptive search methods give maximum temperatures of 45.7 °C, only a 1.7% error compared to the FEM model. The total Lagrangian simulation finds a maximum temperature of 50.2 °C for an 8% error. Since we do not actually calculate the Piola-Kirchhoff stress, error is incurred. Initially the temperature for all four methods is within close agreement. Once significant plastic deformation occurs, the total Lagrangian approximation that we use starts to diverge from the FEM results. This behaviour can be seen in detail in Figure 9-13.
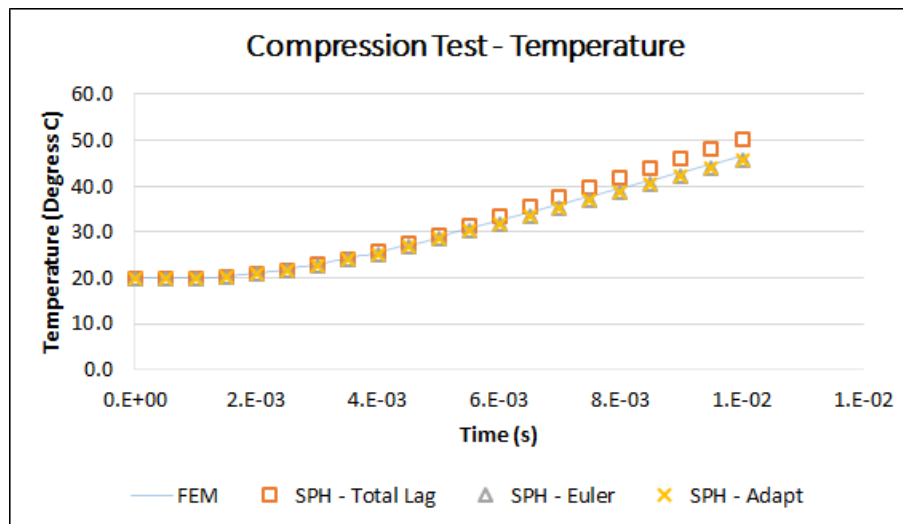

Figure 9-13 – Temperature comparison for the compression test (maximum temperature at center of aluminum specimen)
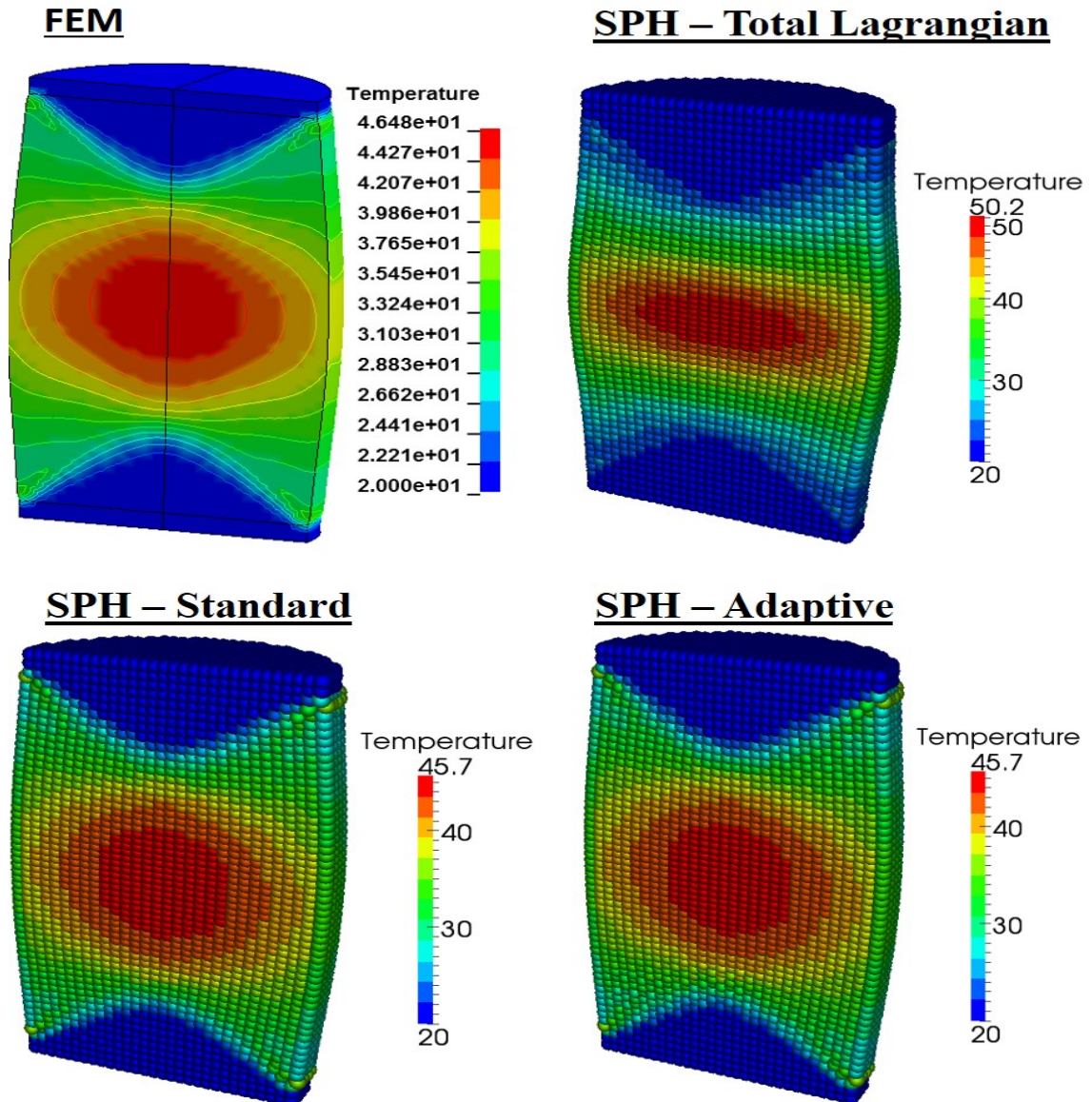
Figure 9-14 – Temperature results for the compression test. Top left: FEM results, top right: SPH with total Lagrangian approach, bottom left: standard SPH approach and bottom right: SPH with adaptive search method

The effective stress (commonly referred to as von Mises stress) is compared for the different models in Figure 9-15. We can see excellent agreement of all three SPH approaches with the FEM results. Note that the effective stress decreases slightly towards the end of the simulation because of thermal softening.

Figure 9-15 – Effective stress comparison for the compression test (taken from center of specimen)

Figure 9-16 – SPH elements that are processed by the adaptive search (in red)
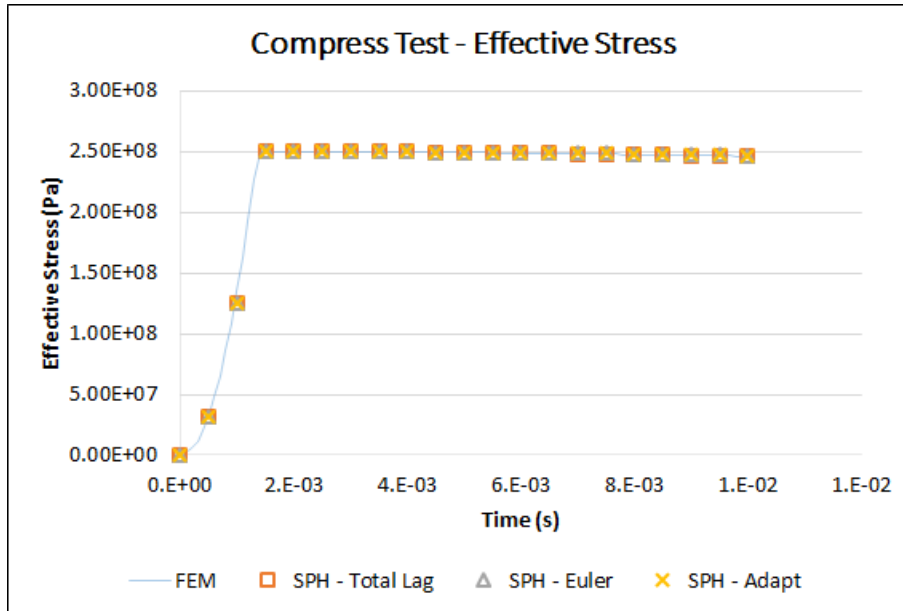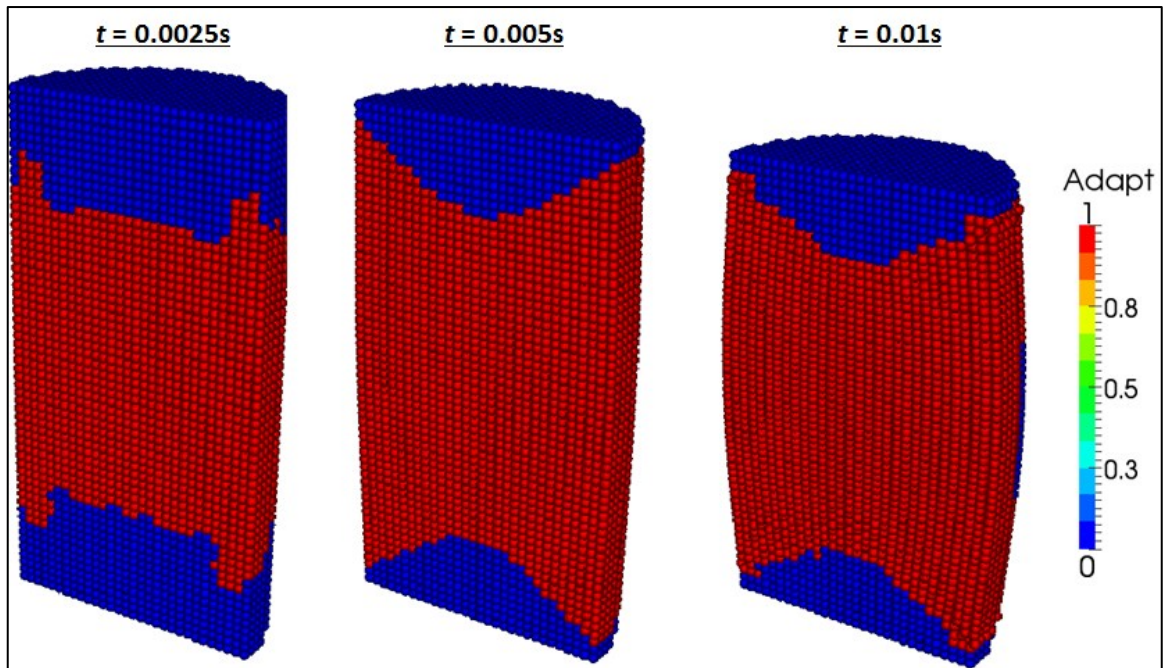
The efficiency of the adaptive search comes from performing the neighbor search only where it really needs to be executed. As the simulation progresses, the SPH elements to be adaptively searched are updated as shown in Figure 9-16.

GPU Performance

The aluminum cylinder is scaled in size in order to compare the performance of the SPH models on the GPU to the CPU. The three models are shown in Figure 9-17. The 42960 element model is the baseline model. The 143730 element model is created by scaling the baseline model 1.5x and the 339360 element model is a 2.0x scale of the baseline model. All three models use the same 10mm inter-particle spacing. For each of the models, the following versions are executed:

- SPH simulation on the CPU using the standard neighbor search every 5 time steps

- SPH simulation on the GPU using the standard neighbor search every 5 time steps

- SPH simulation on the GPU using the total Lagrangian approach

- SPH simulation on the GPU using the adaptive neighbor search

The total time to complete the simulations was recorded and is shown in Figure 9-18. A log based scale is used for the y-axis since the CPU times are considerably higher than those on the GPU. We find that the SPH simulation with the standard searching algorithm every 5 time steps is 19.1 times faster than the CPU on average. With the adaptive search approach, the average speed-up factor is 34.2x. And with the approximate total Lagrangian approach, an average speed-up of 44.3x is achieved. Figure 9-19 gives an overview of the performance of the SPH code on the GPU.
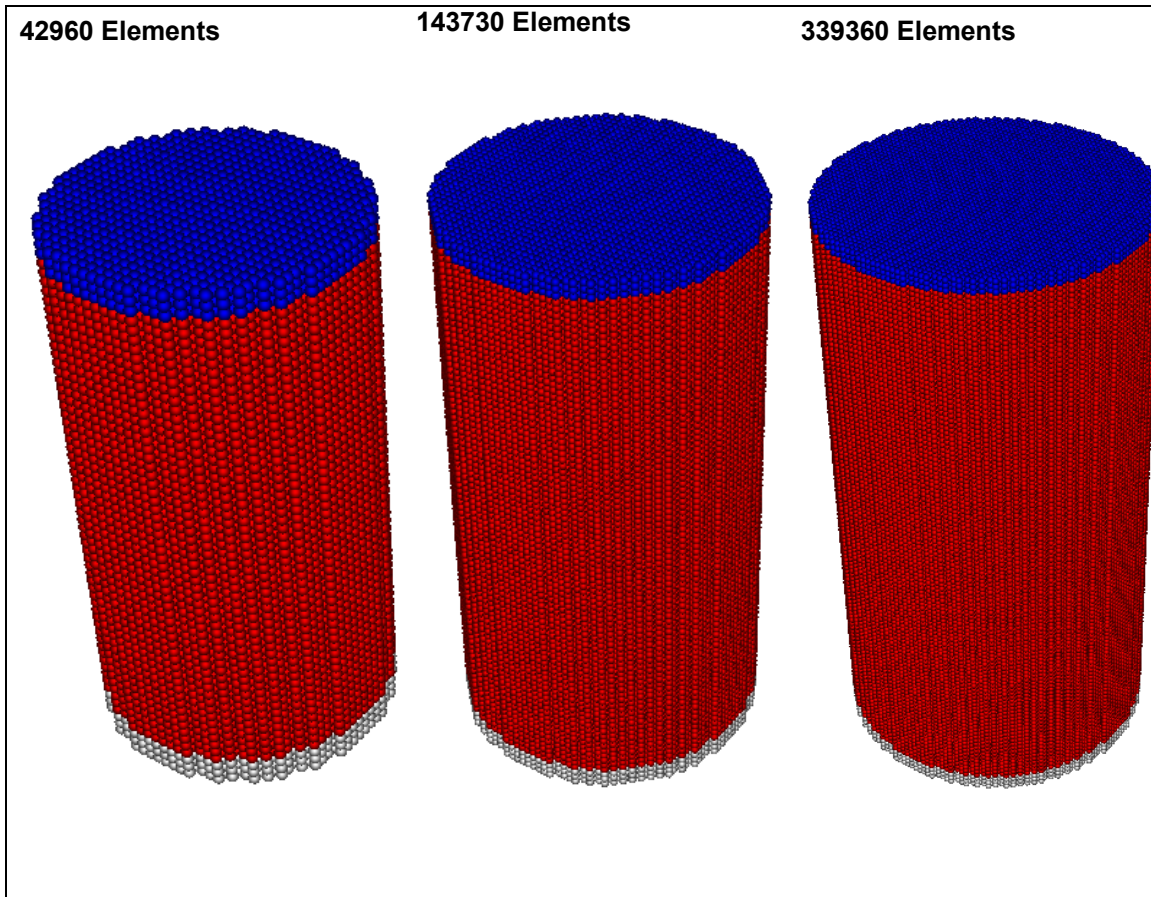
42960 Elements   **143730 Elements**   **339360 Elements**

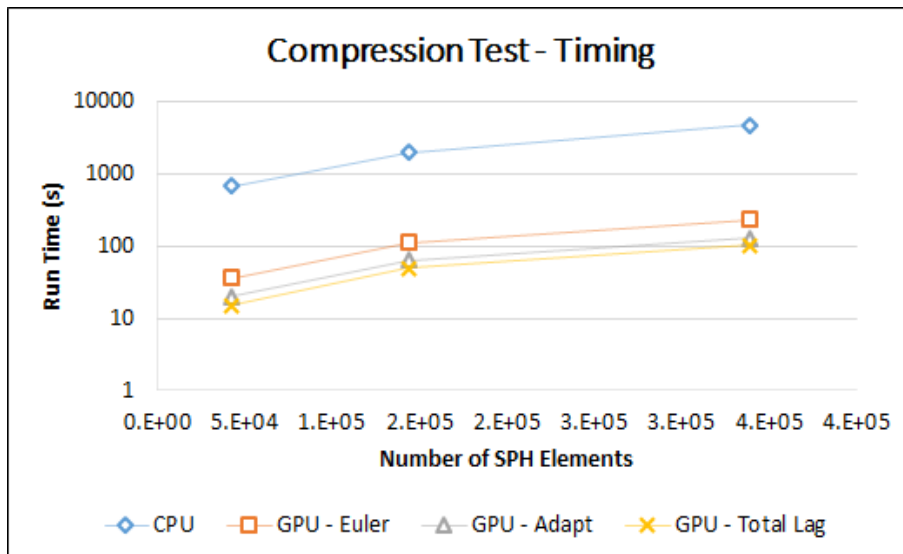**Figure 9-17 – Three models used for performance testing**

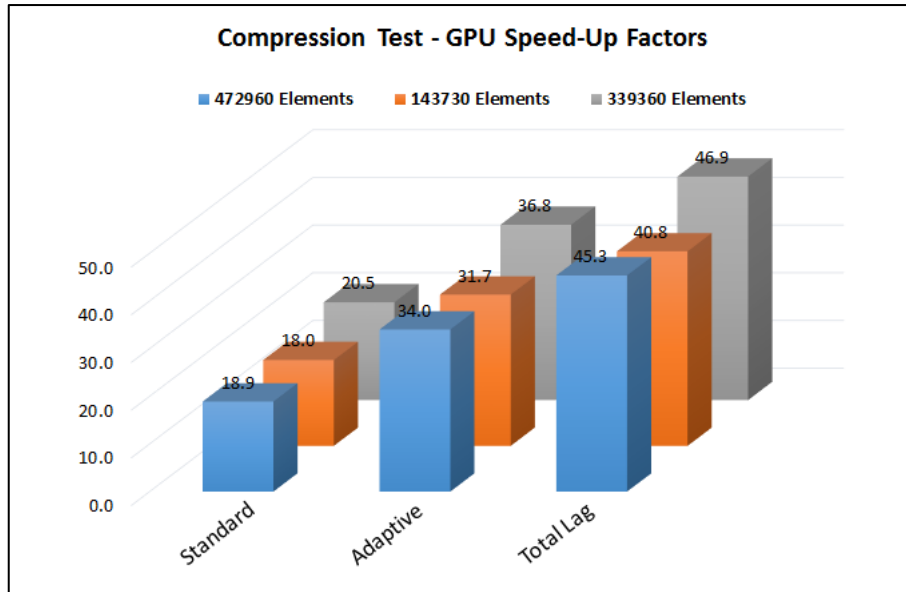**Figure 9-18 – Timing results for the compression test**

Figure 9-19 – Speed-up factors on the GPU

CONVECTION AS A SURFACE INTEGRAL

Previously, the heat source and sink terms (convection, radiation, and friction) that are important for the FSW process were developed by converting the surface integrals into volume integrals. Note that the plastic deformation energy is naturally a volume integral, and should not be cast otherwise.

In some cases, such a volume approximation may be unacceptable. The heat source terms can just as easily be cast into the energy equation by inclusion of the surface integral:

$$\frac{dT_i}{dt} = \frac{1}{\rho_i C_{p_i}} \left[\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \frac{4 k_i k_j}{k_i + k_j} \frac{(T_i - T_j)}{|x_{ij}|^2} x_{ij} \frac{\partial W_{ij}}{\partial x_i{}^\beta} + \chi_{PW} \left(\bar{\bar{\sigma}} : \frac{d\bar{\bar{\varepsilon}}^p}{dt} \right) \right.$$

$$\left. + \sum_{j \in surface} \frac{m_j}{\rho_j} \left(q^\alpha{}_j \hat{n}_{SPH}{}^\alpha{}_j \right) A_j W_{ij_{2D}} \left[\frac{J}{m^3 \, s} \right] \right]$$

(9-17)

Here, $W_{ij_{2D}}$ is an appropriate smoothing function in two dimensions (see section 3.2.2 and Table 3-1). The slight complication with this approach is that a list of the neighbors of the surface particles must be established. The heat flux term, $\bar{q}$ can represent

convection:

$$\bar{q}_{conv} = h_{conv}(T_\infty - T_s)\hat{n}_{SPH}$$

(9-18)

radiation:

$$\bar{q}_{rad} = \epsilon_r \sigma_{SB}\left(T_{surr}{}^4 - T_s{}^4\right)\hat{n}_{SPH}$$

(9-19)

or friction:

$$\bar{q}_{friction} = \epsilon_r \sigma_{SB}\left(T_{surr}{}^4 - T_s{}^4\right)\hat{n}_{SPH}$$

(9-20)

Although this approach is more involved compared to that presented in section 3.9, the results can be improved for problems involving strong flux gradients and/or highly irregular free surfaces.

FULL IMPLICIT SMOOTHED PARTICLE METHOD (FISPM)

The standard SPH method for solid mechanics must be integrated explicitly since there is not a simple way to update implicitly the accelerations with the stresses at the next time step. The main drawback to the explicit time stepping approach is that it is conditionally stable and requires time step sizes that are approximately 50-100 nanoseconds (proportional to the speed of propagation of a stress wave in the material). In this section, a new, true meshfree approach that uses an incremental deformation formulation that is ideal for quasi-static problems is developed. This formulation must be solved implicitly and is not subject to a limit on the time step.

In FSW, the cooling phase will take many minutes compared to a few seconds for the active weld phase. This new formulation will allow us to treat the cooling phase (quasi-static) following the welding phase more efficiently. The ultimate goal of the development is to discretize the set of solid mechanics equations into a meshfree form that can be solved on the GPU. The new approach is called the "full implicit smoothed particle method" (FISPM).

To develop the solid mechanics formulation of FISPM, we will start from the differential form of the dynamic equilibrium equation for a solid body (see Liu [312] or Smith *et al.* [313]):

$$\rho \frac{D\bar{v}}{Dt} = \nabla \cdot \bar{\bar{\sigma}} + \bar{b} \tag{9-21}$$

This equation describes the rate of change of velocity, $\bar{v}$, of a solid body subject to internal forces due to stress, $\bar{\bar{\sigma}}$, and external body forces, $\bar{b}$. In this development, only quasi-static problems will be considered, so, $\partial \bar{v}/\partial t$ will be zero:

$$\nabla \cdot \bar{\bar{\sigma}} + \bar{b} = 0 \tag{9-22}$$

The equation requires a balance between the internal forces (stress divergence) and the external body forces). It is best known as the static equilibrium equation. $\bar{\bar{\sigma}}$ is defined as $\bar{\bar{\sigma}}^T = [\sigma^{11} \quad \sigma^{22} \quad \sigma^{33} \quad \sigma^{23} \quad \sigma^{13} \quad \sigma^{12}]$. The stress in the solid body can be related to the strain, $\bar{\varepsilon}$, using Hooke's law:

$$\bar{\bar{\sigma}} = \boldsymbol{C}\bar{\bar{\varepsilon}}$$

(9-23)

Again, $\bar{\bar{\varepsilon}}$ is defined by the same conventions as $\bar{\bar{\sigma}}$. $\boldsymbol{C}$ is the elasticity tensor; for isotropic materials it is symmetric and takes on the values:

$$C^{\alpha\alpha} = \frac{E(1-v)}{(1-2v)(1+v)} if \ \alpha = \ 1,2,3$$

$$C^{\alpha\alpha} = G, \alpha = \ 4,5,6$$

(9-24)

$$C^{\alpha\beta} = \frac{Ev}{(1-2v)(1+v)} if \ \alpha = \ 2,3 \ , \beta = \ 2,3 \ , \alpha \neq \beta$$

$$C^{\alpha\beta} = 0, otherwise$$

E is the modulus of elasticity, v is poisons ratio, and G is the shear modulus. The material strain is found from the kinematic equation:

$$\bar{\bar{\varepsilon}} = \frac{1}{2}(\nabla\otimes\bar{u} + (\nabla\otimes\bar{u})^T)$$

(9-25)

where $\bar{u}$ is the deformation vector and $\otimes$ is the tensor product (dyadic). For convenience, this can be written as:

$$\bar{\bar{\varepsilon}} = B^T\bar{u} = 0$$

(9-26)

B^T is the transpose of the 6x3 differential operator matrix defined in three dimensions by:

$$B^T = \begin{bmatrix} \partial/\partial x^1 & 0 & 0 & 0 & \partial/\partial x^3 & \partial/\partial x^2 \\ 0 & \partial/\partial x^2 & 0 & \partial/\partial x^3 & 0 & \partial/\partial x^1 \\ 0 & 0 & \partial/\partial x^3 & \partial/\partial x^2 & \partial/\partial x^1 & 0 \end{bmatrix}$$

(9-27)

Using equation (9-23) and (9-26) in (9-22) gives the matrix form of the equilibrium equation:

$$B^T CB\bar{u} + \bar{b} = 0$$

(9-28)

In order to provide a smoothed particle approximation of the static equilibrium equation, a relationship for the mixed derivatives is needed. In this work, the Laplace operator for a scalar field is given by [218]:

$$\nabla \cdot (k \nabla f(x_i)) = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \frac{(4k_i k_j)}{(k_i + k_j)} (f(x_i) - f(x_j)) \frac{x_{ij}{}^\alpha}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\alpha} \tag{9-29}$$

and the Laplacian of a vector field can be approximated as:

$$\nabla \cdot (k \nabla f(x_i{}^\alpha)) = \sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \frac{(4k_i k_j)}{(k_i + k_j)} (f(x_i{}^\alpha) - f(x_j{}^\alpha)) \frac{\partial^2 W_{ij}}{\partial x_i{}^\alpha \partial x_i{}^\beta} \tag{9-30}$$

The mixed derivatives are defined by:

$$\frac{\partial^2 W_{ij}}{\partial x_i{}^\alpha \partial x_i{}^\beta} = \frac{x_{ij}{}^\alpha}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\beta} = \left(\frac{1}{h}\right) \left(\frac{x_i{}^\alpha - x_j{}^\alpha}{r^2}\right) \left(\frac{x_i{}^\beta - x_j{}^\beta}{r}\right) \frac{dW}{dR} \tag{9-31}$$

Using equations (9-30) and (9-31), the static equilibrium equation in the x direction can be written as:

$$u_i{}^1 \sum_{j=1}^{N_i} (\tilde{C}^{11} + \tilde{G}^{23}) - \sum_{j=1}^{N_i} u_j{}^1 (\tilde{C}^{11} + \tilde{G}^{23}) + u_i{}^2 \sum_{j=1}^{N_i} \widetilde{CG}^{12} - \sum_{j=1}^{N_i} u_j{}^2 \widetilde{CG}^{12} + u_i{}^3 \sum_{j=1}^{N_i} \widetilde{CG}^{13}$$

$$- \sum_{j=1}^{N_i} u_j{}^3 \widetilde{CG}^{13} + F^1 = 0 \tag{9-32}$$

Equation (9-31) involves a number of abbreviations:

$$\tilde{C}^{\alpha\alpha} = V_j C_{ij}{}^{11} \frac{x_i{}^\alpha - x_j{}^\alpha}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\alpha} , \alpha = 1,2,3$$

$$\tilde{G}^{\alpha\beta} = V_j G_{ij} \left(\frac{x_i{}^\alpha - x_j{}^\alpha}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\alpha} + \frac{x_i{}^\beta - x_j{}^\beta}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\beta}\right) , \alpha \neq \beta$$

$$\widetilde{CG}^{\alpha\beta} = V_j \left(C_{ij}{}^{12} + G_{ij}\right) \frac{x_i{}^\alpha - x_j{}^\alpha}{r^2} \frac{\partial W_{ij}}{\partial x_i{}^\beta} , \alpha \neq \beta$$

$$V_j = \left(\frac{m_j}{\rho_j}\right)^2 \tag{9-33}$$

$$F_i{}^\alpha = \left(\frac{m_j}{\rho_j}\right) b^\alpha$$

$$C_{ij}{}^{\alpha\beta} = \frac{\left(4 C_i{}^{\alpha\beta} C_j{}^{\alpha\beta}\right)}{\left(C_i{}^{\alpha\beta} + C_j{}^{\alpha\beta}\right)} , \alpha = 1, \beta = 1,2$$

$$G_{ij} = \frac{\left(4G_i G_j\right)}{\left(G_i + G_j\right)}$$

Similar equations can be written in the y and z directions as well. Equation (9-32) can be succinctly written as a matrix equation of the form:

$$\bar{\bar{K}}\bar{u} = \bar{F} \tag{9-34}$$

$\bar{\bar{K}}$ is the global stiffness matrix in units of force per length, $\bar{u}$ is the incremental deformation vector, and $\bar{F}$ is the total incremental force vector that comprises the applied force, thermal expansion force, and the plastic corrector force. The stiffness matrix is 3Nx3N and both $\bar{u}$ and $\bar{F}$ are vectors of length 3N. The set of equations must be solved using an appropriate iterative solution procedure. Although the set of equations are sparse, each row of the stiffness matrix will involve ~56 entries in 3D (using $h_{scale} = 1.2$). Because of this, inverting the stiffness matrix is extremely computationally expensive for even small simulation models. Once the deformations are known, the strain increment can be found from equation (9-26) then the stresses can be found from (9-23).

Because of the kernel truncation at the boundary of the solid, the resulting stiffness' will be incorrect along the free surface. This can be remedied by correcting the kernel gradient used in equation (9-31). The kernel gradient correction was explained in section 3.8. The corrected mixed derivatives are then approximated by:

$$\frac{\partial^2 W_{ij}}{\partial x_i{}^\alpha \partial x_i{}^\beta} = \frac{x_{ij}{}^\alpha}{r^2}\widehat{\nabla_\beta W_{ij}} \tag{9-35}$$

This correction breaks the symmetry of the stiffness matrix, meaning that the linear system must be solved by an iterative solver for non-symmetric sparse positive definite matrices. The bi-conjugate gradient stabilized (BiCGstab) method is a good choice in this case.

A similar development can be performed for the heat transfer equations. In this case, the goal is to update the temperatures in the model implicitly. Starting from the heat diffusion equation and using a backwards difference to discretize the time derivative provides an implicit relationship between the material point temperature at time step m and m+1:

$$\frac{T_i^{m+1} - T_i^m}{\Delta t} = \frac{1}{\rho_i C_{p_i}} \left[\sum_{j=1}^{N_i} \frac{m_j}{\rho_j} \frac{(4k_i k_j)}{(k_i + k_j)} \frac{(T_i^{m+1} - T_j^{m+1})}{|x_{ij}|^2} x_{ij} \frac{\partial W_{ij}}{\partial x_i^\beta} + \dot{q} \right] \tag{9-36}$$

Re-arranging the equation to have the temperature at time step m+1 on the left side and m on the right side:

$$T_i^{m+1} \left(1 - \sum_{j=1}^{N_i} A_{ij} \right) + \sum_{j=1}^{N_i} A_{ij} T_i^{m+1} = T_i^m + \frac{\Delta t \dot{q}}{\rho_i C_{p_i}} \tag{9-37}$$

where

$$A_{ij} = \frac{\Delta t}{\rho_i C_{p_i}} \left[\frac{m_j}{\rho_j} \frac{(4k_i k_j)}{(k_i + k_j)} \frac{x_{ij}}{|x_{ij}|^2} \frac{\partial W_{ij}}{\partial x_i^\beta} \right] \tag{9-38}$$

This forms a matrix equation of the form:

$$\bar{\bar{C}} \bar{T} = \bar{Q} \tag{9-39}$$

The coefficient matrix, $\bar{\bar{C}}$ is symmetric sparse positive definite (SPD), for this reason, the conjugate gradient (CG) method can be used solve for the temperature at time step m+1.

The sparsity of the stiffness and thermal coefficient matrices can easily be preserved by ordering the entries according to the neighbor list for the mechanical and thermal domains. The CG and BiCGstab algorithm used in this work is based on the work of Saad [314].

Verification of FISPM

A model of a beam with fixed ends will be used to test the convergence behaviour of FISPM and CFISPM. The beam (shown in Figure 9-3) has a square cross section with depth of $D = 1.0$ m, a length of $L = 10.0$, and a uniform load of $W = 50$ kN/m.
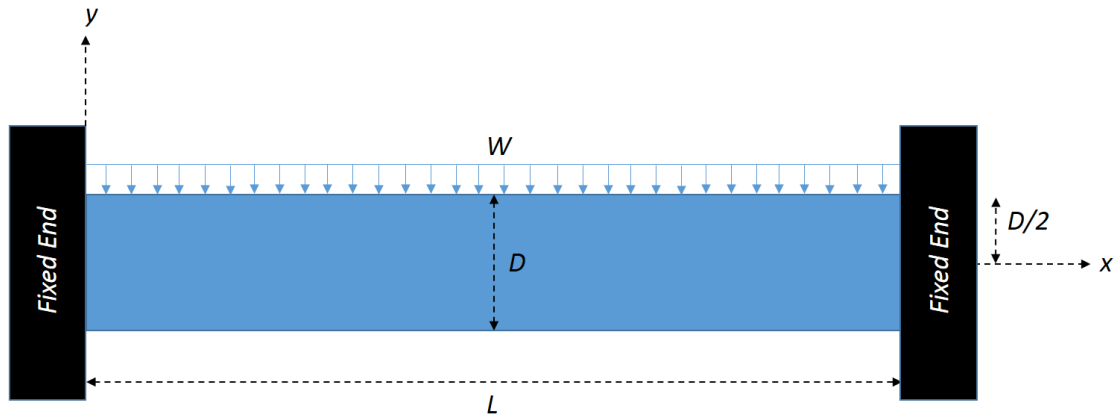
Figure 9-20 – Beam with fixed ends

The analytical solution for the vertical deflection is [311]:

$$u_y(x, y) = -\frac{Wx^2}{24EI}(L^2 - 2Lx + x^2)$$

(9-40)

the horizontal deflection:

$$u_x(x, y) = -\frac{Wxy}{12EI}(L^2 - 3Lx + 2x^2)$$

(9-41)

and the maximum bending stress:

$$S_{xx}(x, y) = \frac{Wy}{12I}(L^2 - 6Lx + x^2)$$

(9-42)

where E and I are the modulus of elasticity and the inertia of the beam respectively. Using the theory calculations, the maximum deflection is mid span and is $u_y(L/2, 0) = 7.5$ mm. The maximum bending stress is $S_{xx}(0 \text{ or } L, \pm D/2) = \pm 25.0$ MPa. The convergence behaviour will be shown using the displacement error norm given by:

$$d_{error} = \sqrt{\frac{\sum_{i=1}^{N_p}\left(u_i|_{theory} - u_i|_{sim}\right)^2}{\sum_{i=1}^{N_p}\left(u_i|_{theory}\right)^2}}$$

(9-43)

where N_p is the number of calculations points in the model. $u_i|_{sim}$ is the displacement at the ith FISPM

calculation point, and $u_i|_{theory}$ is the displacement calculated from theory sampled at the same locations

in the simulation model. Three different particle spacing's are used in this study:

- 6 particles through the thickness, $\Delta s_p = 0.167$ m

- 9 particles through the thickness, $\Delta s_p = 0.111$ m

- 11 particles through the thickness, $\Delta s_p = 0.091$ m

The respective models are shown in Figure 9-4. The blue particles represent the boundary nodes (in

each case three layers of boundary nodes are used), the red and green particles discretize the beam

and the uniformly distributed load is applied to the red particles. The FISPM and CFISPM results will be

compared to the analytical results as well as the standard SPH. The same boundary conditions, loading,

and geometry is used in all cases. The SPH model uses the standard explicit time integration scheme

with the stability criterion:

$$\Delta t_{mech} = CFL \frac{h}{c + |\bar{v}|} \qquad (9\text{-}44)$$

For this test case $CFL = 1.0$. The load is ramped up from zero to full force over 0.5 seconds; then, the

load is held steady from 0.5 to 1.0 seconds (end time for the explicit simulation). This procedure is

required to ensure that the explicit solution reaches a steady state by the end of the calculation.
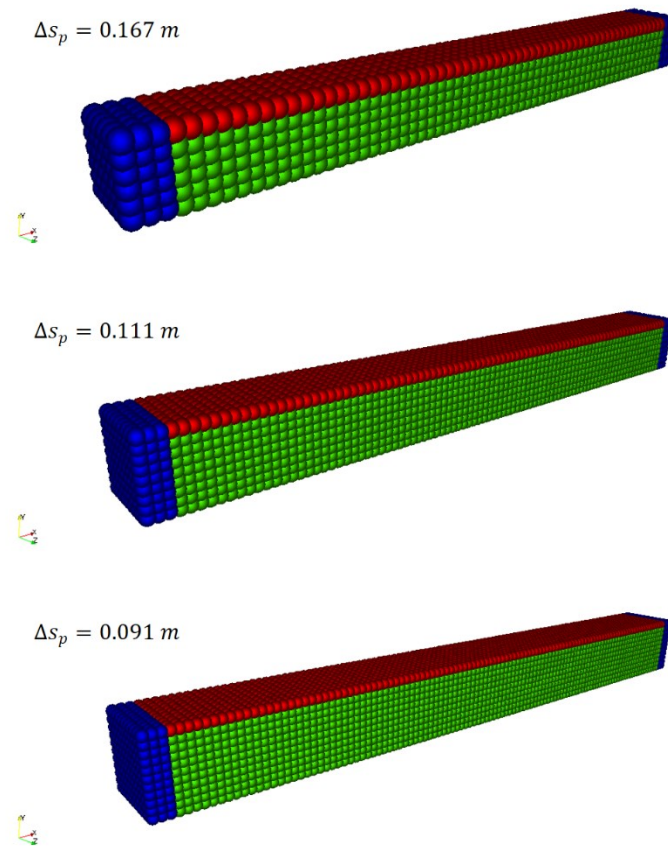
$\Delta s_p = 0.167\ m$

$\Delta s_p = 0.111\ m$

$\Delta s_p = 0.091\ m$

Figure 9-21 – Beam models

The uniformly distributed load is converted into a nodal point load based on:

$$F_{NP} = \frac{W\ L}{N_{LP}} \tag{9-45}$$

where N_{LP} is the number of nodes used to discretize the point load. We have found that at least 6 calculation points through the thickness of the part are required to provide a minimum level of precision in the numerical solution. Certainly, more particles through the thickness will improve the resolution and precision of the solution, however, the memory storage requirements are $\mathcal{O}(N_p{}^3)$, and the solution time is expectantly higher.

A comparison of the bending stress results from theory, FISPM, CFISPM, and standard explicit SPH are shown in Figure 9-5. A close inspection of the FISPM contours show that the solution is affected by the

boundary particles. This can best be noted by looking at the top edge of the beam, here we can see that the stress contours locally decrease close to the boundary. Whereas the CFISPM results are in better agreement with those of the analytical solution. The stress contours in the standard SPH results are not uniform and show spurious stress oscillations in the model.
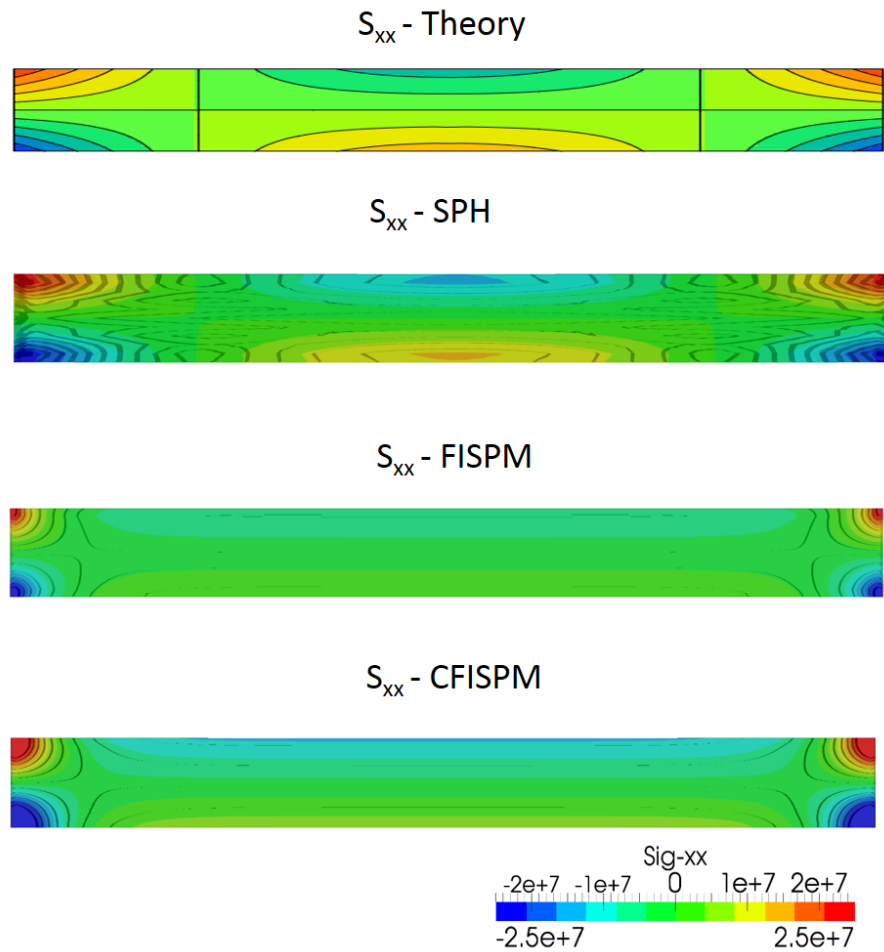

Figure 9-22 – Bending stress (S_{xx}) Comparison between theory, SPH, FISPM, and CFISPM

The displacement error norm results are shown in Figure 9-6 for SPH, FISPM, and CFISPM. The error as a function of particle spacing is provided on the left. The CFISPM shows an improvement of the precision. The convergence rates are 1.4 and 1.5 for FISPM and CFISPM respectively. For this fixed-end beam problem, standard (explicit) SPH has a convergence rate of 0.93. On the right side of Figure 9-23 – Displacement error norm comparison is the error as a function of the smoothing length scale factor. In the case of FISPM, the precision reaches a maximum for $h_{scale} = 1.1$, whereas $h_{scale} = 1.2$

appears to be best suited for the CFISPM approach. Notice that the precision of the CFISPM approach is more uniform than FISPM across the different values of h_{scale}. This is because the corrected method shows less sensitivity to the boundary conditions.
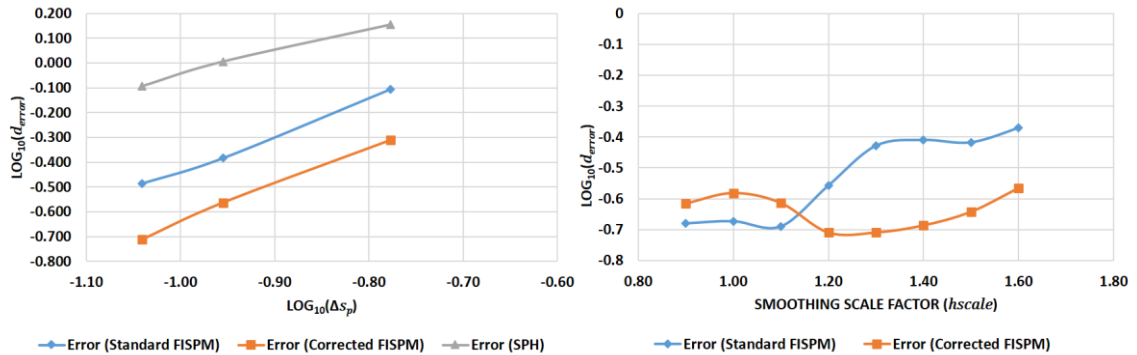

Figure 9-23 – Displacement error norm comparison

The performance of the proposed method is compared to the standard explicit SPH code in Figure 9-24. The graph shows the ratio of calculation times of SPH with FISPM (blue bars), and SPH with CFISPM (orange bars). The best speed up of 37x is achieved by FISPM, whereas the lowest speedup of 23x is from CFISPM. The corrected method requires slightly more calculation time because the BiCGstab solver requires more operations per calculation point.
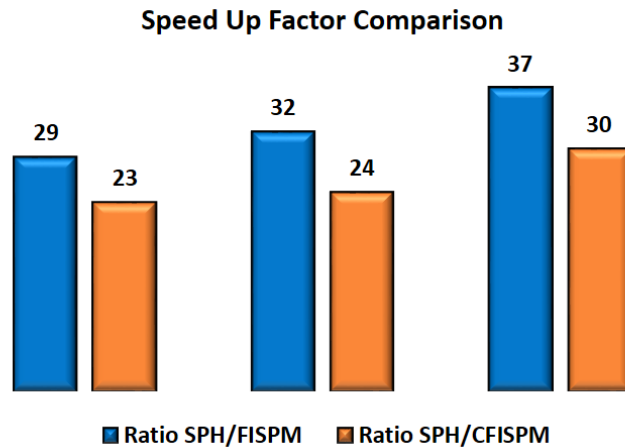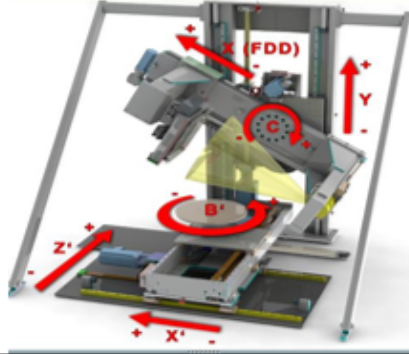

Figure 9-24 – Performance comparison - $h_{scale} = 1.1, 1.2, 1.3$

3D COMPUTED TOMOGRAPHY PARAMETERS - YXLON MULTIPLEX 5500M

Paramètre Général Yxlon Multiplex 5500M

Tube			Détecteur et calibration XRD820 PerkinElmer		
	Voltage (Kv)	220		Calibration d'image	triple gain
	Courrant (mA)	2.4		Nombres d'integration	100
	Puissance (Watt)	528		rafraichissement ip/s	15
	Dimenssion point focal (mm)	0.3		résolution (mm)	0.2
	Filtre	cuivre 1 mm		Dimenssion X (px)	1024
				Dimenssion Y (px)	1024
	Integration (nb) 2D	N/a		Pixel binning	non

Manipulateur		
	X'	23.008
	Y	431
	FDD	964.499
	C (2D seulement)	0.25
	B' (2D seulement)	-97.56
	Z' (2D seulement)	-13.705
	Colimator	non
	BS Collimator_1_Gate_1_	575
	BS Collimator_1_Gate_2_	575
	BS Collimator_1_Gate_3_	575
	BS Collimator_1_mode cmd	1
	BS Collimator_1_speed cmd	100

Paramètres pour acquisition 3D

Paramètres de base			Géométrie 3D		
	type de d'acquisition	qualité		Field of view	39.21
	Intégration (nb)	5		focus detector	1053.14
	Parameter			focus object	573.53
	Projection	720		focus tilt distance	520
	result S(vx)	512		FOV ext center	-642
	rotation angle	360		Magnification	1.836
	Height (mm)	111.53		Object Detection	479.6
	d'acquisition hauteur (Nb)	1		Scan Height	111.53
	Center O (px)	manuel		Vertical center	-26.05
	Ext fiel of view	non			
	Front / left views	non			
	ROI width (px)	360			
	ROI Heigth(px)	1024		Voxel Size	0.07658

Paramètres avancé		
	Beam hardening a0	0
	Beam hardening a1	0
	Beam hardening a2	0
	Beam hardening a3	0
	Protrusion mode	1
	Use GPU	
	Expert	
	Deblurring factor	no

The 3D x-ray images of the welded plates were carried out by Dany Drolet at "Centre de technologie d'aluminum" CNRC-NRC. The parameters used in the study are provided above.