Optimization of Thermo-mechanical Conditions in Friction Stir Welding

by

Cem Celal Tutum

Ph.D. Thesis

Technical University of Denmark
Department of Mechanical Engineering

October, 2009

Optimization of Thermo-mechanical Conditions in Friction Stir Welding

Process Modelling Group
Department of Mechanical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

TM 02-09

ISBN 978-87-89502-89-2

To my mother

Preface

This work has been carried out at the Department of Mechanical Engineering (MEK), Technical University of Denmark (DTU), during the period 2006-2009. The work was supervised by Professor Jesper H. Hattel (MEK), and co-supervised by Associate Professor Henrik N. B. Schmidt (MEK) and Professor Martin P. Bendsøe, the Department of Mathematics (MAT), DTU.

I would like to express my sincere gratitude to Professor Hattel for his unfailing guidance and support throughout my studies, critical review of my work, and most importantly for his great patience and enthusiasm. I would like to thank Dr. Schmidt for providing me irreplaceable inspiration with his state-of-the-art models and being more than a co-supervisor. Special thanks to Prof. Bendsøe for taking intense academic interest in this study as well as providing valuable suggestions that improved the quality of the work.

I am also indebted to Jon Spangenberg and Petr Kotas; this work would not have been finalized without their help and encouragement. I would like to express my thanks to Kim Lau Nielsen and Anders Astrup Larsen together with other INNOJoint Project members for their strong collaboration. Moreover, I would like to express my special gratitude to Dr. Jesper Thorborg and Jens Ole Frandsen for always being ready for any kind of support. Vivek Chidambaram, Jokob Hilgert, Jacob Børby, Martin Asger Haugaard, José Blasques and Ramin Moslemian, thanks for being such companions on our journeys into the wilds of Ph.D. studies.

Sincere thanks are due to the TopOpt (*www.topopt.dtu.dk*) research group members for their fruitful discussions and social activities spent together.

It was a privilege to meet Dr. Ivo Sbalzarini from ETH Zurich and have an opportunity to discuss on the evolutionary optimization and high performance computing. My keen appreciation goes to him for the inspiration he provided. My special thanks are also due to the Danish Center of Applied Mathematics and Mechanics (DCAMM) for offering a wide range of courses that are given by distinguished researchers.

There are other people whom I owe quite a bit of inspiration. Ata Muğan, Erol Şenocak, Haydar Livatyalı and Ümit Sönmez are gratefully acknowledged.

I wish to express my cordial appreciation to my parents as well as Raik and Tülay, for their love, patience and encouragement. My mother Tülin, I devote this thesis to her.

Finally, my thanks go to Danish Government and DTU for their continuous financial support and providing a nice working environment for this Ph.D. work during the three years.

Cem Celal Tutum
Kgs. Lyngby, October 2009

Abstract

The present thesis deals with the challenging multidisciplinary task of combining the manufacturing process of friction stir welding (FSW) with mathematical optimization methods in the search for optimal process parameters. The goals (objectives in optimization parlance) are process related in the sense that they describe or express when the process works in an optimal way or yields final parts that are somehow optimal. These expressions are denoted objective functions and the mathematical optimization algorithm is then searching for a set of the investigated process parameters (in optimization terms denoted design variables) that will either minimize or maximize the objective functions depending on the problem at hand.

The FSW process which has been the subject of the optimization in this study is a relatively new welding process that was invented in 1991 by The Welding Institute (TWI), UK. In short, the process is solid-state, that is, no melting takes place, meaning that a lot of the disadvantages normally associated with traditional fusing welding processes can be avoided. In the FSW process a rotating tool is submerged into the two work pieces and due to frictional and plastic dissipation, the temperature is increased to an extend where the material is sufficiently softened to be stirred together, thereby forming a weld. The process is characterized by multiphysics involving solid material flow, heat transfer, thermal softening, recrystallization and the formation of residual stresses.

In the present work, several models for the FSW process have been applied. Initially, the thermal models were addressed since they in essence constitute the basis of all other models of FSW, be it microstructural, flow or residual stress models. Both analytical and numerical models were used and combined with the Sequential Quadratic Programming (SQP) gradient-based optimization algorithm in order to find the welding speed and the heat input that would yield a prescribed average temperature close to the solidus temperature under the tool, thereby expressing a condition which is favourable for the process.

Following this, several thermomechanical models for FSW in both ABAQUS and ANSYS were developed. They were used for the analysis of the transient temperature and stress evolutions during welding and subsequent cooling, eventually leading to the residual stress state and reduced mechanical properties due to thermal softening. In one case, the subsequent loading situation of a real FSW structure was also taken into account, thus making way for an integrated analysis of the welding process and the loading situation during service of the welded part. Another case combined the predicted stresses with a subsequent uni-axial loading situation in which a damage evolution analysis was carried out in order to predict the final weld's load carrying capacity when subject to tension perpendicular to the weld line.

The thermomechanical models predicting residual stresses were also combined with an evolutionary optimization algorithm (NSGA-II) in the search for the optimal combination of the process parameters that one essentially controls in practice, namely the welding speed and the rotational speed, which would minimize residual stresses and maximize the welding speed divided by rotational speed (known as advancement per revolution and expressing a desired feature of the process).

Finally, some more theoretical investigations regarding several well-known unconstrained and constrained multi-objective-optimization (MOO) benchmark problems were carried out. This was done in order to investigate some of the difficulties that a multi objective evolutionary algorithm may have to tackle. Specifically, three elitist algorithms, i.e. the MOGA-II, the NSGA-II (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB with two versions including the one with an extra Pareto-optimal set archieve strategy for post-processing purposes), were employed for this purpose. The results (especially for the constrained case) show that the cNSGA-II shows a good performance for having both a converged and a well-spread distribution of the Pareto-optimal set with less computational cost.

Table of Contents

Chapter 1

Introduction

This Ph.D.thesis is the result of the project entitled "Thermo-mechanical Modelling and Optimization of Dynamic Process Conditions in Friction Stir Welding" which has been carried out at the Process Modeling Group, Department of Mechanical Engineering, Technical University of Denmark (DTU) in the period from July 2006 to October 2009.

1.1 Motivation of the work

The objective of this study is to investigate and control the thermomechanical conditions, mainly residual stresses and service-load conditions, in the friction stir welding (FSW) process by means of numerical modeling and optimization techniques, respectively. The FSW process is not yet fully understood and since the 1990s, many theoretical/experimental investigations and EU supported industrial projects, e.g. JOIN-DMC and DEEPWELD, have been carried out in order to get more knowledge about the many physical aspects of the FSW process. FSW is used already in routine, as well as in critical applications, for the joining of structural components mainly made of aluminium and its alloys. However, further studies are still being carried out in order to obtain a more robust "process window".

The field of modelling FSW is broad ranging from models concerning e.g. microstructure evolution, material flow, heat flow, heat generation, residual stresses, mechanical loads (tool forces etc.) and strength of the joint. In order to get more insight into some of the fundamental mechanisms involved in the FSW process, the physical phenomena have been divided into sub-systems, i.e. thermal and mechanical behavior, followed by the integrated modeling approach combining coupled behavior of these two physical aspects with the actual load case.

This Ph.D. study is part of the project named *INNOJoint* (Innovative Joining Processes Applying Integrated Modelling) a major research project at DTU Mechanical Engineering which is an interdisciplinary project that focuses on determining appropriate welding parameters as early as possible in the design process. Five professors and one associate professor from DTU, two postdocs, six PhD students and six multinational companies, e.g. EADS (Airbus), BOSCH, Volvo, Danfoss, Grundfos and Danstir are also involved in the project.

The majority of the work carried out in the present Ph.D. project has been related to

computational models of different aspects of the FSW process. Especially, the thermal profiles and transient/residual stresses evolved during and after welding, respectively, have been investigated numerically and compared with published experimental results. Thus, important information regarding the evolution of residual stresses which is a function of the main FSW process parameters, i.e. tool rotational speed and traverse welding speed, etc., is gathered by applying the general purpose finite element (FE) software ANSYS (Ansys Inc., 2007), ABAQUS (Dassault Systèmes, 2007) and COMSOL (COMSOL AB, 2007). Following the modeling studies, numerical optimization routines, i.e. both classical and evolutionary algorithms, have been used in order to find optimum process parameters aiming at process-specific goals. These goals are formulated as single and multi-objective optimization problems which need to be solved with appropriate numerical techniques which utilize different solution strategies. For this purpose, the commercial programming language MATLAB (The Math Works Inc., 2007) and the multi-objective optimization software modeFRONTIER (ESTECO s.r.l., 2007), have been used.

1.2 The Friction Stir Welding process

Friction Stir Welding (FSW) is an efficient solid-state, i.e. without melting in the workpiece, joining technique that is intended to be used for joining of especially the aluminum alloys, besides dissimilar welds, which are difficult to weld with traditional welding techniques. Joining of large panels, which cannot be easily heat treated post weld to recover temper characteristics, is another common interest for FSW to be preferred for industrial applications besides its other benefits. It was invented and experimentally proven by Wayne Thomas and his colleagues at The Welding Institute, Cambridge, UK in December 1991, and then TWI filed for world-wide patent protection in December of that year (TWI, 1991; Wikipedia, 2009a; Bhadeshia, 2003).

In FSW, a cylindrical-shouldered tool, with a cylindrical/profiled, threaded/unthreaded probe (pin) is rotated at a constant speed and moved at a constant traverse rate in the joint line between two pieces of sheet or plate material, which are butt welded together as shown in Figure 1.1. The parts have to be clamped rigidly onto a backing plate in order to prevent the abutting joint faces from being forced apart but also to support the high plunging forces applied by the FSW machine head. The length of the pin is slightly less than the required weld depth and the tool shoulder should be in direct contact with the surface of the workpiece. The probe is submerged into the workpiece and then the tool is moved along the weld line with a tilt angle of 2-4 degress increasing the pressure under the tool shoulder.

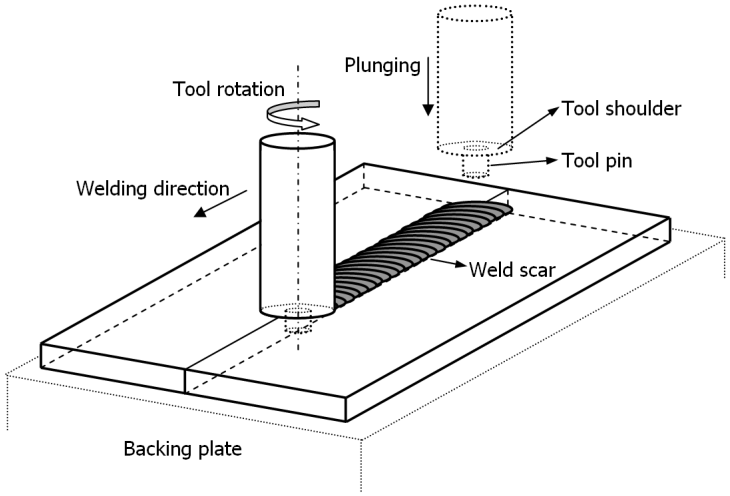
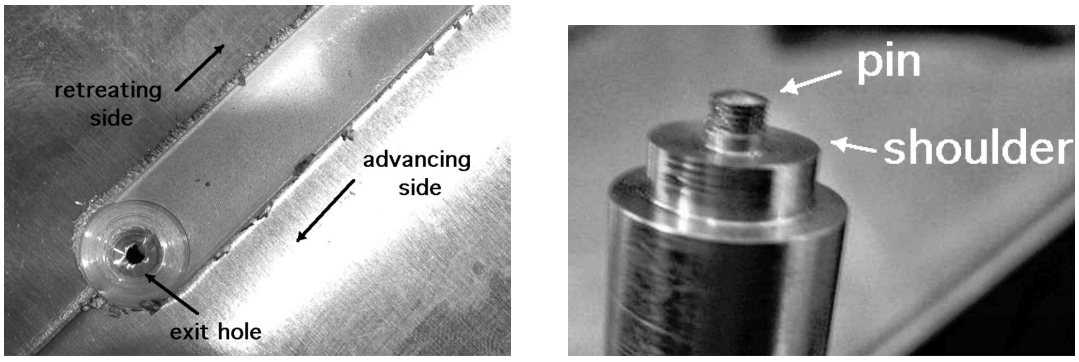


Figure 1.1: Schematic view of the FSW process [ref. PAPER-III] (Tutum et al., 2009).

Frictional heat is generated between the wear-resistant welding tool shoulder and pin on one side, and the material of the work pieces on the other. This heat, along with the heat generated by the plastic dissipation due to the mixing process, causes the stirred materials to soften without reaching the melting point (hence cited a solid-state process), allowing the traversing of the tool along the weld line in a plasticised tubular shaft of metal. As the pin is moved in the direction of welding, the leading face of the pin, often assisted by a special pin profile, forces plasticised material to the back of the pin while applying a substantial forging force to consolidate the weld metal. The welding of the material is facilitated by severe plastic deformation in the solid state, involving dynamic recrystallization of the weld nugget (Murr et al., 1997).



(a) (b)

Figure 1.2: (a) An FSW weld between aluminium sheets (Nandan et al., 2008a). (b) An actual tool, with a threaded-pin (Nandan et al., 2008a).

The solid-state nature of the FSW process, in combination with its unusual tool and asymmetric nature, results in a very characteristic microstructure. While some regions are common to all forms of welding some are unique to the FSW process. While the terminology is varied the following is representative of the consensus.

- The *stir zone* (also nugget, dynamically recrystallised zone), D in Fig. 1.3, is a region of heavily deformed material that roughly corresponds to the location of the pin during welding. The grains within the stir zone are roughly equiaxed and often an order of magnitude smaller than the grains in the parent material (Murr et al., 1997). A unique feature of the stir zone is the common occurrence of several concentric rings which has been referred to as an *onion-ring* structure. The precise origin of these rings has not been firmly established, although variations in particle number density, grain size and texture have all been suggested.

- The $flow\ arm$ is on the upper surface of the weld and consists of material that is dragged by the shoulder from the retreating side of the weld, around the rear of the tool, and deposited on the advancing side of the weld.

- The *thermo-mechanically affected zone* (TMAZ), C in Fig. 1.3, is situated on either side of the stir zone. In this region the strain and temperature are lower than in the stir zone and the effect of welding on the microstructure is correspondingly smaller. Unlike the stir zone the microstructure is recognizably that of the parent material, albeit significantly deformed and rotated. Although the term TMAZ technically refers to the entire deformed region it is often used to describe any region not already covered by the terms stir zone and flow arm.

- The *heat-affected zone* (HAZ), B in Fig. 1.3, is common to all welding processes. As indicated by the name, this region is subjected to a thermal cycle but is not deformed during welding. The temperatures are lower than those in the TMAZ but may still have a significant effect if the microstructure is thermally unstable. In fact, in age-hardened aluminium alloys this region commonly exhibits the poorest mechanical properties.

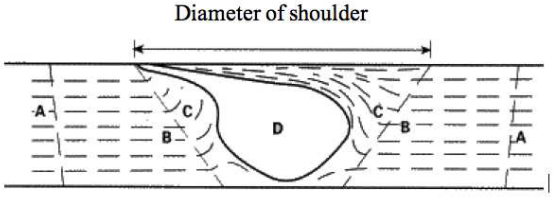


Figure 1.3: Schematic view of the weld zone with different regions (Frigaard and Grong Ø, 2001).

Some of the advantages of FSW are improved mechanical properties of the welds, reduced distortion and residual stresses, and benign environmental characteristics (Mishra and Ma, 2005). The requirement for lighter and load resistant structures, especially in aerospace and automotive industries, emphasizes the need for investigating the efficient choice of the important process parameters that control the FSW welding procedure.

The solid-state nature of FSW immediately leads to several advantages over fusion welding methods since any problems associated with cooling from the liquid phase are immediately avoided. Issues such as porosity, segregation, solidification cracking and liquation cracking are not an issue during FSW. In general, FSW has been found to produce a low concentration of defects and is normally very tolerant to variations in parameters

and materials.

Nevertheless, FSW is associated with a number of unique defects. For example, insufficient weld temperatures, due to low rotational speeds or high traverse speeds, meaning that the weld material is unable to accommodate the extensive deformation during welding. This may result in long, tunnel-like defects running along the weld which may occur on the surface or subsurface. Low temperatures may also limit the forging action of the tool and thus reduce the continuity of the bond between the material from each side of the weld. The light contact between the material has given rise to the name *kissing bond*. This defect is particularly worrying since it is very difficult to detect using nondestructive methods such as X-ray or ultrasonic testing. If the pin is not long enough or the tool rises out of the plate then the interface at the bottom of the weld may not be disrupted and forged by the tool, resulting in a lack-of-penetration defect. This is essentially a notch in the material which can be a potential source of fatigue cracks.

A number of potential advantages of FSW over conventional fusion-welding processes can be summarized as below:

- Good mechanical properties in the as welded condition
- Improved safety due to the absence of toxic fumes or the spatter of molten material.
- No consumables - conventional steel tools can weld over 1000 m of aluminium and no filler or gas shield is required for aluminium.
- Easily automated on simple milling machines - lower setup costs and less training.
- Can operate in all positions (horizontal, vertical, etc), as there is no weld pool.
- Generally good weld appearance and minimal thickness under/over-matching, thus reducing the need for expensive machining after welding.
- Low environmental impact.

However, some disadvantages of the process can be listed below:

- Exit hole left when tool is withdrawn.
- Large down forces required with heavy-duty clamping necessary to hold the plates together.
- Less flexible than manual and arc processes (difficulties with thickness variations and non-linear welds).
- Often slower traverse rate than some fusion welding techniques although this may be offset if fewer welding passes are required.

The shipbuilding and marine industries are two of the first industry sectors which have adopted the FSW process for commercial applications. The process is suitable for the following applications: panels for decks, sides, bulkheads and floors, aluminium extrusions, hulls and superstructures, helicopter landing platforms, offshore accommodation, marine and transport structures, masts and booms, e.g. for sailing boats and finally refrigeration plants.

At present the aerospace industry is welding prototype and production parts by FSW. Opportunities exist to weld skins to spars, ribs, and stringers for use in military and civilian aircraft. The Eclipse 500 aircraft, in which $\simeq$ 60% of the rivets are replaced by friction stir welding, is now in production. This offers significant advantages compared to riveting and machining from solid, such as reduced manufacturing costs and weight savings. Longitudinal butt welds in Al alloy fuel tanks for space vehicles have been friction stir welded and successfully used. The process could also be used to increase the size of commercially available sheets by welding them before forming. The FSW process can therefore be considered for: wings, fuselages, empennages, cryogenic fuel tanks for space vehicles, aviation fuel tanks, external throw away tanks for military aircraft, military and scientific rockets, repair of faulty MIG welds, various primary and secondary structural components.

Several examples on the commercial production of high speed trains made from aluminium extrusions which may be joined by FSW are given in Skillingberg and Green (2007); TWI (1991). Applications include: high speed trains, rolling stock of railways, underground carriages, trams, railway tankers and goods wagons as well as container bodies.

The FSW process is currently also being used commercially in the automotive industry. Existing and potential applications include: engine and chassis cradles, wheel rims, attachments to hydroformed tubes, tailored blanks, e.g. welding of different sheet thicknesses, space frames, e.g. welding extruded tubes to cast nodes, truck bodies, tail lifts for lorries, mobile cranes, armour plate vehicles, fuel tankers, caravans, buses and airfield transportation vehicles, motorcycle and bicycle frames, articulated lifts and personnel, bridges, skips, repair of aluminium cars, magnesium and magnesium/aluminium joints.

Finally, FSW can also be considered for other applications, such as electric motor housings (in production), refrigeration panels, cooking equipment and kitchens, white goods, gas tanks and gas cylinders and connecting of aluminium or copper coils in rolling mills.

1.3 Structure of the thesis

The thesis constitutes of 5 chapters followed by 4 appended papers and 2 technical reports.

In *Chapter* 1, the motivation of the work has been given with an overview of the needs for investigations in the FSW process followed by some examples of industrial applications.

Chapter 2, is overall speaking devoted to modeling of the FSW process and it is initiated by the thermal modelling and then combined with the mechanical modeling where transient thermal as well as residual stresses together with a real-world service load application, is presented.

Analytical models, precribed heat source models and the thermal-pseudo-mechanical heat source model constitute the backbone of the thermal modeling section, which is also the main driver for the mechanical models presented in the subsequent section of the modeling chapter. Except for the steady-state analytical (thin plate) thermal model presented in `PAPER-I`, the rest of the thermal models are dealing with both steady-state and transient cases. `PAPER-I` also includes a comparison of a relatively simple analytical solution with a more complex 2-D Eulerian model in order to be used for obtaining an average temperature under the tool shoulder.

The mechanical modeling section is mainly based on `PAPERs-II-III and IV` although some of them are also optimization oriented. An overview of the governing equations for the thermo-mechanical modeling and the evolution of residual stresses for welding applications are given first and then an analytical thermo-elastic solution for the moving line-type heat source is presented in order to initiate the discussions for the fundamental issues in thermo-mechanics. Next, numerical finite element models for the residual stress calculations investigating the different clamping conditions and different levels of temperature dependency of the yield stress, using prescribed and TPM-heat source based thermal models (which are implemented using commercial software, i.e. ABAQUS and ANSYS) are given in detail. Following these numerical models, mechanical properties of friction stir welded structures is presented briefly with an accompanying softening (dissolution) model.

As a final topic of the modeling chapter the integrated modeling approach, applied to FSW for simulation of the service-load performance of a welded structure and a damage evolution in a FSW joint specimen during in-service conditions, is presented.

Chapter 3, considers mainly optimization studies aiming at process-specific goals or success criteria satisfying limitations related to the production. A brief overview on optimization methodologies is given initially, including deterministic traditional techniques and nature inspired undeterministic approaches (i.e. genetic algorithms) as well as hybrid methodolgies aiming at combining the advantages of both strategies for multi-objective optimization cases in particular. A brief description of the algorithms, i.e. the Sequential Quadratic Programming (SQP) as the classical gradient based technique and the Non-dominated Sorting Genetic Algorithm (NSGA-II) as the multi-objective evolutionary algorithm, is given. Following these introduction sections, two optimization problems based on thermal and thermo-mechanical (residual stress) models, are presented in detail.

The first deals with both analytical and numerical thermal models combined with the Sequential Quadratic Programming (SQP) gradient-based optimization algorithm in order to find the welding speed and the heat input that would yield a prescribed average temperature close to the solidus temperature under the tool, thereby expressing a condition which is favourable for the process.

The second applies the thermomechanical models predicting residual stresses in combination with the evolutionary optimization algorithm (NSGA-II) in the search for the optimal combination of the process parameters of welding speed and rotational speed, which would minimize residual stresses and maximize the welding speed divided by rotational speed (known as advancement per revolution and expressing a desired feature of the process).

Chapter 4 is dedicated to summary of appended 4 articles and 2 technical reports.

Chapter 5 presentes the conclusions of the thesis and suggestions for future research directions.

Finally, in the two technical reports, focus is put on the well-known benchmark problems for multi-objective optimization with and without constraints. In these, a comparison of three algorithms, i.e. the MOGA-II and the NSGA-II in the commercial multi-objective optimization software modeFRONTIER together with the custom implementation by the author of the NSGA-II in MATLAB, is carried out.

Chapter 2

Modeling

This chapter focuses on the thermal and thermo-mechanical modeling studies in the FSW process ranging from 2-D analytical models to 3-D numerical models. First, thermal models which are mainly the driving force for the thermo-mechanical models, i.e. residual stress models, are presented in detail. Following that, thermo-mechanical models including residual stress calculations and integrated modeling studies, i.e. combining with service load performance of the welded structure, are presented.

2.1 Thermal Modeling of FSW

In the FSW process shown schematically in Fig. 1.1, heat is generated by friction and by the plastic deformation that occur between the tool and the work piece. The heat flows into the work piece as well as the tool. The amount of heat that is conducted into the work piece influences the quality of the weld, as well as the distortion and residual stresses in the work piece. For example, insufficient heat generation could lead to failure of the tool pin since the work piece material is not soft enough. Therefore, understanding the heat flow aspects of the FSW process is extremely important, not only for understanding the physical phenomena but also for improving the process efficiency in order to obtain good weld properties.

Thermal modelling has since the late nineties been a central part of modelling of FSW in general (Schmidt and Hattel, 2008a). One of the reasons for this is that many of the properties of the final weld are directly a function of the thermal history which the workpiece has been exposed to. Secondly, the FSW process itself is highly affected by the heat generation and heat flow and thirdly, from a modelling viewpoint, thermal modelling of FSW can be considered the basis of all other models of the process, be it microstructural, Computational Fluid Mechanics (CFD) or Computational Solid Mechanics (CSM) models.

In the FSW process the welding parameters, i.e. tool rotational speed, traverse speed, etc., are all chosen such that the softening of the workpiece material enables the mechanical deformation and material flow. However, unlike many other thermomechanical processes, the mechanisms of FSW are fully coupled meaning that the heat generation is related to material flow and frictional/contact conditions and vice versa. Thus, in theory a thermal model alone cannot predict the temperature distribution/history without a pre-knowledge about the heat generation, since the fundamental mechanisms of FSW are

not part of a pure thermal model. For this reason, several analytical expressions have been given in literature for the heat generation for a given weld as a function of tool geometry and welding parameters, e.g. tool radius and rotational speed (Schmidt and Hattel, 2008a; Hattel et al., 2008).

2.1.1 Governing Equations

All thermal models, independent of their complexity, rely on the solution of the classical heat conduction equation, Eq. (2.1), which is a scalar partial differential equation (PDE) in mathematics parlance (Betounes, 1998; Carslaw and Jaeger, 1959). This time-dependent problem can be solved equivalently in both Lagrangian and Eulerian reference frames with an appropriate set of initial and boundary conditions. In the former case, the thermal field is described with respect to a fixed coordinate system whereas in the latter case the coordinate system moves with the heat source as schematically shown in Figure 2.1. The temperature change in a region over time is formulated in the Lagrangian reference frame as

$$\rho c_p \frac{\partial T}{\partial t} = \nabla(k \nabla T) + q_{vol} \tag{2.1}$$

where ρ denotes the material density $\left(\frac{kg}{m^3}\right)$, c_p the specific heat capacity $\left(\frac{J}{kg\,K}\right)$, T the temperature field (K), k the thermal conductivity $\left(\frac{W}{m\,K}\right)$, and q_{vol} the volumetric heat source term $\left(\frac{W}{m^3}\right)$. Under the assumption of constant thermal conductivity, independent of temperature, Eq. (2.1) can be reformulated,

$$\frac{\partial T}{\partial t} = \frac{k}{\rho c_p} \nabla^2 T + \frac{q_{vol}}{\rho c_p} \tag{2.2}$$

where the constant term $\left[\frac{k}{\rho c_p}\right]$ is called the thermal diffusivity of the material which is a measure of how fast heat is conducted in a solid (Carslaw and Jaeger, 1959; Lindgren, 2007).

In case of describing the heat flow in a Eulerian reference frame, a convective term is added to Eq. (2.1) as given by

$$\rho c_p \frac{\partial T}{\partial t} = \nabla(k \nabla T) + q_{vol} - \rho c_p u \nabla T \tag{2.3}$$

where u is the velocity field vector which can contain the tool welding velocity as well as the material flow field around the tool probe, i.e. in the shear layer.

If transient effects are negligible, the time-dependent term, i.e. $\frac{\partial T}{\partial t}$ vanishes and Eq. (2.1) in the Lagrangian reference frame is reduced to

$$0 = \nabla(k \nabla T) + q_{vol} \tag{2.4}$$

and correspondingly in the Eulerian reference frame Eq. (2.5) is obtained.

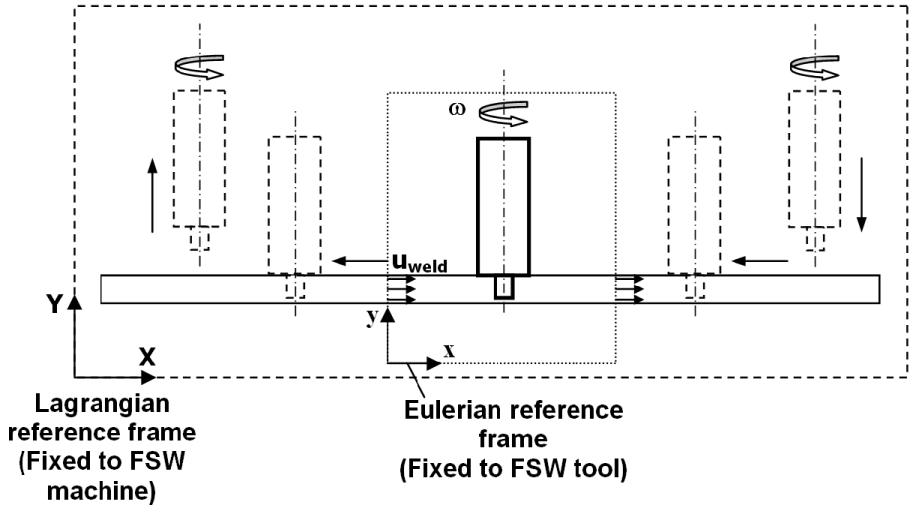


Figure 2.1: Schematic view of the Lagrangian and the Eulerian reference frames.

$$0 = \nabla(k\nabla T) + q_{vol} - \rho c_p u \nabla T \tag{2.5}$$

Although the use of the former approach is quite common, the latter can take advantage of the steady state conditions that may exist with respect to the moving heat source (Lindgren, 2007). These formulations are summed up in tabular form to give an overview in Table 2.1. Besides some numerical applications of the latter, it also results in important analytical formulations to be presented in detail in the following section. The combination of the choice of reference frame and degree of enmeshed modeling level offers a huge range of possibilities, and the "correct" choice depends on the objective of each model. As an example, a residual stress model would need a temperature field coming from a transient Lagrangian thermal model (Schmidt and Hattel, 2008a).

Table 2.1: Governing (heat conduction) equations in thermal models with respect to different reference frames and time domains [ref. `PAPER-IV`] (Hattel et al., 2008)

Reference Frame	Transient	Steady-state
Lagrangian	$\rho c_p \frac{\partial T}{\partial t} = \nabla(k\nabla T) + q_{vol}$	$0 = \nabla(k\nabla T) + q_{vol}$
Eulerian	$\rho c_p \frac{\partial T}{\partial t} = \nabla(k\nabla T) + q_{vol} - \rho c_p u \nabla T$	$0 = \nabla(k\nabla T) + q_{vol} - \rho c_p u \nabla T$

2.1.2 Analytical thermal models in FSW

An accurate solution of the complete heat conduction equation considering heat transfer by both conduction and convection is complicated. As a first step, it is often useful to discuss a simplified solution considering only conduction heat transfer. Following that, other simplifications can be sorted as using constant (temperature independent) material properties, i.e. k, ρ and c_p, simplifications in the heat sources (Gaussian distribution of the heat source, double elipsoid heat source as well as point and line heat sources),

time dependence (assuming steady-state conditions) and boundary conditions (infinite or semi-infinite domains). These simplifications are attractive since analytical solutions can be obtained for the heat conduction equation in many situations and these solutions can provide interesting insight about the welding process (Feng, 2005). In the next section, an analytical model particularly applicable for FSW process will be described in detail and it will constitute the basis for Section 3.3.1, relating closely to `PAPER-I` (Tutum et al., 2007).

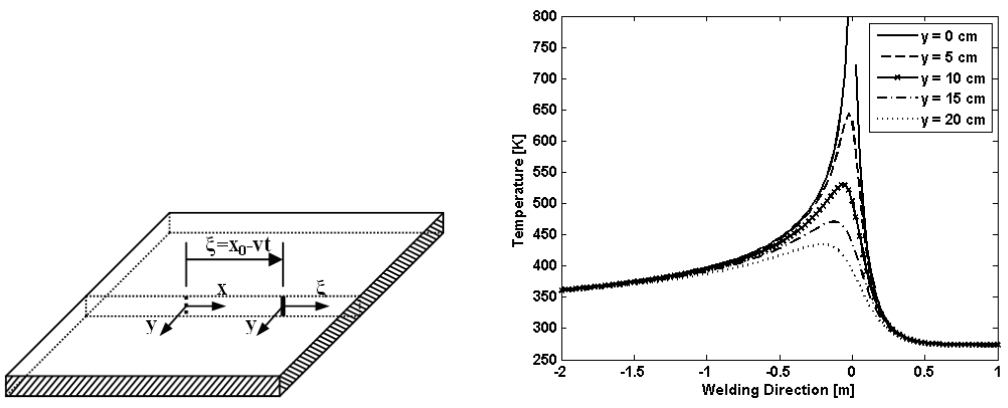
2.1.2.1 Application of the Thin-Plate Solution in FSW

A schematic view of a line-type heat source moving with a constant speed along a line in an infinitely extended plate is shown in Fig. 2.2a. The temperature field is governed by the same heat conduction partial differential equation, including the convection term, as given in Eq. (2.3). The closed form solution can be obtained under quasi-steady state assumptions as given by Eq. (2.6) (Rosenthal, 1946),

$$T = \left(\frac{Q_{tot}}{2\pi k d}\right) e^{\frac{-u_{weld}\xi}{2a}} K_0\left(\frac{u_{weld}}{2a} r\right) + T_0 \tag{2.6}$$

where (Q_{tot}/d) defines the heat input per unit thickness, ξ denotes the moving coordinate axis ($r = \sqrt{\xi^2 + y^2}$), a is the heat diffusivity, K_0 denotes the Modified Bessel Function of the second kind and zeroth order and T_0 defines the ambient temperature.

Fig. 2.2b shows the resulting temperature profiles along the welding axis at different distances, i.e. $y = 0, 5, 10, 15$ and 20 cm away from the joint line, while the heat source is passing by $x = 0$ m. Notice the mathematical singularity of infinite temperature at the location of the heat source, which of course is not possible in real FSW applications where the solidus temperature is the limit for peak temperatures.

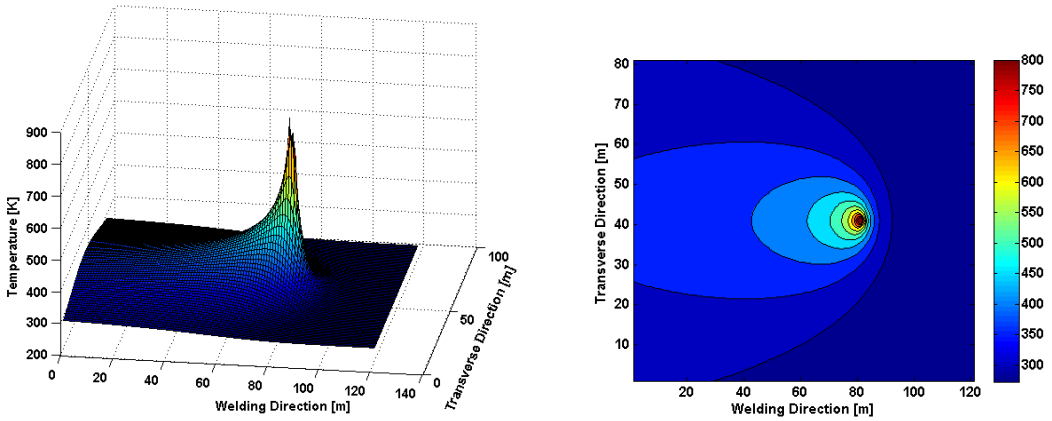


(a) (b)

Figure 2.2: (a) Schematic view of a line-type heat source moving along the joint line of the work piece [ref. `PAPER-I`] (Tutum et al., 2007). (b) Temperature profile along x-axis at different y-points (Tutum and Hattel, 2008).

Fig. 2.3a shows the thermal field where only the 1.2 m x 1 m region surrounding the heat source is considered. Notice from Fig. 2.3b that the analytical model does not see the effects of the boundary conditions because of the infinitely large domain description,

while this is not the case for the numerical model having a finite domain, see Section 2.1.3.1, Fig. 2.8.



(a) (b)

Figure 2.3: (a) Temperature field on a 3 mm-thick plate (Tutum and Hattel, 2008). (b) Contour plot of the temperature field (Tutum and Hattel, 2008).

It should be mentioned that a modified version of this solution given by Eq. (2.6) is used later in Chapter 3 for obtaining optimal values of the heat input and the welding velocity. Details about this are given in Section 3.3.1.

The Rosenthal solutions described above have been widely used, especially in the early modelling of FSW. In Gould and Feng (1998) and McClure et al. (1998) the 3D Rosenthal solution is used to develop a circular heat source resembling the shoulder of the tool by placing sources in a ring around the tool center and integrating to obtain the full temperature field. The heat is assumed to be generated by Coulomb friction between the tool and the workpiece, i.e. $\delta = 0$, see Eq. (2.7) for definition. Russell and Shercliff (1999) presented an analytical model using the Rosenthal point heat source solution as a thermal model in order to be used as an input for the subsequent microstructure (softening) model based on the method of Myhr and Grong (1991a,b), which has proven to be sufficiently accurate for FSW of many Al-alloys. The softening of the material (dissolution of the precipitates in the alloy) is mainly governed by the thermal cycles, which are integrated over time, rather than the peak temparatures. Fonda and Lambrakos (2002) presented a more detailed analytical model based on the 3D Rosenthal solution involving finite plate thickness and distribution of several point heat sources around the tool. By using an inverse modelling technique, the temperature field described by the peak temperatures, is adjusted by scaling the power of the heat sources with respect to the hardness measurements. Vilaça et al. (2005) developed an analytical thermal model for 2D and 3D cases using modified Rosenthal solutions, again based on a reverse engineering approach, i.e. modifying the process parameters iteratively along with thermal measurements, enabling simulation of the asymmetric temperature profile induced by the tool rotations. Finally, Larsen et al. (2009) used the 2D Rosenthal solution to optimize a FSW thermal model by using space mapping and manifold mapping in which a coarse model is used along with a fine FE-based model to be optimized.

2.1.3 Prescribed heat source models in FSW

In most pure thermal models of FSW, the heat generation from both frictional and plastic dissipation is modelled via a surface flux boundary condition at the tool/matrix interface. The main unknown parameters in these surface flux expressions are either the friction coefficient under the assumption of sliding and the material yield shear stress under the assumption of sticking.

Along this line, Schmidt and Hattel (2004) have proposed the following generally adopted equation for the total heat generation $[W]$,

$$\begin{aligned} Q_{total} &= \delta Q_{sticking} + (1-\delta)\, Q_{sliding}, \\ &= \frac{2}{3}\pi\omega \left[\delta\tau_{yield} + (1-\delta)\,\mu p\right] \left[\left(R^3_{shoulder} - R^3_{probe}\right)(1-\tan\alpha) + R^3_{probe} + 3R^2_{probe}H\right] \end{aligned} \tag{2.7}$$

where δ is the dimensionless slip rate between the tool and the workpiece (1 for full sticking and 0 for full sliding).

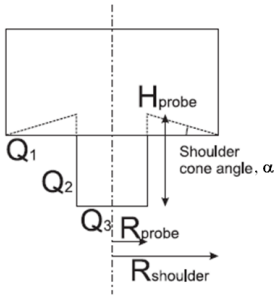


Figure 2.4: Schematic view of analytical tool geometry (including conical shoulder and cylindrical probe and heat generation contributions from shoulder Q_1, probe side Q_2 and probe tip Q_3 (Schmidt and Hattel, 2004, 2005).

However, when implementing this into a numerical model using a position dependent surface flux [W/m^2], it is typically used on the following form (Schmidt and Hattel, 2005),

$$\begin{aligned} q_{total} &= \dot{\gamma}\tau_{friction} + (\omega r - \dot{\gamma})\,\tau_{yield}, \\ &= \omega r \left[\delta\tau_{yield} + (1-\delta)\,\mu p\right] \end{aligned} \tag{2.8}$$

which in fact is the basis for deriving equation, Eq. (2.7). Furthermore, when combining Eq. (2.7) and (2.8) and then assuming the simple tool geometry of a flat shoulder, one obtains the following well-known expression for the heat generation (Schmidt and Hattel, 2005),

$$q_{total} = \frac{3Q_{total}r}{2\pi R^3_{shoulder}} \tag{2.9}$$

which can be applied as a radius dependent surface flux in the model, under the assumption of a constant contact condition close to sliding or in cases of sticking where the

shear layer is very thin.

From Eq. (2.9) it can be seen that Q_{total} can be considered an input parameter in the same manner as the friction coefficient, the pressure and the material's yield shear stress in Eq. (2.7). However, it should be mentioned that having Q_{total} as part of the input parameters for the model in some situations could conflict with the very objective for thermal modelling of FSW. This is especially the case when the model should predict temperatures for conditions not supported by measurements of the heat input Q_{total}. In these cases it is not straightforward to predict or simulate the effect of e.g. a change in welding or rotational speed, because the total heat generation Q_{total} is a function of these changes in parameters and hence can be considered an internal function of the FSW process; unlike other welding process, e.g. TIG where the heat input is controlled externally.

This dilemma can obviously be overcome by supporting the thermal model with a thermo-mechanical CSM or CFD based model which "includes" the underlying physics of the process, namely the material flow producing heat generation by plastic dissipation in the shear layer and frictional contact at the tool/workpiece. As being pretty obvious from the later section on thermo-mechanical modeling, the efforts for such simulations are highly demanding from both a computational and a human resource point of view, thus leaving pure thermal modelling of FSW as a valuable "simple" alternative for simulation of heat flow having in mind the limitation of such models, which as earlier mentioned primarily lie in the evaluation of Q_{total}.

Having said that, it should be noted that the most utilized procedure for evaluating Q_{total} typically is to rely on experimental findings by simply performing the actual welds and measuring Q_{total} with a dynamometer, and thereby accepting the inherent limitation of the resulting thermal model to predicting only temperatures for a known total heat generation.

Since the prescribing of the heat generation is the most single important parameter in thermal models of FSW, Schmidt and Hattel (2005) proposed a classification of different heat sources as shown in Fig. 2.5. One characteristic in this classification is how "detailed" the tool heat generation is resolved. Three levels were evaluated, i.e. i) including shoulder heat generation, without probe heat generation, ii) including shoulder and probe heat generation, with probe material and iii) shoulder and probe heat generation but without probe material.

A second characteristic was whether the convective contribution due to the material flow in the shear layer was taken into account. Two extreme contact conditions were evaluated, i.e. full sliding and full sticking. In the case of sliding the heat was applied as a surface flux and in the case of sticking the heat was applied as a volume flux in a shear layer. This shear layer was prescribed analytically assuming a uniform thickness with a linear velocity profile ramped between the tool velocity at the contact interface and the welding velocity outside the shear layer.

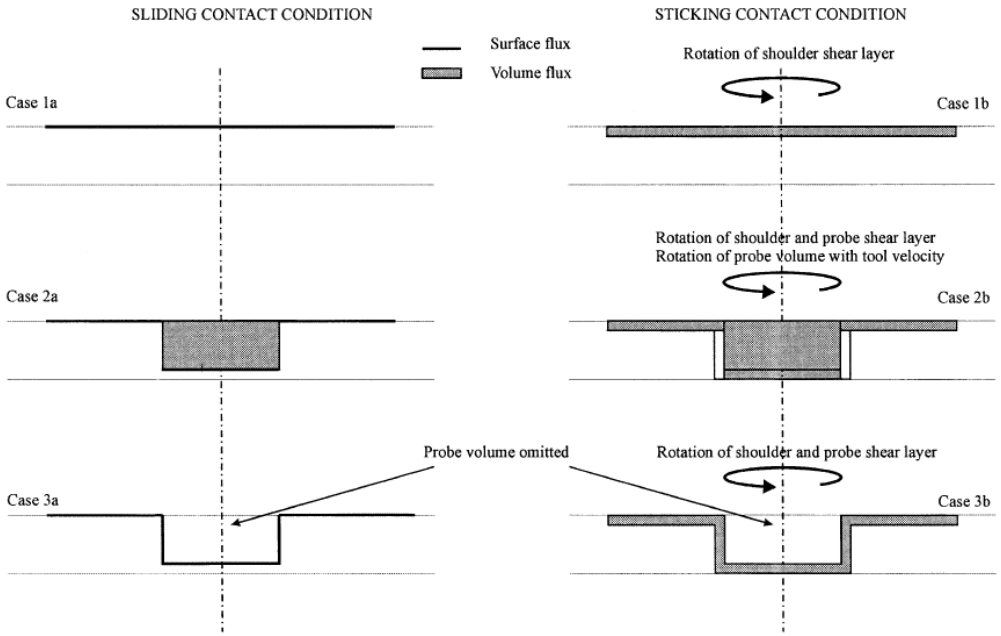


Figure 2.5: Schematic view of configuration of six cases of prescribing the heat generation in FSW: the left-hand column is for sliding condition and the right-hand column for sticking condition; the first row shows no probe heat generation, the second row is with probe heat generation, and the third row is for the probe volume removed, and heat generation in the shear layer (Schmidt and Hattel, 2005).

It was concluded that for analysing the temperature field in the volume under the tool, special attention should be paid to the heat generation and how the material "flows" around the tool probe. The main effect of including the probe in the thermal model is to change the material properties to those of the tool; secondly, rotating the probe and modelling the shear layer around it such that it resembles a flow model.

One new procedure to obtain the heat generation is to couple the traditional analytical expression for the heat generation with a constraint based on experimental or phenomenological considerations. It is well-known from the constitutive behaviour of a solid (representative for those alloys used in FSW), that the yield stress dramatically decreases once the temperature approaches the solidus temperature. Above the solidus temperature, the material acts as a fluid. As a consequence, the material close to the tool/matrix interface will reduce its heat generation to negligible values if exceeding the solidus temperature subsequently reducing the temperature level, allowing the material to recover its strength. A selfstabilizing effect will thus establish at a temperature level around the solidus temperature, hence this could be used as an average temperature constraint in a pure thermal model. This is what is done in the thermal model by Tutum et al. (2007) [ref. `PAPER-I`], where the heat generation in FSW of AA2024 is controlled by using an optimization scheme such that an average temperature of 500 °C at the tool/-workpiece interface is obtained. This procedure has given promising results for obtaining temperature fields without prior knowledge regarding Q_{total}; in fact this is an output of the optimization analysis. A more thorough description of this approach is given in Section 3.3.

2.1.3.1 Steady-state Eulerian models with precribed heat source

Modeling the whole welding process, i.e. plunge, dwell and pull out periods, holds some notable complexities. In order to reduce the computational cost regarding the moving heat source, meanwhile preserving the applicability, only the welding period is taken into account and a moving coordinate system which is located on the heat source is applied. Eq. (2.2) describes the heat transfer in the plate in general and it is modified into Eq. (2.5) as described in Section 2.1.1. The traverse motion of the tool is modeled by prescribing a material flow through the rectangular plate region. Due to this flow prescription, the equation includes a convective term in addition to the conductive term. If the heat flow problem involves material flow, the Eulerian formulation is often advantageous. In Fig. 2.6, the boundary conditions are described for the steady state Eulerian model in the simulation tool COMSOL.

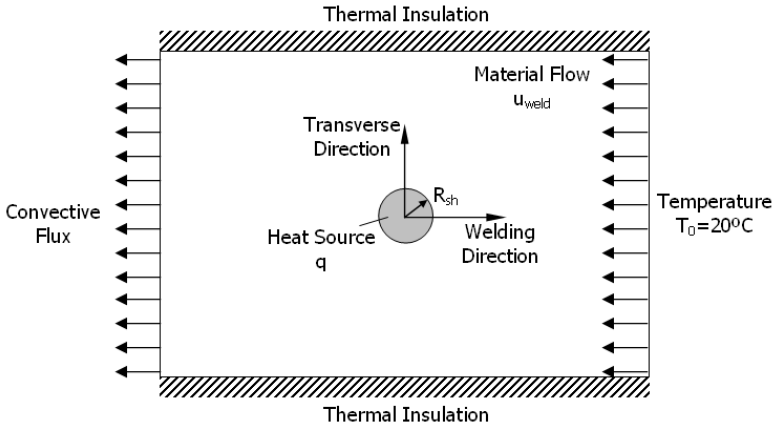


Figure 2.6: Schematic view of the numerical model with boundary conditions [ref. `PAPER-I`] (Tutum et al., 2007).

The circle with a radius of R_{sh}=10 mm, in the middle of the rectangular region, represents the tool shoulder where the stationary and uniform heat source is applied as a volume flux in the shear layer. The tool is stationary while the material flows through it with a velocity of u_{weld} in the opposite welding direction, in Fig. 2.6. The room temperature (20 °C) is defined at the right edge of the rectangular region. The heat flux on the left edge of the plate region, where the aluminum leaves the computational domain, is dominated by convection. On the upper and lower edges of the plate boundaries, thermal insulation is enforced.

A schematic view of the applied volume heat flux as shown in Fig. 2.7 , is given by,

$$q_{total} = \frac{3Q_{total}r}{2\pi R^3_{shoulder}d} \tag{2.10}$$

where Q_{tot} is the total heat input, d denotes the thickness of the actual shear layer under consideration and r is the radius originating at the center of the tool. This volume heat flux corresponds to a surface heat flux from the tool shoulder (without the tool probe) (Schmidt and Hattel, 2005), distributed through the plate thickness under the assumption of 2D heat flow. This corresponds to Case 1b in Fig. 2.5 for a thin plate.

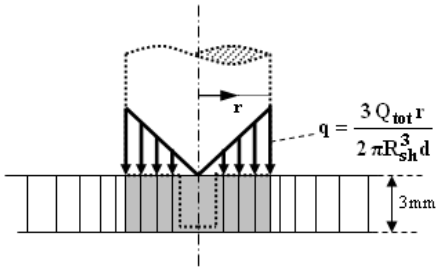


Figure 2.7: Schematic view of tool geometry with applied linear heat source on tool shoulder in 2D thermal model [ref. PAPER-I] (Tutum et al., 2007).

Fig. 2.8 shows the contour plot of the resulting temperature field of the model (isotherms in 50 K increments). The modeling includes aluminum properties for the workpiece material for a 3 mm-thick plate. The welding speed is 2 mm/s. Notice that the maximum temperature is around 800 K, which is a result of having higher temperatures than a desired average temperature over the shoulder area of 500 °C. This is not physically possible since the solidus temperature, T_{sol} , for this case was chosen to be 500 °C ($\simeq$ 773 K); however this is a result of the use of a simple heat source model.

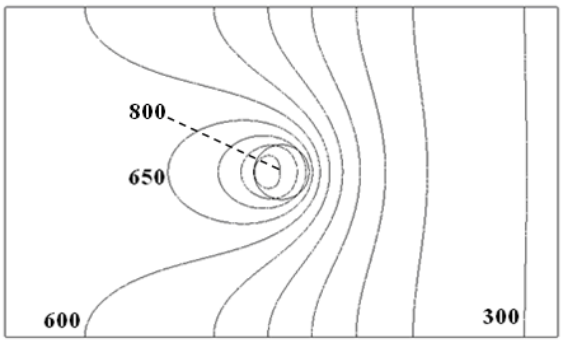
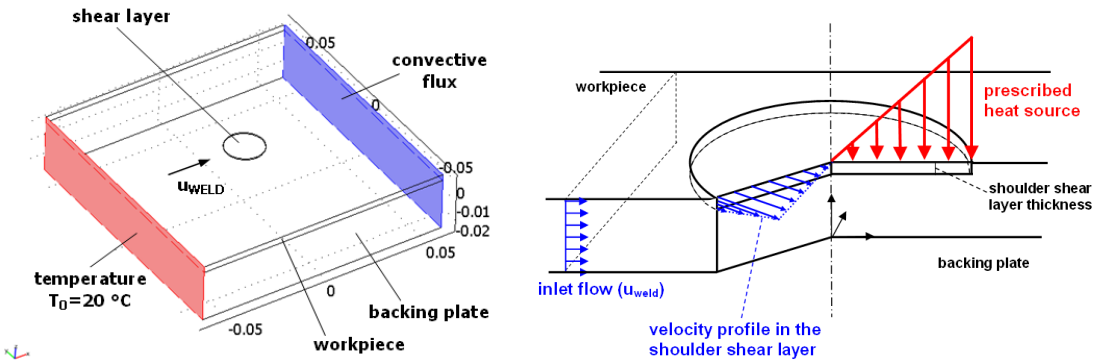


Figure 2.8: Contour plot of the temperature field. Increment between isotherms is 50 K [ref. PAPER-I] (Tutum et al., 2007).

This 2D Eulerian thermal model can be extended into the following 3D Eulerian thermal model, shown schematically in Fig. 2.9, including the prescribed material flow under the tool shoulder (without probe) and the steel backing plate which absorbs the heat from the aluminum workpiece. The convective heat transfer induced by the material flow which is a function of the contact condition, i.e. sliding, sticking or partial sliding/sticking, affects the local temperature distribution close to the tool/matrix interface. Some refinements of the shoulder heat source are particularly examined for the two extreme contact conditions, i.e. full sliding and full sticking, further discussions are given by Schmidt and Hattel (2005), see Fig. 2.5.

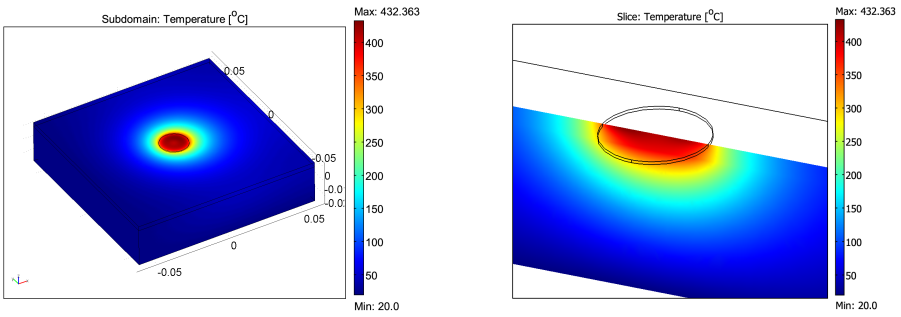
Case-1a considers the full sliding contact condition at the tool shoulder/matrix interface where the contact shear stress described by Coulomb's law ($\tau_{contact} = \tau_{friction} = \mu P$) is less than the material yield shear stress, thus resulting in no workpiece deformation. The heat source is defined as a surface heat flux acting on the top surface of the thin



(a) (b)

Figure 2.9: (a) Geometry of the 3D Eulerian thermal model with boundary conditions (case 1a and 1b). (b) Closer view of the shear layer and prescribed velocity field (inspired by Schmidt and Hattel (2005)).

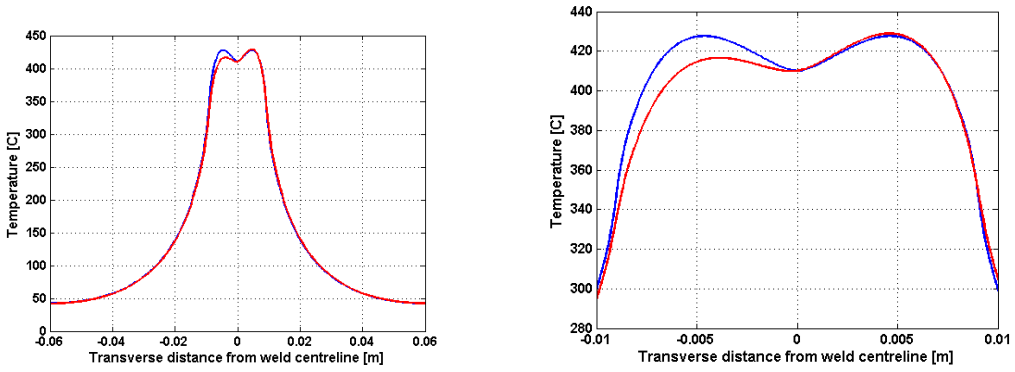
cylindrical volume representing the shear layer shown in Figs. 2.10a and 2.10b. As discussed before in Section 2.1.3, the surface heat flux contributes to the steady-state heat conduction equation, Eq. (2.5), as a boundary condition. Case-1b takes the full sticking contact condition into account at the tool shoulder/matrix interface where the friction shear stress is higher than the matrix yield shear stress denoting that the matrix is deformed with the same rate as the tool rotational speed (a segment at the top surface of the matrix sticks to the tool surface). Since the base matrix is stationary relative to the shoulder/matrix interface, a deformation layer is established. The effect of this layer can be visualized using a ramping factor which scales the velocity of the segments inside the volume in the vertical direction as seen in Fig. 2.9b. In this case the heat source is applied in the thin shear layer as a volumetric heat flux, q_{vol} in Eq. (2.5). The resulting thermal field for this Case-1b is shown in Fig. 2.10a and the asymmetric temperature distribution around the shear layer can be seen in Fig. 2.10b.



(a) (b)

Figure 2.10: (a) Temperature field (Case-1b) (Tutum, 2007b). (b) A cross-sectional view of the temperature field around the shear layer, notice the asymmetric temperature distribution due to the material flow in the shear layer (Tutum, 2007b).

The difference in the temperature distribution between the two cases of 1a and 1b is noticable from the thermal profiles given in Figs. 2.11a and 2.11b. As can be seen it is approximately 15 °C at a distance from the center equal to the half of the tool shoulder radius, particularly in the retreating side.



(a) (b)

Figure 2.11: (a) Comparison of the thermal profiles in Case-1a (symmetric) and Case-1b (asymmetric) along the transverse direction through the center of the heat source at the top surface (Tutum, 2007b). (b) Closer view of the thermal profiles in the diameter of the tool shoulder (Tutum, 2007b).

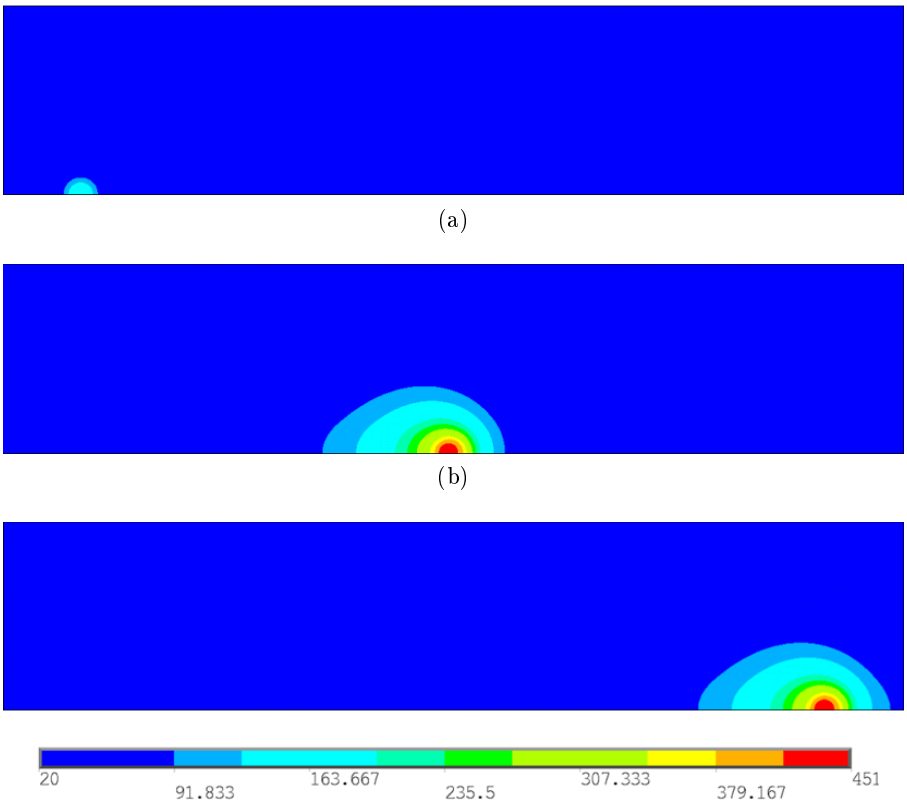
2.1.3.2 Transient Lagrangian models with precribed heat source

Steady-state Eulerian thermal models with prescribed heat sources with respect to different contact conditions have been described in the previous section. However, in order to take the transient effects into account, preparing a Lagrangian model is more efficient compared to a transient Eulerian model, as discussed in Section 2.1.1.

The model presented below represents the welding of two flat plates by considering the bead on plate. Due to symmetry assumptions, i.e. neglecting the asymmetric heat source caused by the asymmetric shear layer formation, only one of the plates is modeled. The dimension of the workpiece is 600 mm x 100 mm x 3 mm. Due to the very low contribution to the heat generation coming from the tool pin, only the tool shoulder is considered in the heat source and it is modelled as a surface heat flux moving along the joint line on the top surface (1.5 mm offset from the middle section) of the shell finite element model. The diameter of the tool shoulder is 20 mm. The moving heat source starts and stops at 50 mm away from the left and the right edges of the plate, respectively. The total heat input in the surface heat flux, given by Eq. (2.9), is 650 W. The effect of the thermal contact with the backing plate is modelled by an equivalent heat transfer coefficient of 500 W/m^2K at the bottom of the workpiece (1.5 mm offset from the middle section) and with an ambient temperature of 20 °C.

The transient evolution of the thermal field is shown in Figs. 2.12a through 2.12c, corresponding to times at 1 s, 125 s and 250 s, respectively. The maximum temperature obtained during welding, at steady-state condition, is 451 °C. Fig. 2.13 shows the thermal profiles with respect to time, recorded at a few positions which makes it clear to see the

transient and steady-state intervals. The peak temperatures keep increasing until 50 s. and then stabilizes at 451 °C until the heat source is turned off at 250 s. and the plate cools down to the room temeperature (20 °C) for a period of 500 s.



(a)

(b)

(c)

Figure 2.12: Temperature distribution in the plate at (a) time=1 s. (b) time=125 s. (c) time=250 s.

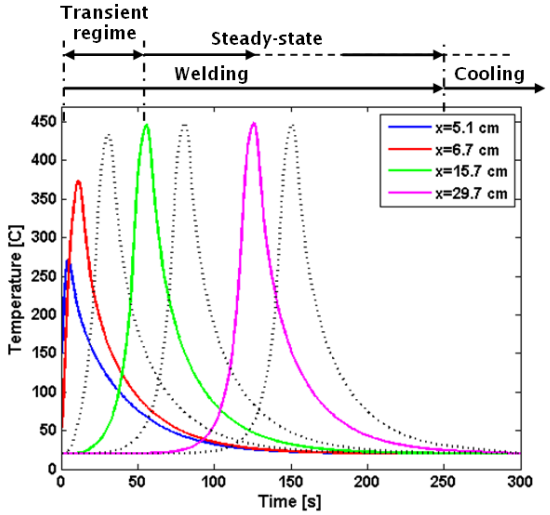


Figure 2.13: Transient and steady-state regimes during welding.

2.1.4 Thermal-Pseudo-Mechanical model

The Thermal-Pseudo-Mechanical (TPM) model, which is proposed by Schmidt and Hattel (2008a,b), is in this section presented in order to simulate the thermal effect of the welding process. As the name indicates, although being purely thermal, the model still takes some mechanical effects into account without actually solving for any mechanical field. The main advantage of this is that the obvious simplicity of thermal models is still maintained in the TPM model. The temperature dependent yield stress of the workpiece material is the main driver for the heat source and is part of the numerical solution itself. This makes the model nonlinear. However, the computational cost is still affordable as compared to thermal calculations that are based on analytical heat source models and it is only a fraction of that of a full thermo-mechanical model. The good correlation of the thermal profiles obtained from experiments with those obtained from the TPM model serves as a validation of the model as well (Schmidt and Hattel, 2008a,b). The model presented here constitutes the basis of the residual stress model which is presented in `PAPER-III` (Tutum et al., 2009).

For the analysis problem, the heat conduction equation, in the form of Eq. (2.1), has to be solved together with an appropriate set of initial and boundary conditions in order to obtain the transient thermal field. In this case q_{vol} is zero because the heat generation is modelled as a surface heat flux from the tool shoulder (without the tool probe). The surface heat flux, q, which is a function of the tool radius and the temperature, is given as follows

$$q(r,T) = \omega r \tau(T) = \left(\frac{n\ 2\pi}{60}\right)\ r\ \frac{\sigma_{yield}(T)}{\sqrt{3}},\ for\ 0 \leq r \leq R_{shoulder} \tag{2.11}$$

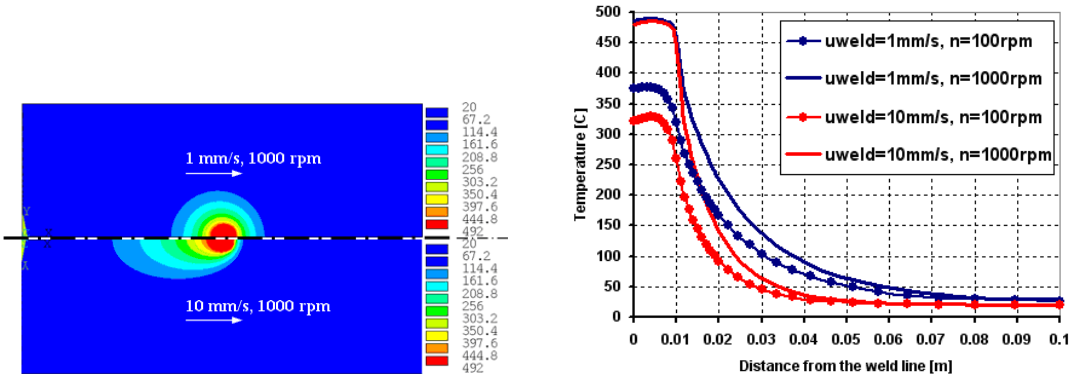
where n is the number of tool revolutions per minute, r is the radial position originating in the tool center and $R_{shoulder}$ is the tool shoulder radius. Finally, the temperature dependent yield stress σ_{yield} is defined as

$$\sigma_{yield}(T) = \sigma_{yield,ref}\ \left(1 - \frac{T - T_{ref}}{T_{melt} - T_{ref}}\right) \tag{2.12}$$

where $\sigma_{yield,ref}$ is the yield stress at room temperature, T_{ref} is 20 °C and T_{melt} is the solidus temperature (500 °C). Once temperatures reach the solidus temperature, i.e. T becomes equal to T_{melt} in Eq. (2.12), the "self stabilizing effect" causes the heat source to "turn itself off", i.e. the material looses all its resistance, and the heat generation decreases automatically due to thermal softening. The effect of the thermal contact with the backing plate is modelled by an equivalent heat transfer coefficient of 700 W/m^{2}K at the bottom of the workpiece and with an ambient temperature of 20 °C. Fig. 2.14a shows the contour plots of the resulting temperature field of the symmetric models (with increments of 47.2 °C) for a chosen welding speed of 1 mm/s and 10 mm/s, respectively, for a rotational speed of 1000 rpm. Fig. 2.14b shows a parameter study for the thermal profiles along the transverse direction and at the middle of the plate with two welding speeds, i.e. 1 and 10 mm/s, and two rotational speeds, i.e. 100 and 1000 rpm. These results clearly show the main characteristics of the TPM model:

- Increasing the welding speed for a fixed rotational speed yields lower temperatures in general, but the decrease in peak temperature vanishes for rotational speeds that are higher than 500 rpm.

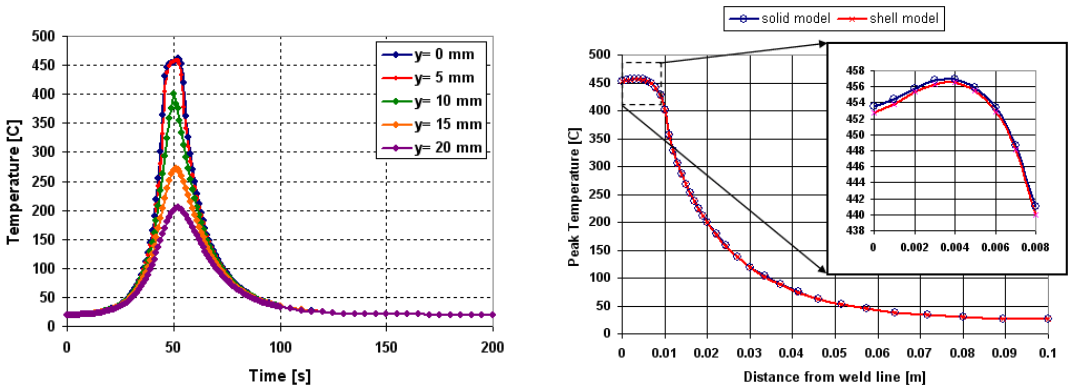
- Increasing the rotational speed for a fixed welding speed results in higher peak temperature and the increase in peak temperature vanishes for higher rotational speeds.
- The gradients in thermal profiles along the transverse direction of the plate become higher with increasing welding speed, while more uniform and wider thermal profiles are obtained for lower welding speeds.



(a) (b)

Figure 2.14: (a) Contour plots of the temperature field with increments of 47.2 °C [ref. PAPER-III] (Tutum et al., 2009). (b) Temperature profiles for different process variables through the middle of the heat source along the transverse direction [ref. PAPER-III] (Tutum et al., 2009).

Transient thermal profiles at different distances in 5 mm increments from the weld line are shown in Fig. 2.15a; here the heat source is located at the middle of the plate. Due to the high heat transfer from the workpiece to the backing plate, the thermal profiles are settled at 20 °C at the end of the cooling session. The accuracy of the thermal simulation using shell elements has been compared with a three-dimensional solid linear 8-node element model. Fig. 2.15b shows the thermal profiles obtained at the middle of the plate along the transverse direction while the tool is traversing with 2 mm/s and rotating with 300 rpm. There is at most a 1 °C difference between two thermal models. The application of the TPM model for the case presented here was published in PAPER-III (Tutum et al., 2009).



(a) (b)

Figure 2.15: (a) Thermal History [ref. PAPER-III] (Tutum et al., 2009). (b) Comparison of temperature profiles obtained by shell (red) and solid (blue) models [ref. PAPER-III] (Tutum et al., 2009).

2.2 Thermo-mechanical Modeling of FSW

Despite the relatively low heat generation during the FSW process, the rigid clamping used in the process gives rise to high reaction forces acting on the plates so as to ensure keeping the plates in place as well as avoid shrinkage of the weld center region. As a result, this generates a high amount of yielding in compression at high temperatures, finally resulting in longitudinal and transverse residual stresses. These residual stresses will act as pre-stresses on the final structure and this might be critical for the fatigue performance during service (James et al., 2007). Although the level of residual stresses resulting from the FSW process is somewhat lower as compared to traditional welding techniques such as fusion welding, it has been shown that the residual stresses play a major role for the fatigue crack growth (Bussu and Irving, 2003) and buckling behavior (Murphy et al., 2007; Bhide et al., 2006), etc. in welded structures obtained using FSW.

The maximum tensile residual stresses are typically found on, or at either side of, the weld line. The mechanisms behind the evolution of residual stresses in welding, in general, are the same, only the magnitudes and distribution of these show some difference depending on the modeling of the heat sources. A schematic view of these thermo-mechanical mechanisms are shown in Fig. 2.16 on a half-plate (under symmetry assumption) clamped on the sides and described with respect to a fixed coordinate system (Lagrangian reference frame) represented by an observer standing on the lower-right side of the workpiece. The thermal history profiles are shown below the workpiece together with the corresponding longitudinal stress-(mechanical) strain curves in the welding direction for convenience since they are dominating. The heat source, i.e. the welding tool, is assumed to be moving from left to right with a constant speed. When the heat source is approaching the observer, where the workpiece material is still at room temperature which is relatively colder than the heat source, the material in front of the tool is heated up and expands meanwhile softening, but since it is constrained by the surrounding colder material, this causes compressive stresses as well as compressive plastic strains after exiting the yield limit. It should be mentioned that the stress-strain curves shown at the lower row of the graphs in Fig. 2.16 for simplicity are schematically drawn under the assumption of ideal

plasticity, i.e. no hardening after yielding as well as no temperature dependency of the yield stress, which is not the case in real applications, but still representative.

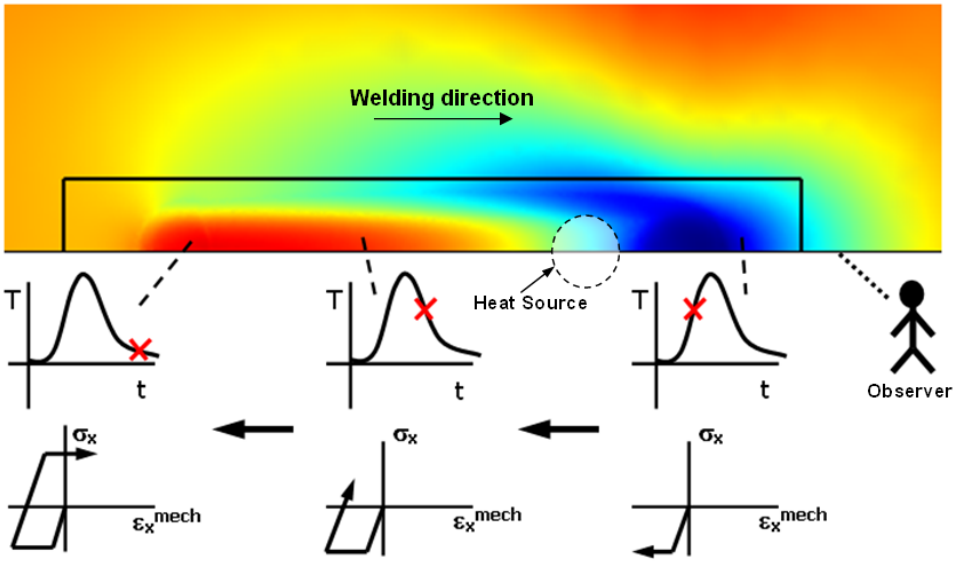


Figure 2.16: Schematic view of how residual stresses evolve. Blue denotes compression and red denotes tension (Tutum, 2007a).

After the heat source passes by, the material in the joint line starts to cool down as seen from the schematic graphs in the middle column in Fig. 2.16. Following the cooling, tensile stresses (or shrinkage forces in other words) start to evolve due to negative thermal strains, i.e. positive mechanical strains due to the constraint, in the longitudinal direction. At the last, left-most graphs, the workpiece cools to the reference room temperature and the tensile stresses, which have been following the stress-strain curve in the elastic regime, eventually reach to the critical level where the material yields in tension. These tensile stresses, so-called residual stresses, lower the loading capacity of the component and the compressive plastic misfit situated at the end of the welding process causes distortion, i.e. shrinkage, in the plate unless some removal techniques, i.e. thermal and mechanical tensioning (Michaleris and Sun, 1997; Michaleris et al., 1999; Richards et al., 2008a), in-process thermal-stretching (Dong et al., 1998) and local-dynamic cooling (Richards et al., 2008b) are applied.

To illustrate the basic formation process of transient thermal stresses and plastic strains in the weld heat-affected zone during heating and cooling stages as well as the significance of the cumulative plastic strains on the final state of residual stresses and distortion, a uniformly heated bar which is fixed at both ends as shown in Fig. 2.17 is considered (Satoh, 1972). The axial stress in this case corresponds to the longitudinal stress in the welding process. The constraints on both ends represent the resistance of the cold material against expanding material due to the heat input. The softening of the surrounding resisting material can also be included by changing the rigid constraint at one end with a spring which is corresponding to the stiffness of the material (Lindgren, 2007; Feng, 2005).

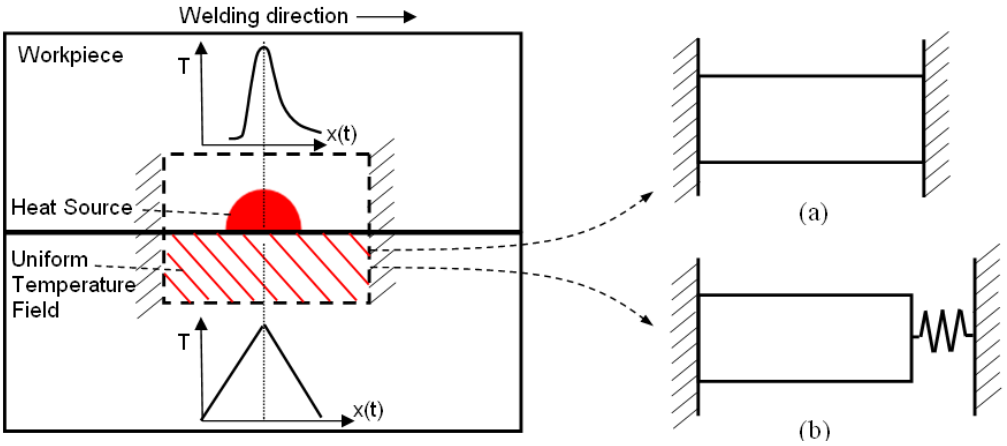


Figure 2.17: Satoh test illustrating the evolution of transient thermal stresses by a both-ends-fixed bar analogy (Tutum and Hattel, 2008).

The perfect rigid constraint (case-a in Fig. 2.17) is an extreme case compared to real welding applications. On the other hand, the modified version (case-b in Fig. 2.17) can be calibrated with respect to experiments, i.e. using both thermal and stress/strain measurements. Besides its theoretical returns and simplicity, the Satoh Test can also be used as a very useful validation tool for testing housebuilt or commercial codes.

2.2.1 Governing Equations

For evaluation of the residual stress field in the workpiece, a standard mechanical model based on the solution of the static force equilibrium equation in the three directions, Eq. (2.13), should be solved for the unknown displacements,

$$\sum_{i=1}^{3} F_i = 0 \quad \Rightarrow \quad \sigma_{ij,i} + p_j = 0 \tag{2.13}$$

where p_j is the body force at any point within the plate and σ_{ij} is the stress tensor. For a strain driven formulation, where the strains are computed from the displacements, the total strain can be decomposed into an elastic part (ε_{ij}^{el}), a plastic part (ε_{ij}^{pl}), a thermal part (ε^{th}), and a visco-plastic part (ε_{ij}^{vp}) as given in Eq. (2.14).

$$\varepsilon_{ij}^{tot} = \varepsilon_{ij}^{el} + \varepsilon_{ij}^{pl} + \delta_{ij}\varepsilon^{th} + \varepsilon_{ij}^{vp} \tag{2.14}$$

For time independent plasticity, which is considered for all thermo-mechanical models in this context, the visco-plastic part is neglected. It might be argued in the FSW process that the strain rates in the deformation zone, where the pin is stirring the two workpiece materials, are very high, but if the thermal driven residual stress field is considered only, then the overall traversing speed of the tool, i.e. heat source, is slower compared to other welding applications. This assumption leads to ε_{ij}^{vp}=0 and Eq. (2.14) is reduced to Eq. (2.15) as,

$$\varepsilon_{ij}^{tot} = \varepsilon_{ij}^{el} + \varepsilon_{ij}^{pl} + \delta_{ij}\varepsilon^{th} \tag{2.15}$$

The total strain can be formulated as given in Eq. (2.16), for the small-strain theory assumption where the nonlinear terms are neglected,

$$\varepsilon_{ij}^{tot} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) \tag{2.16}$$

where u_i is the total displacement vector. The thermal strain is written as $\delta_{ij}\varepsilon^{th}$ since it is volumetric, i.e. it acts only in the x, y, z-axes, and it is computed in increments as in Eq. (2.17) (which is then summed leading to the total form),

$$\Delta\varepsilon^{th} = \int_{T_1}^{T_2} \alpha dT \Rightarrow \varepsilon^{th} = \sum_{i}^{n_{incr}} \Delta\varepsilon^{th} \tag{2.17}$$

where α is the thermal expansion coefficient and T_1, T_2 denote the bounding temperatures in the increment. The well-known generalized Hooke's law for thermo-elastic systems is applied to relate elastic strains to stresses using the relation given in Eq. (2.15) as follows,

$$\sigma_{ij} = L_{ijkl}^{el}\varepsilon_{kl}^{el} = L_{ijkl}^{el}\left(\varepsilon_{kl}^{tot} - \varepsilon_{kl}^{pl} - \delta_{kl}\varepsilon^{th}\right) \tag{2.18}$$

where C_{ijkl}^{el} is the elastic constitutive (stiffness) tensor. Eq. (2.18) can be rewritten by splitting the total strain into the mechanical part and the thermal part, i.e. $\varepsilon_{kl}^{tot} = \varepsilon_{kl}^{mech} + \varepsilon_{kl}^{th}$,

$$\begin{aligned}\sigma_{ij} &= C_{ijkl}^{el}\left(\varepsilon_{kl}^{mech} - \varepsilon_{kl}^{pl}\right), \\ &= \frac{E}{1+\upsilon}\left[\frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \frac{\upsilon}{1-2\upsilon}\delta_{ij}\delta_{kl}\right]\left(\varepsilon_{kl}^{mech} - \varepsilon_{kl}^{pl}\right)\end{aligned} \tag{2.19}$$

where υ is Poisson's ratio and E is the Young's modulus in the case of isotropic material.

The plastic strain is based on the standard J_2-flow theory with a temperature dependent yield surface, i.e. f$(\sigma_e, \varepsilon_e^{pl}, T)$ in Eq. (2.20), seperating the stress space into the elastic region inside ($f < 0$), and the plastic region on the surface ($f = 0$),

$$f(\sigma_e, \varepsilon_e^{pl}, T) = \sigma_e(s_{ij}) - \sigma_y(\varepsilon_e^{pl}, T) \tag{2.20}$$

where ε_e^{pl} is the equivalent plastic strain, σ_y is the yield stress and σ_e is the effective von Mises stress which only considers the deviatoric part of the stress tensor (s_{ij}) hence the hydrostatic part does not contribute to the plastic deformation (i.e. $s_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}$). The plastic deformation outside this yield surface ($f > 0$) is not physically possible without imposing the evolution of the yield surface (Fig. 2.18) and the *consistency condition* given in Eq. (2.22).

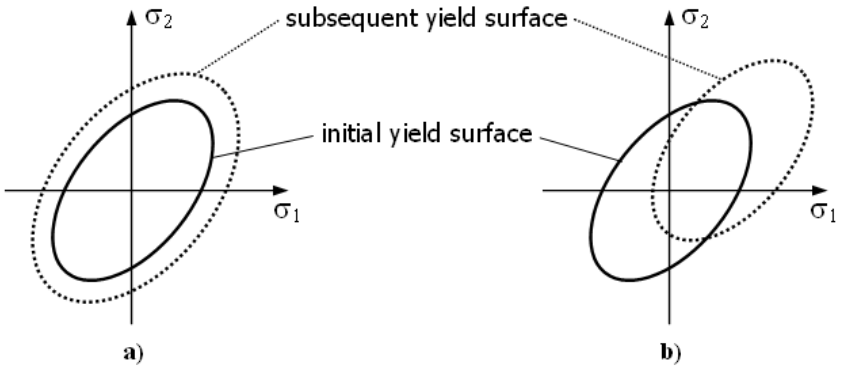


Figure 2.18: a) Isotropic hardening. b) Kinematic hardening.

The *Associated flow rule* determines how the plastic strain develops by using the yield criterion function (yield surface) as a plastic dissipation potential and this relation is formulated as,

$$d\varepsilon_{ij}^{pl} = d\lambda \frac{\partial f}{\partial \sigma_{ij}} \tag{2.21}$$

where $d\lambda$ is a scalar, the so-called plastic multiplier which stands for the size of the plastic strain increment whereas the $\frac{\partial f}{\partial \sigma_{ij}}$ term gives the direction of this increment, i.e. normal to the yield surface in isotropic hardening, as seen in Fig. 2.19.

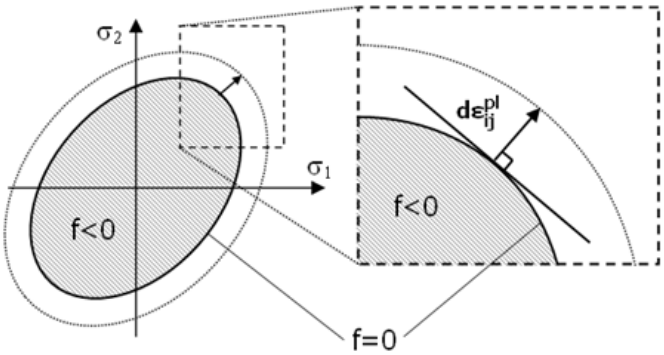


Figure 2.19: Plastic strain increment normal to the yield surface.

The plastic multiplier is determined from the *consistency condition*,

$$\begin{aligned} &\dot{f}(\sigma_e, \varepsilon_e^{pl}, T) = 0, \\ &\dot{f}(\sigma_e, \varepsilon_e^{pl}, T) = \frac{\partial f}{\partial s_{ij}} \dot{s_{ij}} + \frac{\partial f}{\partial \varepsilon_e^{pl}} \dot{\varepsilon}_e^{pl} + \frac{\partial f}{\partial T} \dot{T} = 0 \end{aligned} \tag{2.22}$$

which, alongside with satisfying Eq. (2.20), implies that the stress state must stay on the yield surface during plastic deformation.

For the implementation of the aforementioned plasticity models, there exist numerous stress-update algorithms and techniques such as operator splitting, the standard predictor, predictor-corrector, sub-incrementation, pure incremental or forward Euler scheme,

generalised trapezoidal or mid-point algorithms, a backward Euler (return-map), etc. and further details can be found in Simo and Hughes (1998); Belytschko et al. (2000); Crisfield (1991); Cook et al. (2001) (for computer implementation issues in detail), Lubliner (1990); Khan and Huang (1995); Tvergaard (2001); Boley and Weiner (1960) (for more theoretical oriented issues), and more specificly, thermo-mechanical modeling of welding processes in Lindgren (2007); Radaj (1992) and Feng (2005). Moreover, a detailed study regarding the implementation of physically based models for plastic deformation coupled to microstructure evolution, i.e. dislocation processes, and implementation of return-map algorithm can be found in (Domkin, 2005).

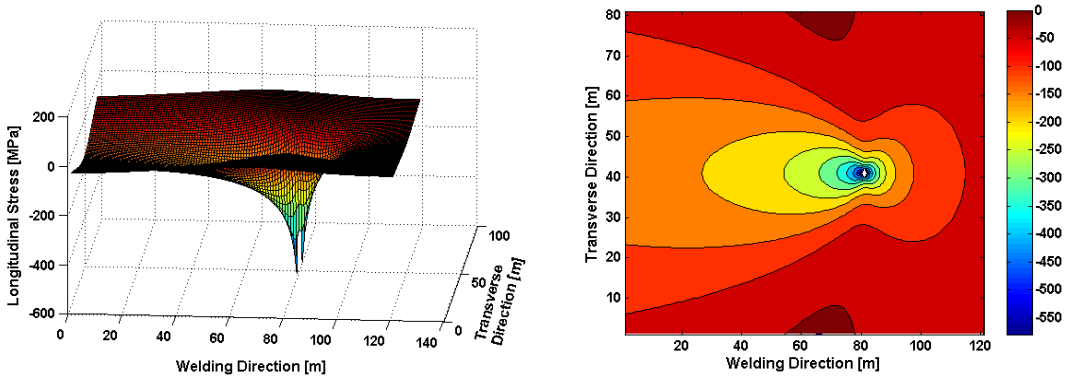
2.2.2 Analytical thermo-mechanical model

The thermal stress field during and after welding is as earlier discussed to a large extent non-linear and inelastic. In continuum mechanics, however, the solution for a nonlinear inelastic field problem is generally obtained by proceeding from a linearized elastic field problem. Fundamental questions relating to the stress field can be clarified with the linear-elastic solution. During welding, the stress field obtained from a linear elastic solution might resemble the real case to some extent. Morover, this elastic solution may be the starting point for further solution steps in the non-linear range of an elastic-plastic analysis. However, after cooling, it should be kept in mind that stresses become zero in this case, which of course is totally inapplicable to welding residual stresses.

The closed form solution for the quasi-static elastic thermal stress field, i.e. longitudinal stress, σ_x, transverse stress, σ_y, and shear stress, τ_{xy}, of a line-type heat source moving with a constant speed along a line in an infinitely large plate is given in Eq. 2.23 (Radaj, 1992),

$$\begin{aligned}
\sigma_x &= -\frac{\alpha E q}{4\pi k d}\left(e^{\frac{-vx}{2a}}\left[K_0(\frac{vr}{2a}) - \frac{x}{r}K_1(\frac{vr}{2a})\right] + \frac{2a}{v}\frac{x}{r^2}\right) \\
\sigma_y &= -\frac{\alpha E q}{4\pi k d}\left(e^{\frac{-vx}{2a}}\left[K_0(\frac{vr}{2a}) + \frac{x}{r}K_1(\frac{vr}{2a})\right] - \frac{2a}{v}\frac{x}{r^2}\right) \\
\tau_{xy} &= \frac{\alpha E q}{4\pi k d}\left(e^{\frac{-vx}{2a}}\frac{y}{r}K_1(\frac{vr}{2a}) - \frac{2a}{v}\frac{y}{r^2}\right)
\end{aligned} \tag{2.23}$$

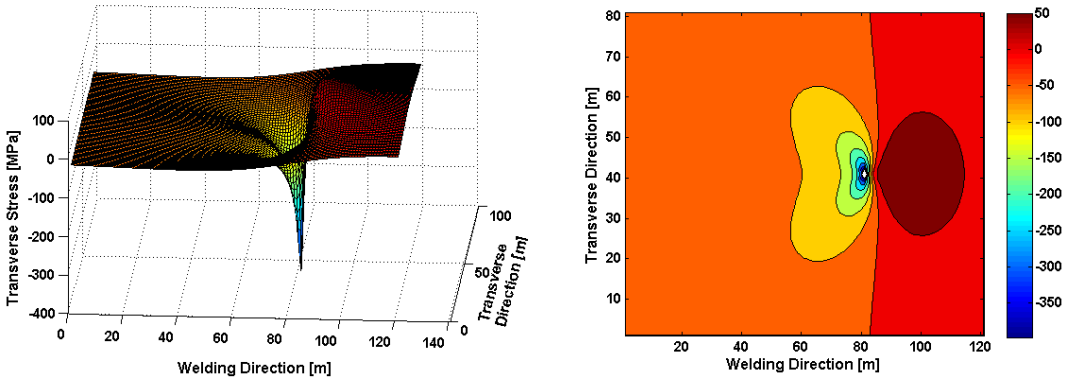
where K_0 and K_1 are Modified Bessel Functions of the second kind and zero and first orders, respectively, E is the Young's modulus and α is the thermal expansion coefficient. This solution is obtained from solving the 2D thermo-elastic problem based on the temperature field given by Eq. (2.6).



(a) (b)

Figure 2.20: (a) Longitudinal stress field on a 3-mm thick plate (Tutum and Hattel, 2008). (b) Contour plot of the longitudinal stress field (Tutum and Hattel, 2008).

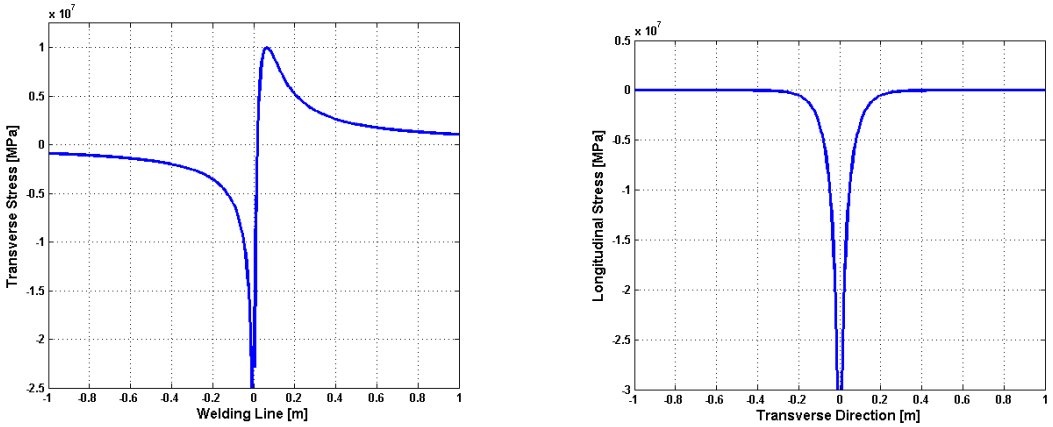
The longitudinal stress field shown in Figs. 2.20a and 2.20b resembles the reversed thermal field of the 2-D Rosenthal solution, Eq. (2.6), presented in Section 2.1.2.1, in Figs. 2.3a and 2.3b, respectively. There is again a singularity at the location of the heat source due to K_0 and K_1. Moreover, since only the steady-state condition is considered in this analytical model, compressive stresses , i.e. negative stress values shown in the contour scale, in front of the heat source are pronounced, but the tensile stresses evolving after accumulation of plastic strains are not captured, which as earlier mentioned is the main missing link to the more realistic thermo-elasto-plastic numerical models.



(a) (b)

Figure 2.21: (a) Transverse stress field on a 3-mm thick plate (Tutum and Hattel, 2008). (b) Contour plot of the transverse stress field (Tutum and Hattel, 2008).

Notice from Fig. 2.22b that the longitudinal stress profile, σ_x, is symmetric accross the joint line which can be predicted beforehand due to the symmetric thermal field, but also it is not in equilibrium (if finite dimensions are considered) in the transverse direction of the infinite large plate. The transverse stress profile along the joint line, as shown in Fig. 2.22a, is asymmetric due to a similar discussion as for the longitudinal stresses, but on the other hand, the equilibrium is more visible.



(a) (b)

Figure 2.22: (a) Transverse stress profile along the joint line (Tutum and Hattel, 2008). (b) Longitudinal stress profile along the transverse direction (Tutum and Hattel, 2008).

2.2.3 Numerical thermo-mechanical models

As presented in the previous section, analytical models in the field of thermo-mechanics, those which are particularly used in welding applications, are limited primarily due to the linearity in the material description, i.e. no temperature dependency, no hardening or softening behavior, etc. and no time dependency, i.e. exclusion of transient effects. If the residual stresses and distortions in the welded structure are the primary concern, numerical models, which are capable of dealing with the problems mentioned previously, should be used.

In the next two subsections, prescribed and TPM heat source based numerical thermo-mechanical models, in which transient thermal models are sequentially (weakly) coupled to the quasi-static mechanical models, are presented. These models constitute the basis of `PAPER-II` (Tutum et al., 2008a) and `PAPER-III` (Tutum et al., 2009) attached at the end of this Ph.D. thesis.

2.2.3.1 Prescribed heat source-based residual stress models

The model presented here represents the welding of two flat plates resembling the bead on plate. As earlier with bead on plate, due to symmetry only one of the plates is modelled. The dimension of the workpiece is 300 mm x 150 mm x 3 mm and the material properties, which are given in Table 2.4 and Fig. 2.23, do not correspond to any specific commercial Al-alloy but should be understood as a benchmark material. This assumption of a benchmark material is in agreement with the study presented by Zhu and Chao (2002), which concludes that the temperature dependent yield stress has a significant effect on the residual stress and distortion, and except for this yield stress, using material properties at room temperature gives reasonable predictions of the transient temperature fields, the residual stresses and distortions. The welding speed is 2 mm/sec and the diameter of the tool shoulder is 20 mm. The moving heat source starts and stops at 50 mm from

Table 2.2: Summary of the thermal model boundary conditions for Case-A, B and C.

Load Steps	Step-I (Welding)	Step-II (Cooling)	Step-III (Releasing)
Case-A, B, C	Moving surface heat flux Adiabatic B.C.s	No heat source Adiabatic B.C.s	No heat source Adiabatic B.C.s

Table 2.3: Summary of the mechanical model boundary conditions for Case-A, B and C. Note that the rollers are applied at the bottom surface of the plate.

Load Steps	Step-I (Welding)	Step-II (Cooling)	Step-III (Releasing)
Case-A	Clamp both ends, Apply the rollers	Keep the clamps, Keep the rollers	Clamp the corner, Release the clamps and rollers
Case-B	Clamp the side, Apply the rollers	Keep the clamp, Keep the rollers	Clamp the corner, Release the clamp and rollers
Case-C	Clamp the corner, Apply the rollers	Keep the clamp, Keep the rollers	Keep the clamp, Release the rollers

the left and the right edges of the plate, respectively, so the welding step is defined for 100 seconds. Different clamping conditions, corresponding to the mechanical boundary conditions defined in the following subsections from Case-A to Case-C, are tested in order to visualize the effect of them on transient and residual stress states. Both thermal and mechanical boundary conditions for these three cases are summarized in Tables 2.2 and 2.3, for convenience.

The model is composed of two regions, i.e. the weld region and the outer region, see Fig. 2.24a, that are enmeshed using a structural scheme; in the weld region: elements of 2 mm x 2 mm x 1.5 mm (corresponding to 150 x 10 x 2 elements along the edges), in the outer region: elements of 2 mm x 3.25 mm x 1.5 mm (corresponding to 150 x 40 x 2 elements), i.e. a total of 18,000 C3D8T linear coupled temperature-displacement elements in ABAQUS, see Fig. 2.24b.

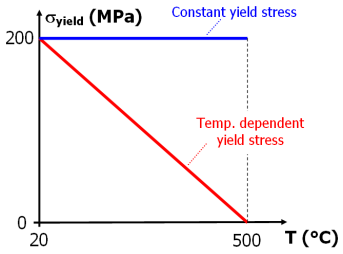


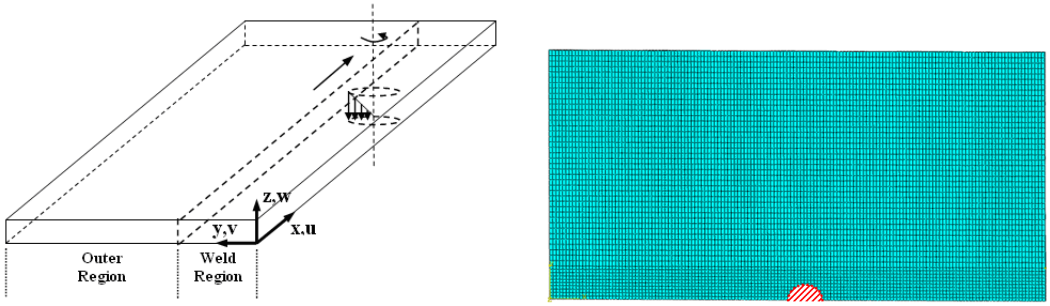
Figure 2.23: Constant and temperature dependent yield stress curves.

In order to simulate the moving heat source during the welding step, a position

Table 2.4: Temperature Independent Material Properties of Benchmark Material [ref. PAPER-II] (Tutum et al., 2008a)

Heat Conduction, k	$[W/(m\,K)]$	160
Heat Capacity, c_p	$[J/(kg\,K)]$	900
Young's Modulus, E	$[GPa]$	70
Tangent Modulus, E_t	$[GPa]$	7
Thermal Expansion Coefficient, α	$[1/K]$	$2.3\text{x}10^{-5}$
Density, ρ	$[kg/m^3]$	2700

and time dependent distributed surface heat flux is applied on element faces using the *DFLUX user subroutine in ABAQUS. A total heat power of 600 W and a radius of the tool shoulder of 10 mm are used for the corresponding constants in Eq. (2.9), Section 2.1.3. The heat source is turned off during cooling and releasing. The heat transfer from the plate to the environment, which plays an important role in usual applications, is neglected (corresponding to adiabatic boundary conditions) in order to simplify the benchmark study and focus on the effect of the mechanical boundary conditions. Since all the thermal boundary conditions are the same in the three cases of the benchmark model, only the mechanical boundary conditions are described in detail in the following three sections.



(a) (b)

Figure 2.24: (a) Schematic view of Model-1 [ref. PAPER-II] (Tutum et al., 2008a). (b) Enmeshment of Model-1 [ref. PAPER-II] (Tutum et al., 2008a).

2.2.3.1.1 Case-A: Case-A considers clamping of both ends of the plate perpendicular to the symmetry line. Rollers, which resemble the vertical displacement restriction due to the backing plate, are applied at the bottom of the plate. These clamping and roller constraints are kept constant also during the cooling step and the releasing step is applied as defined in the previous section.

At the end of the cooling step, the temperature field is stabilized at 107.2 °C because of the adiabatic boundary conditions. Longitudinal and transverse stresses, which are noted at the middle of the plate on a path defined between the points of (0.15, 0.0, 0.0015) and (0.15, 0.15, 0.0015), are shown in Fig. 2.26. The workpiece material, that has already

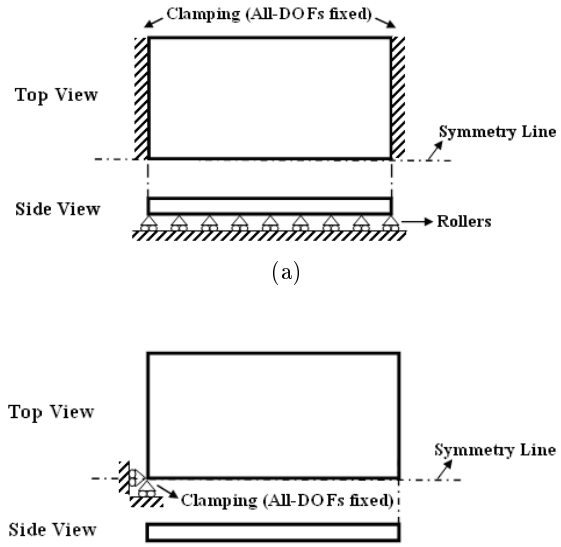


Figure 2.25: (a) Step-1 and 2 for Case-A [ref. PAPER-II] (Tutum et al., 2008a). (b) The mechanical boundary conditions at Step-3 of each case for Model-1 [ref. PAPER-II] (Tutum et al., 2008a).

yielded in compression, transforms into tensile stresses by the decrease of thermal strains during the cooling period and approximately 60 MPa is obtained on the center line while 175 MPa compressive stress is observed at approximately 60 mm away from the center line in the transverse direction. Following the cooling step, the clamps and rollers are released at both ends and the bottom surface of the plate, respectively, and the stresses reestablish as shown in Fig. 2.26, so that they shift approximately 125 MPa all over in the positive axis. Transverse stresses are quite low compared to longitudinal stresses as expected due to the geometry of the plate, welding direction and mechanical boundary conditions.

Both longitudinal and transverse stresses get higher when the temperature dependent yield stress, see Fig. 2.23, is taken into account as seen in Fig. 2.27. The maximum tensile stress obtained in the center line shifts from 200 MPa to 275 MPa, due to much higher plastic strain observed, at the end of releasing step. The increase in the maximum value of the transverse stresses along the transverse direction of the plate compared to those observed in Fig. 2.26 is approximately 25 MPa. The maximum compressive longitudinal stress obtained close to the middle of the plate is around 60 MPa at the end of the releasing step. It can be seen by comparing Figs. 2.26 and 2.27 that the general characteristics of both the longitudinal and transverse stress are preserved while changing in magnitude, as expected.

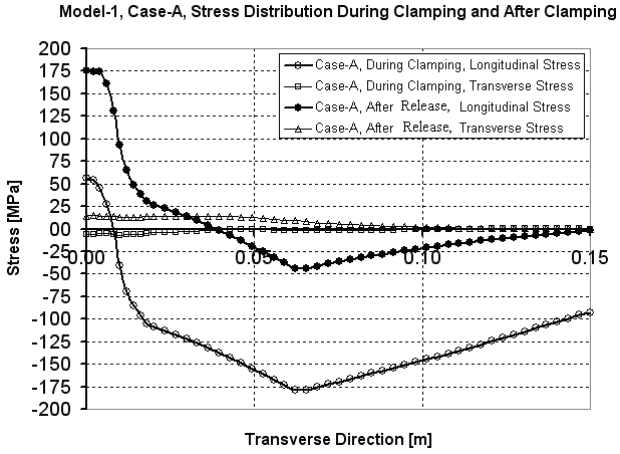


Figure 2.26: Stress distribution along the transverse direction of the plate for Case-A for the temperature independent yield stress [ref. PAPER-II] (Tutum et al., 2008a).

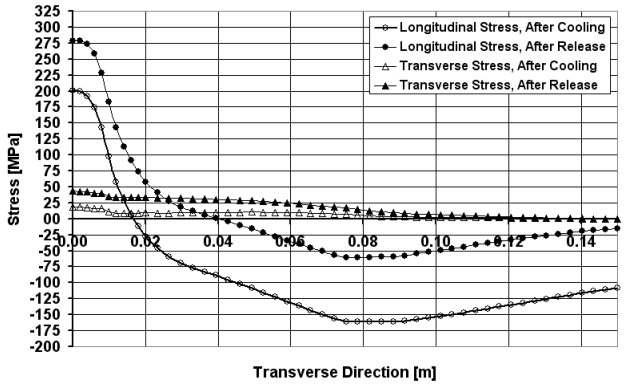


Figure 2.27: Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-A [ref. PAPER-II] (Tutum et al., 2008a).

2.2.3.1.2 Case-B: Case-B considers the type of clamping mostly used in industrial applications, that the side edge of the plate parallel to the symmetry line is fixed. Rollers are applied at the bottom of the plate.

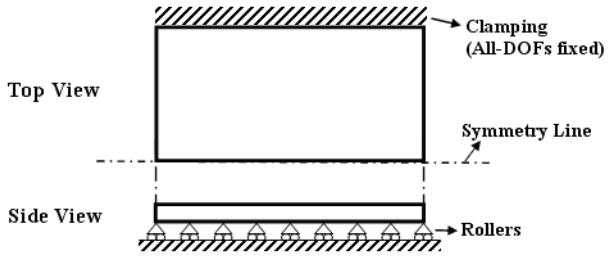


Figure 2.28: Step-1 and 2 for Case-B [ref. PAPER-II] (Tutum et al., 2008a).

The same temperature field is obtained through all three cases, as should be, due to adiabatic boundaries using the same thermal model parameters, i.e. total heat input,

thermal expansion coefficient, adiabatic thermal boundary conditions, etc. Longitudinal and transverse stresses are shown in Fig. 2.29 and graphed using the same path defined in the previous section. In this case, the side of the plate is clamped and the stress evolution directly captures this resistance by building a localized compressive zone close to the transverse edge of the plate during the welding step. The maximum compressive longitudinal stresses obtained are approximately 170 MPa, while 135 MPa is observed also for the maximum transverse stresses. In the subsequent releasing step, equilibrium in the longitudinal stresses is observed and the maximum tensile stress reaches approximately 125 MPa while almost negligible values are obtained for the transverse stresses, as expected.

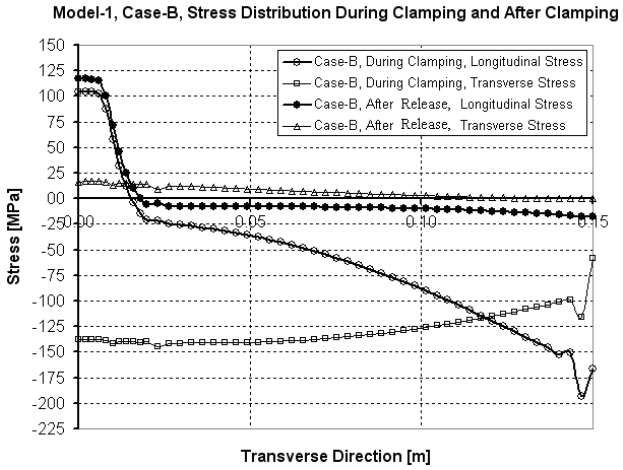


Figure 2.29: Stress distribution along the transverse direction of the plate for Case-B for the temperature independent yield stress [ref. PAPER-II] (Tutum et al., 2008a).

The temperature dependence of the yield stress results in higher longitudinal stress due to higher plastic strain, as mentioned in the previous section. Fig. 2.30 shows the resulting longitudinal and transverse stresses along the transverse direction for Case-B. The maximum longitudinal and transverse stresses are 180 MPa and 20 MPa, respectively, on the centre line at the end of the releasing step.

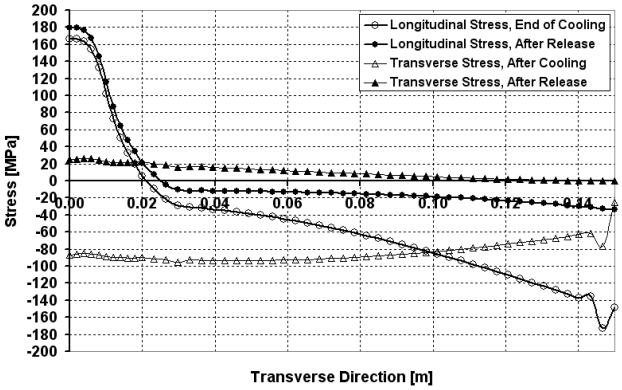


Figure 2.30: Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-B [ref. PAPER-II] (Tutum et al., 2008a).

2.2.3.1.3 Case-C: Case-C considers a theoretical case with the boundary conditions, shown in Fig. 2.31, mainly applied to account for the investigation of the effect of the mechanical constraints, which corresponds to bead on plate with clamping preventing out of plane deformations.

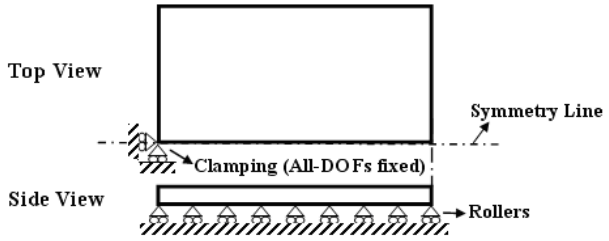


Figure 2.31: Step-1 and 2 for Case-C [ref. PAPER-II] (Tutum et al., 2008a).

This is the simplest case out of the three cases due to the simplicity of the mechanical boundary conditions. After both the cooling and releasing steps, the stress evolution almost does not change since the force equilibrium remains almost constant in these two steps.

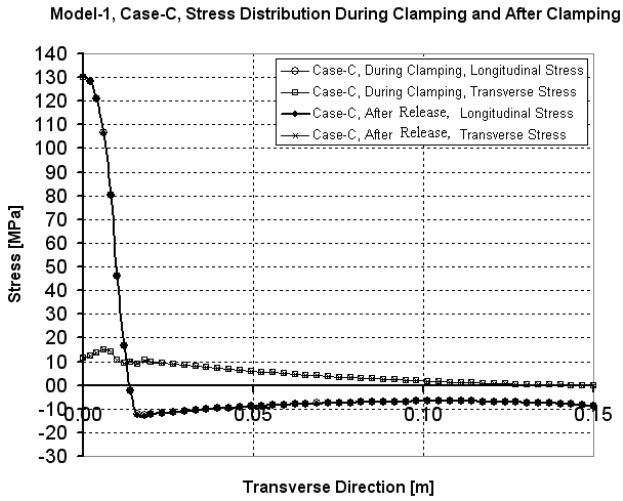


Figure 2.32: Stress distribution along the transverse direction of the plate for Case-C for the temperature independent yield stress [ref. PAPER-II] (Tutum et al., 2008a).

Fig. 2.33 shows the stress distribution along the transverse direction of the plate regarding the temperature dependent yield stress model of Case-C excluding the results obtained after the cooling step since they are the same as the ones obtained after the releasing step. The maximum longitudinal tensile stress observed in the center line is approximately 230 MPa which is 100 MPa higher than the results for the constant yield stress shown in Fig. 2.32.

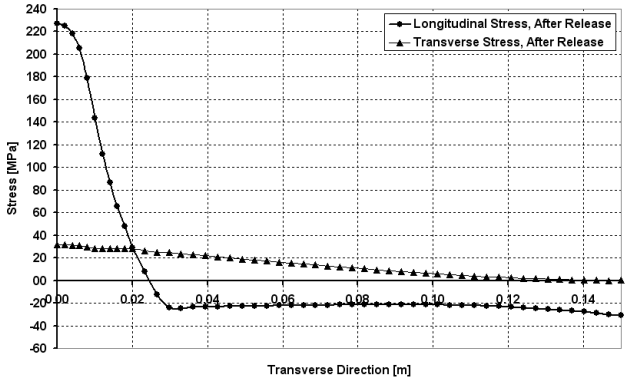


Figure 2.33: Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-C [ref. PAPER-II] (Tutum et al., 2008a).

The three cases shown here were investigated in order to analyze some principal effects of mechanical boundary conditions on residual stresses of FSW. These results were presented at the 7^{th} *International FSW Symposium*, Japan, and published in the conference proceedings [ref. PAPER-II] (Tutum et al., 2008a).

2.2.3.2 TPM heat source-based residual stress models

The de-coupled thermo-mechanical model of the FSW process used in this section consists of a transient TPM heat source based-thermal model, as presented in detail in Section

2.1.4, and a quasi-static elasto-plastic mechanical model, which is implemented by utilizing the commercial finite element software ANSYS. This thermo-mechanical model presented here is used to predict residual stresses in the plate to be further investigated for the optimization study given in PAPER-III (Tutum et al., 2009).

As several times before, the model represents the welding of two flat plates by considering the bead on plate, and hence only one of the plates is modeled. The dimension of the workpiece is 300 mm x 100 mm x 3 mm. This means that the thermally induced out-of-plane stresses will be negligible and a plane-stress analysis is reasonable. As before, the temperature independent material properties, which are given in Table 2.4, do not correspond to any specific commercial aluminum alloy but should be understood as a representative aluminum benchmark material. The temperature dependence of the yield stress for both thermal and mechanical analyses is shown in Fig. 2.34 as a linear function of temperature with a negative slope, decreasing from 200 MPa to 0 MPa at 20 °C and 500 °C, respectively. There is an exception for the mechanical analysis that 475 °C is chosen to be the cut-off temperature and the yield stress is kept constant at 10 MPa above this temperature. This engineering simplification provides a substantial convenience for controlling computational cost as it reduces the nonlinearities that do not have a significant effect on the global behavior of the thermo-mechanical model (Zhu and Chao, 2002). This material description is considered to be adequate for the FSW simulation where the first order effect is the yield strength of the material as a function of temperature.

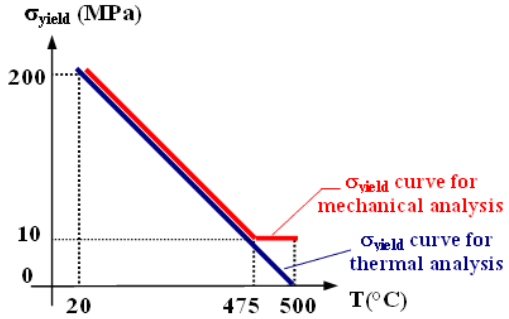
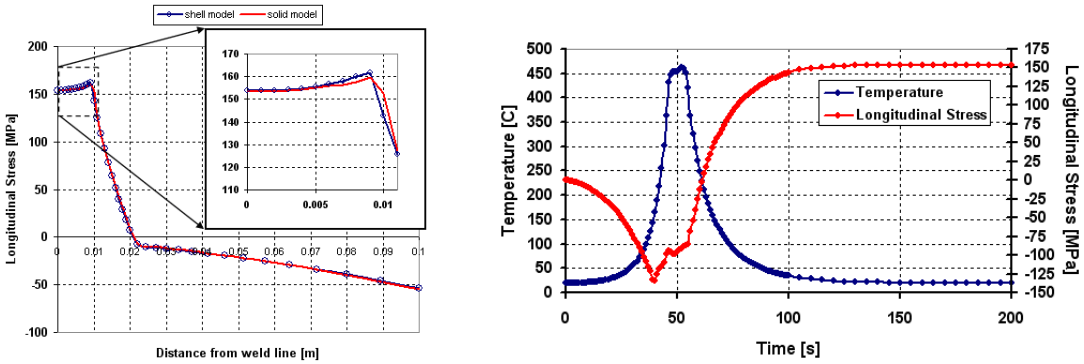


Figure 2.34: Temperature dependent yield stress curves considered for the thermal (red) and mechanical (blue) analyses, respectively [ref. PAPER-III] (Tutum et al., 2009).

Due to the very low contribution to the heat generation coming from the tool pin, only the tool shoulder is considered in the heat source. The diameter of the tool shoulder is 20 mm. The mechanical effects of the tool are not included, and thus residual stresses are assumed to be primarily a function of the thermal load history (Chao and Qi, 1998; Peel et al., 2003). The moving heat source starts and stops at 50 mm away from the left and the right edges of the plate, respectively. The SHELL 131, 4-Node layered thermal shell element is used for the transient thermal analysis while the PLANE 182, 2-D 4-node structural solid element is used for the quasi-static mechanical analysis and the same structured finite element mesh is used in both cases.

The accuracy of the mechanical simulation using shell and plane stress models has been compared with a three-dimensional solid linear 8-node element model. Here, SOLID 45 elements are used for the quasi-static mechanical analysis. Referring to Fig. 2.35a,

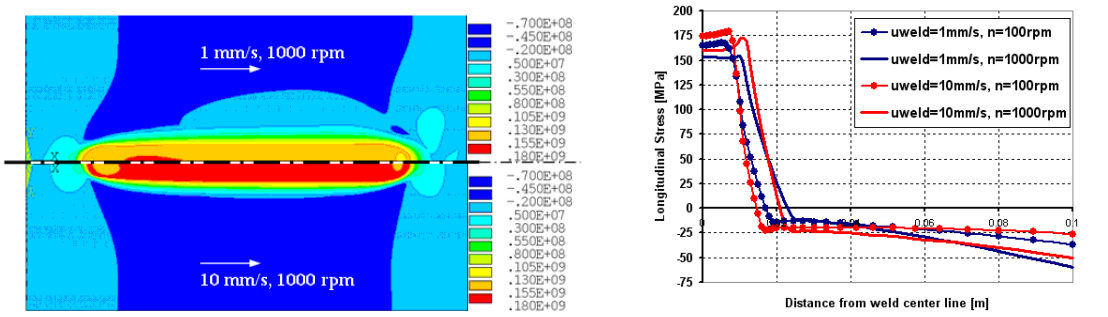
the longitudinal stress profiles obtained with the two models are in good agreement, e.g. the maximum difference is 3 MPa. Thus, it can be concluded that for the present case the sequentially coupled shell and plane-stress models can be used for the optimization studies while being both accurate and computational efficient.



(a) (b)

Figure 2.35: (a) Comparison of longitudinal stress profiles obtained by shell (red) and solid (blue) models [ref. PAPER-III] (Tutum et al., 2009). (b) Transient longitudinal stress and temperature profiles as a function of time in the weld line, at the middle of the plate [ref. PAPER-III] (Tutum et al., 2009).

Fig. 2.35b shows the transient thermal profile along the weld line together with the transient longitudinal stress; this is the main contribution to the residual stresses evolved locally along the centre line, i.e. in the tension zone.



(a) (b)

Figure 2.36: (a) Contour plot of the longitudinal stress field with increments of 22 MPa [ref. PAPER-III] (Tutum et al., 2009). (b) Residual normal stress in longitudinal direction as a function of distance from weld line [ref. PAPER-III] (Tutum et al., 2009).

Fig. 2.36a shows the contour plots of the resulting longitudinal stress field of the symmetric models (isotherms in 25 MPa) for a chosen welding speed of 1 mm/s and 10 mm/s, respectively, for a rotational speed of 1000 rpm. A parameter study regarding longitudinal stress profiles along the transverse direction and at the middle of the plate with two welding speeds, i.e. 1 and 10 mm/s, and two rotational speeds, i.e. 100 and

1000 rpm, is presented in Fig. 2.36b. These results clearly show the main characteristics of the model of the residual stresses in the FSW process:

- Increasing the welding speed for a fixed rotational speed results - in general - in higher stress levels in the tension zone.
- Increasing the rotational speed for a fixed welding speed yields lower peak residual stress.
- The gradients in residual stress profiles along the transverse direction of the plate become higher with increasing welding speed, while the model results in wider residual stress profiles for lower welding speeds.

In this entire Section 2.2.3, a presentation of the numerical thermo-mechanical models was given. Although giving interesting results, common for all the models was that no welding defects were considered and moreover the yield stress was a function of the current temperature only, and not thermal history. This will be briefly considered in the next section.

2.2.4 Mechanical Properties of FS welds

As mentioned above, it is well known that the thermal conditions during FSW highly affect the probability of getting welding defects as well as the resulting mechanical properties of the final weld. In Fig. 2.37 the effect of a hot condition (a), a stable condition (b) as well as a cold condition (c) in FSW is presented. In the hot condition overstirring is observed in which the material under the shoulder rotates more than in the lower weld region leading to potential welding defects. In the intermediate case a stable shear layer reaching from the shoulder down to the bottom of the pin is observed leading to a defect free weld. In the cold weld condition the temperature is simply too low to facilitate a proper flow of the material. In this case a worm-hole or tunnel-like defect is very likely to occur.

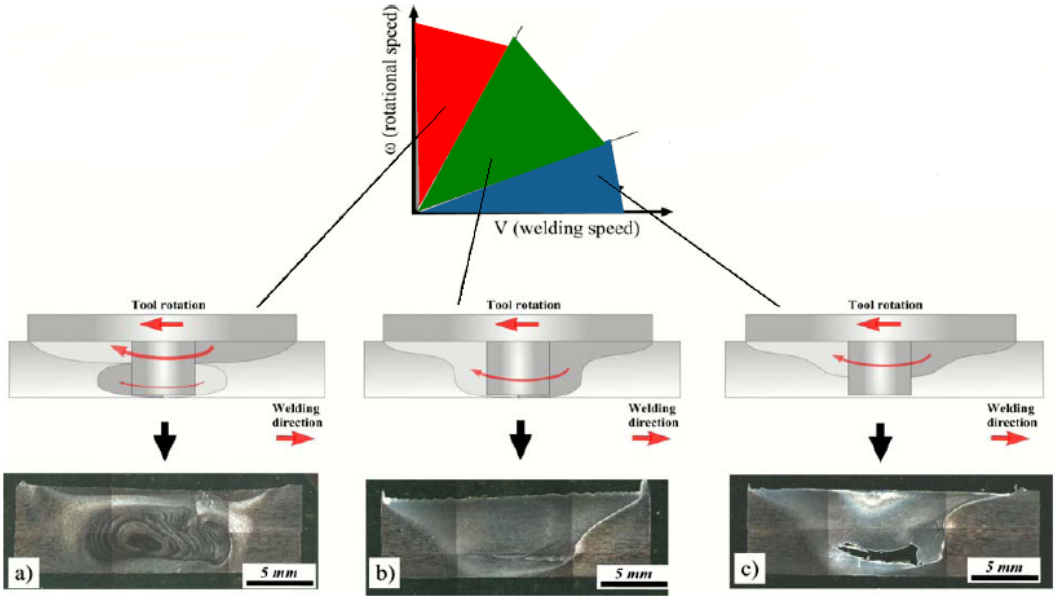


Figure 2.37: (a) Red denotes hot condition (overstirring), low k. (b) Green denotes stable condition, intermediate k. (c) Blue denotes cold condition, high k (k: advancement per revolution (APR) $\Rightarrow k = u_{weld}/n(rpm)$) (Jahazi et al., 2008)

2.2.4.1 Softening model

It is well-known that some heat treated metals loose hardness and strength in the HAZ when subjected to thermal cycles during welding. In particular this is the case for high strength heat treated aluminum alloys, used in e.g aerospace applications. The softening of the precipitate-strengthened aluminum by dissolution has thus been studied previously, particularly in the 6082-T6 aluminum alloy which has high strength due to a fine distribution of $\beta''(Mg_2Si)$ precipitates in the aluminum matrix, by developing a dissolution softening model (Myhr and Grong, 1991a,b). The basis of this model is experiments in which samples are put into an oven and kept there for a specified period of time at a specified temperature. Following this, the samples are tested mechanically for hardness, HV. Based on this results curves for hardness vs. thermal history are constructed.

In the softening model presented here, the fraction of the hardening precipitates dissolved, X_d, is related to the equivalent time of heat treatment, t_{eq}, where $t_{eq} = t/t^*$. Here, t is the time of heat treatment at a temperature T, and t^* is the time for total precipitate dissolution at this temperature, T. The equivalent time of heat treatment can be calculated as in Eq. (2.24) (Myhr and Grong, 1991a),

$$t_{eq} = \sum_{i=1}^{t_{total}} \frac{t}{t^*} = \sum_{i=1}^{t_{total}} \frac{t}{t_{ref} \exp\left[\frac{Q_{eff}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right]} \tag{2.24}$$

where t_{ref} is the time for full dissolution at a reference temperature (T_{ref}), R is the gas constant and Q_{eff} is the effective activation energy for precipitate dissolution. This is achieved by discretising the thermal history curve (as shown in Fig. 2.38) and calculating t_{eq} for each segment, followed by summing these terms to obtain a total t_{eq} value for the

curve, representing its relative kinetic strength with respect to the dissolution of hardening precipitates. The fraction of hardening precipitates, $\frac{f}{f_0}$, is related to the equivalent time of heat treatment by,

$$\frac{f}{f_0} = 1 - X_d = 1 - (t_{eq})^n = 1 - \sqrt{t_{eq}} \tag{2.25}$$

where the exponent, n, is a material constant obtained experimentally, e.g. n=0.5 in this case as given in Table 2.5 from Myhr and Grong (1991a). Finally, the hardness distribution is predicted via linear interpolation between the original state and the fully dissolved state,

$$HV = (HV_{max} - HV_{min}) \frac{f}{f_0} + HV_{min} \tag{2.26}$$

where HV_{max} is the hardness of the material in T6 condition and HV_{min} is the hardness of the fully softened (original state) material. This procedure is schematically shown in Fig. 2.38.

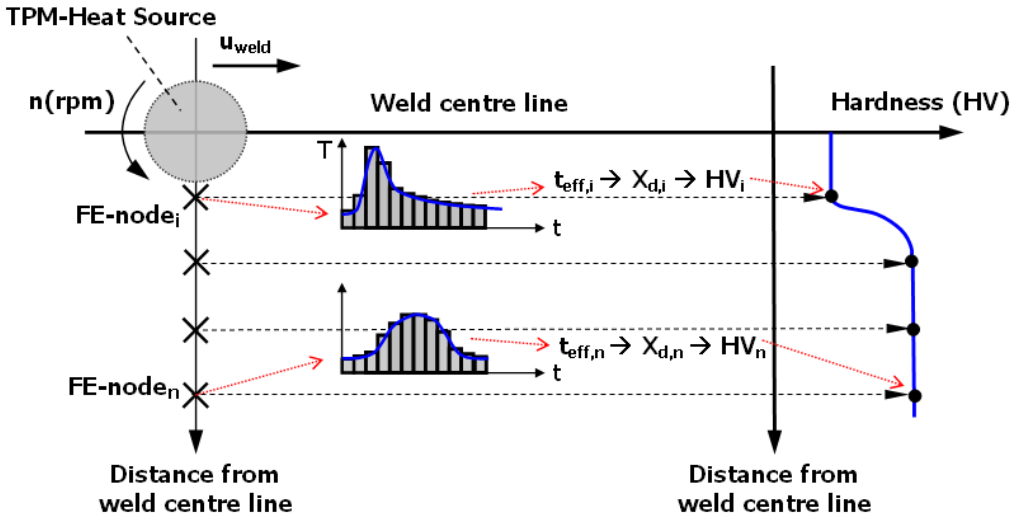


Figure 2.38: Schematic view of the weld hardness profile calculated in a sequential process which is initiated by the TPM-heat source model and followed by the softening model (adapted from Russell and Shercliff (1999))

A 3D thermal finite element model implemented in ANSYS is used to obtain the thermal field which is then used as an input for the softening model to predict the hardness distribution in the plate (300 x 125 x 5 mm^3). The TPM heat source model is applied together with proper boundary conditions, i.e. moving surface heat flux at the top surface of the plate, heat loss through the bottom (h_{bottom}=500 W/(m^{2}K)) and top (h_{top}=10 W/(m^{2}K)) surfaces of the plate. The material constants are shown in Table 2.5, which are previously applied by Myhr and Grong (1991a). The softening model, formulated as in Eqs. (2.24) through (2.26), is also implemented in ANSYS using scripting features.

Table 2.5: Constants used for the softening model (Myhr and Grong, 1991a)

t_{ref}	600 s
Q_{eff}	130 kJ/mol
R	8.314 J/(mol K)
T_{ref}	375 °C
n	0.5
HV_{max}	110
HV_{min}	42

Fig. 2.39 shows the hardness distributions in the cross-sections of the middle of the plates for two welding cases, i.e. a hot weld (the welding speed, u_{weld}=1 mm/s and the tool rotational speed, n=1800 rpm) and a cold weld (the welding speed, u_{weld}=10 mm/s and the tool rotational speed, n=600 rpm). Since the material flow due to the rotation of the tool shoulder is not taken into account, only one of the plates is considered. Hence, the distribution of both thermal and hardness profiles along the weld centre line are symmetric, as expected.

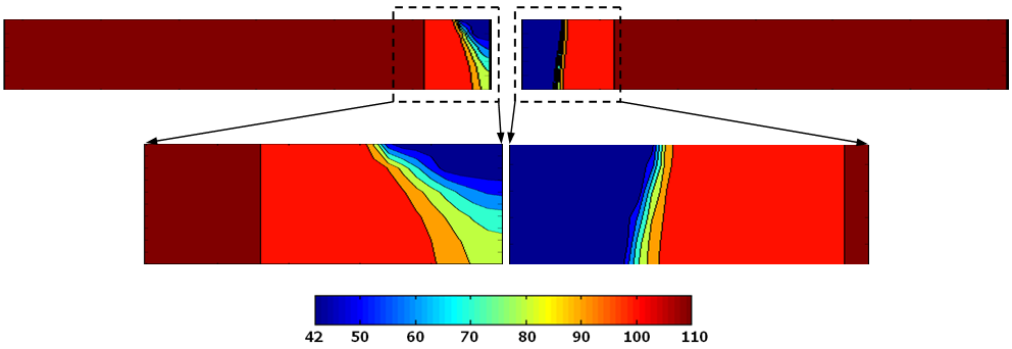


Figure 2.39: Hardness solution fields (cross-sectional view of the plate, i.e. 5 mm-thick and 125 mm-wide) for the cold weld, u_{weld}=10 mm/s and n=600 rpm (on the left), and the hot weld, u_{weld}=1 mm/s and n=1800 rpm (on the right)

As seen in both cases, the highest dissolution of the precipitates, in other words the lowest hardness distribution, is observed at the top surface of the plates since it is directly exposed to the heat generation from the tool shoulder. However, a significant difference in the slopes of the contour lines is visible. Furthermore, the width and the height of the fully softened regions noticeably differ in both cases, i.e. the cold weld results in a narrower HAZ compared to the hot weld in which the material under the tool shoulder is fully softened from the surface to the bottom of the plate.

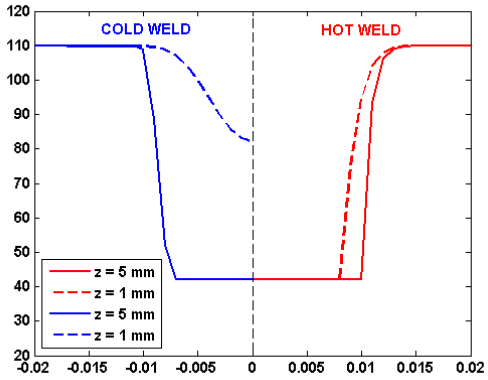


Figure 2.40: Hardness profiles at different heights of the plate for the cold weld, u_{weld}=10 mm/s and n=600 rpm (on the left), and the hot weld, u_{weld}=1 mm/s and n=1800 rpm (on the right)

Similarly, the hardness profiles constructed at different thicknesses, i.e. at 1 mm and 5 mm, are shown in Fig. 2.40 where the characteristics are in good agreement with the experimental results shown in Figs. 2.41 and 2.42 made by Jahazi et al. (2008), one should keep in mind that the material used for the experiments is different.

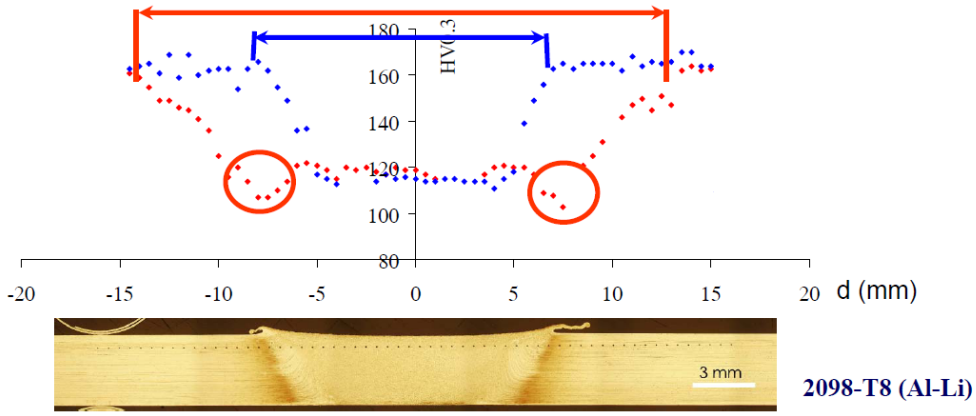


Figure 2.41: Hardness profiles for 2098-T8 (Al-Li) alloy. Blue denotes cold weld (u_{weld}=20 mm/s and n=600 rpm) and red denotes hot weld (u_{weld}=1 mm/s and n=1800 rpm) (Jahazi et al., 2008).

The thermal softening of the workpiece material also has implications for the maximum welding speed achiavable in the FSW process (Russell and Shercliff, 1999). It is desired to have a fully dissolved, i.e. fully softened, region in front of the tool pin in order to avoid pin fracture or minimize tool wear while traversing along the weld line. If the cold weld case shown in Fig. 2.39 is considered, the welding speed is most likely to be higher than a limiting (desirable) speed at which the softened zone just encloses the tool pin.

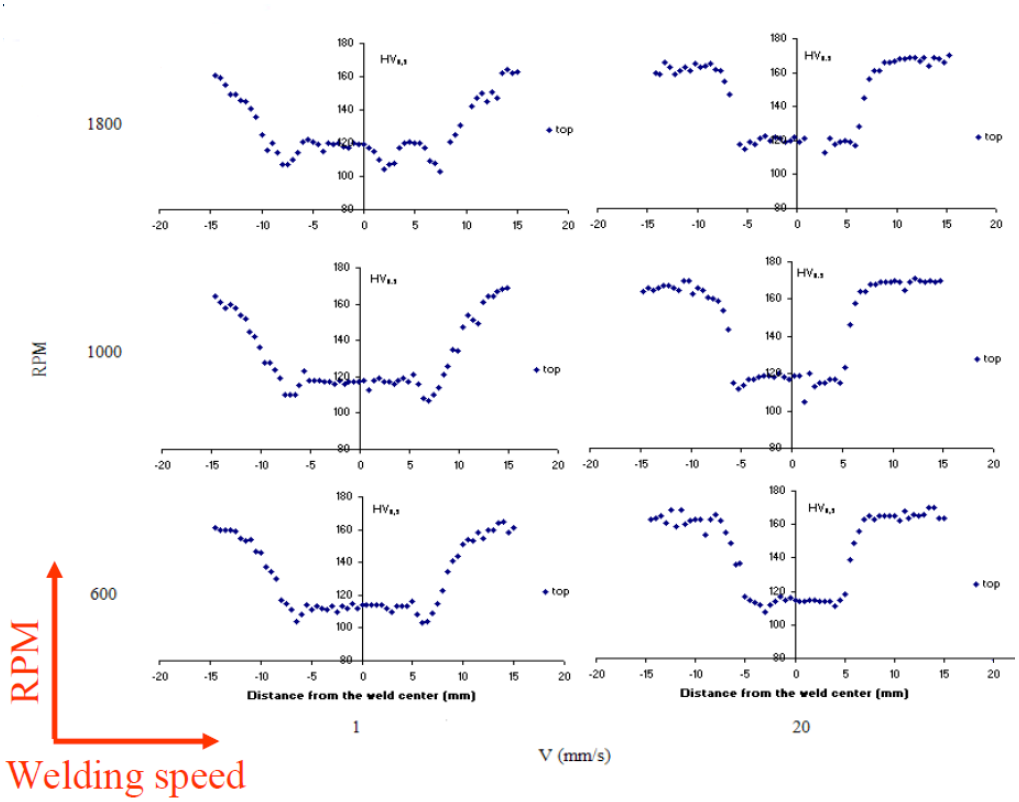


Figure 2.42: DOE studies of hardness measurements vs. FSW process parameters, i.e. welding speed and tool rotational speed, for 2098-T8 alloy (Jahazi et al., 2008).

At relatively high temperatures, i.e. close to the weld centre line, significant dissolution occurs, leading to a lot of alloying elements in solution which reprecipitate after cooling giving rise to some strength recovery. At lower temperatures, however, i.e. further out from the weld centre, less dissolution occurs, and the concentration of solute is too low for significant reprecipitation upon cooling. Furthermore, excess solute may tend to join onto existing precipitates, leading to coarsening and permanent strength loss (Russell and Shercliff, 1999).

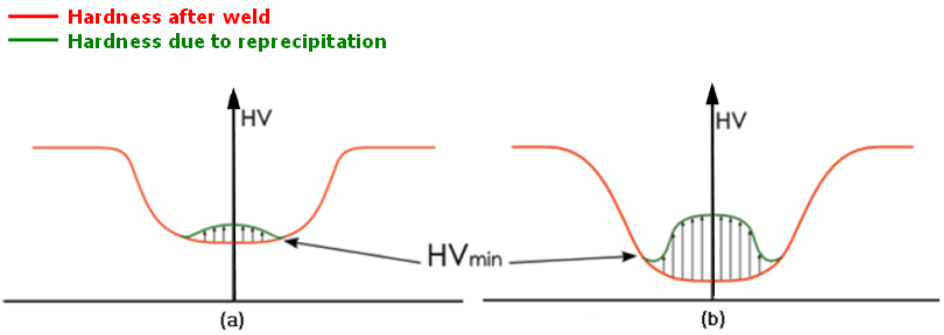


Figure 2.43: Principal hardness profiles (a) Cold weld, (b) Hot weld (Jahazi et al., 2008).

This behavior is schematically shown in Fig. 2.43 for the cold and hot welds, respectively, where the red curve represents the hardness profile observed after welding and the

green curve shows the hardness redistribution after reprecipitation, i.e. after natural ageing (Jahazi et al., 2008). However, the softening model presented here does not include this reprecipitation behavior, as well as the mechanical effects taking place close to the tool.

2.3 Integrated modeling

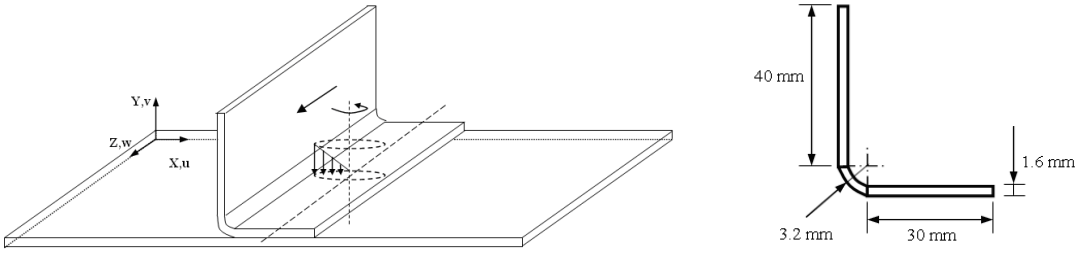
In the field of modelling of manufacturing engineering„ the integration of multiprocess steps and materials modelling has emerged as a new growing field, which attracts some attention in literature, although the amount of articles dealing specifically with the subject is still relatively limited. Crumbach et al. (2004) make a through process modelling of aluminium sheet production following the microstructural evolution during the production steps involving casting, heat treatment and forming processes. Also, Bellini et al. (2006) analyse the heat treatment of cast parts. Kermanpur et al. (2004) simulate the various stages of gas turbine disc manufacture to track defects throughout the entire process chain of melting, homogenisation heat treatment, cogging, forging, final heat treatment and machining. Gandin et al. (2002) make an integrated model of casting, solidification and heat treatment to predict the final yield stress of an Al-Cu cast alloy. Myhr et al. (2004) presents a numerical model for microstructure and strength evolution in Al-Mg-Si alloys during aging, welding and post-heat treatment and Lundbäck et al. (2005) simulate the sequence of TIG welding and post-weld heat treatment of an Inconel plate. Recently, examples of mapping the results to a subsequent load analysis during service have also emerged; for example, Robin et al. (2005) model phase distribution and residual stresses in spot welding and map the results to a subsequent crash calculation and Thorborg et al. (2006) model the process sequence of welding a Stellite 6 layer on a P91 valve followed by machining, heat treatment and in-service conditions in a power plant. For the heat treatment and in-service stages, the microstructural evolution due to diffusion in the Stellite/P91 is also taken into account (Hattel, 2008).

In the following, two examples of integrating the modelling of the FSW process together with the subsequent load analysis and damage evolution during in-service conditions will be presented.

2.3.1 Investigation of the service load performance of a friction stir welded structure by means of integrated modeling approach

This model represents the friction stir welding of an L-shaped stringer on a 300 mm x 300 mm x 1.6 mm plate as shown schematically in Fig. 2.44a, both having the benchmark material properties, see Table 2.4 in Section 2.2.3.2. The cross-sectional dimensions of the stringer are given in Figure 2.44b, representing an industrial application (Marie and Allehaux, 2006).

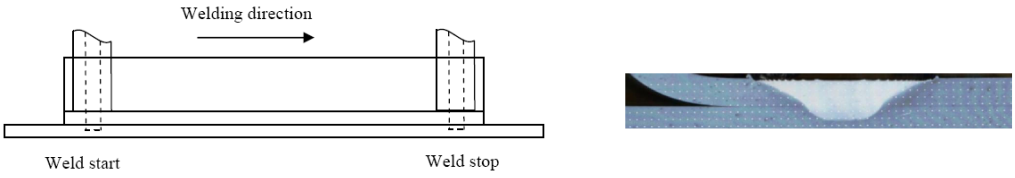
In order to define a benchmark model for a complex case as presented here, the interfacial condition between the plate and the stringer has been assumed to be "perfect", i.e. continuous properties across and along the plate at the interface. Hence the workpiece has been modeled as an extrusion of a continuous cross-sectional profile. The welding speed is 2 mm/sec and the heat input is 1050 W. The diameter of the tool shoulder is 20 mm. The weld starts and stops at 50-mm away from the ends of the plate. The heat transfer from the plate surfaces to the ambient is taken into account. The heat transfer



(a) (b)

Figure 2.44: (a) The friction stir welding of an L-shaped stringer on a plate [ref. PAPERs-II-IV] (Tutum et al., 2008a; Hattel et al., 2008). (b) Dimensions of the cross-section of the L-shaped stringer [ref. PAPER-II] (Tutum et al., 2008a).

coefficients are 20 W/(m^2 K) and 1000 W/(m^2 K) at top and bottom surfaces of the plate, respectively. The ambient temperature is 20 °C.



(a) (b)

Figure 2.45: (a) Side view of the welding process (Marie and Allehaux, 2006). (b) Cross-section of the stiffener after friction stir welded on the plate (Marie and Allehaux, 2006).

The plate is clamped at both sides along the weld direction and rollers are applied at the bottom of the plate due to the displacement restriction into the backing plate during the welding and cooling steps. In the releasing step, one of the corners of the plate is constrained and all of the previous mechanical boundary conditions are deactivated. At the end of the releasing step, the plate is clamped along one of the edges perpendicular to the welding line and a service load, which is defined as a bending load of 25 kg at the other end of the plate, is applied using a traction stress distributed over the cross section area of the "free" end.

The benchmark case defined above, i.e. having a total heat input of 1050 W, welding speed of 2 mm/s, plate dimensions of 300 mm x 150 mm x 1.6 mm, and a stringer geometry which is shown in detail at Fig. 2.44a, results in a maximum temperature of 500.6 °C during the whole welding step and 20 °C, which is the reference temperature, at the end of the cooling step of the simulation. An asymmetric longitudinal stress distribution is obtained, as shown in Fig. 2.46, since an asymmetric temperature field is observed due to the addition of the L-shaped stringer on the plate. Longitudinal and transverse stresses are reported along the dashed line in Fig. 2.47 (at the second column of the Step-4).

Some specific characteristics are noticed in Fig. 2.46, i.e. relatively higher compressive stress just near the right edge of the stringer and tensile stresses along the center line

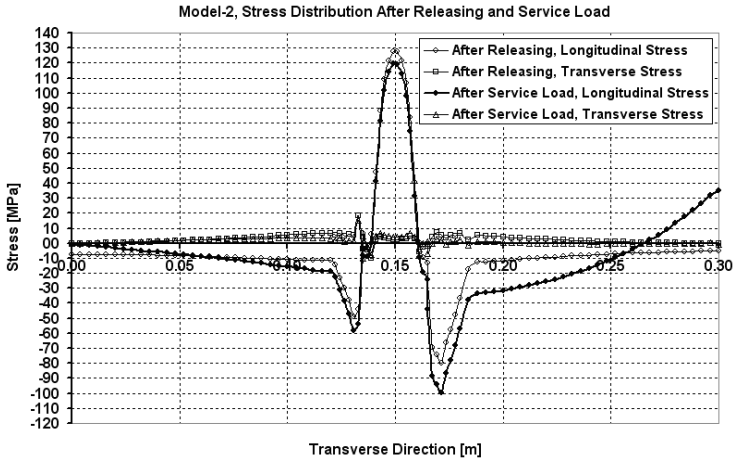


Figure 2.46: Stress distribution along the transverse direction of the plate with stiffener [ref. PAPER-II] (Tutum et al., 2008a).

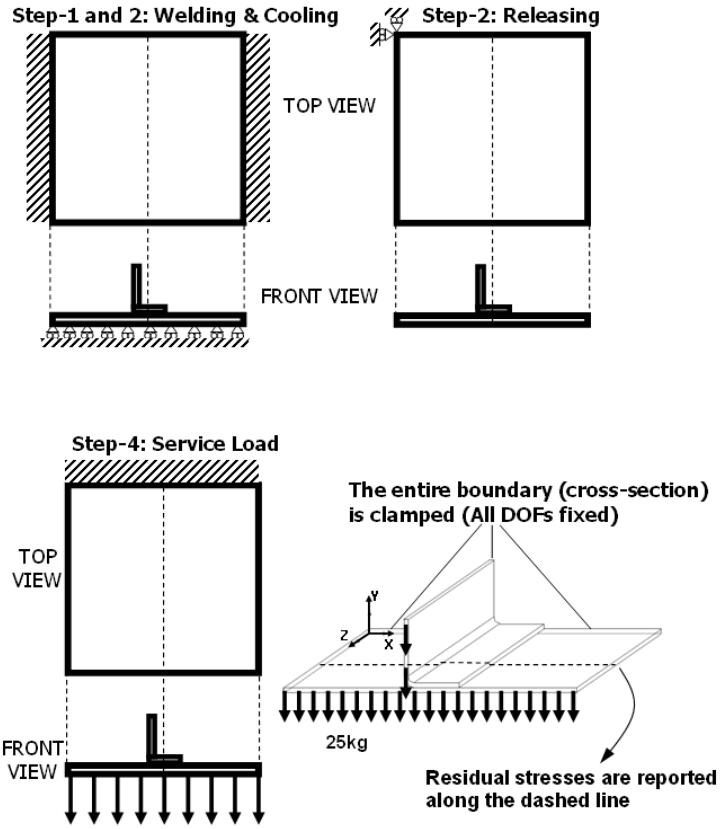


Figure 2.47: The 4 steps in the coupled residual stress and in-service load analyses [ref. PAPERs-II-IV] (Tutum et al., 2008a; Hattel et al., 2008).

of both the plate and the stringer, after releasing the clamping and the rollers. Similar characteristics are preserved after the application of the service load, which as mentioned before is defined as a bending load of 25 kg at the end of the plate while being clamped at the opposite edge. There is a 10 MPa decrease in the tension zone which could be considered as a relatively beneficial result from loading a welded structure with residual

stresses, while having a 20 MPa increase in the compression zone that should also be considered. Although the 10 MPa decrease might seem to be low, it should be recognized that for a 300 mm-long plate it is quite important since in industrial applications much longer plate-stringer constructions are used; as a result more than a 10 MPa decrease in longitudinal stress can be expected when considering the appropriate loading for a real welded structure having residual stresses.

From Fig. 2.48 it is seen that addition of temperature dependence for the yield stress results in higher longitudinal stresses as seen in previous sections while preserving the same level of transverse stresses in the transverse direction of the plate. Approximately 10 MPa in difference in longitudinal stress, can also be noticed after release and service load steps. It can be noted that the asymmetry in the stress distribution, which is explained as a result of adding a stringer on the thin plate, has almost disappeared and similar levels of compressive stresses at both sides of the stringer root have been obtained.

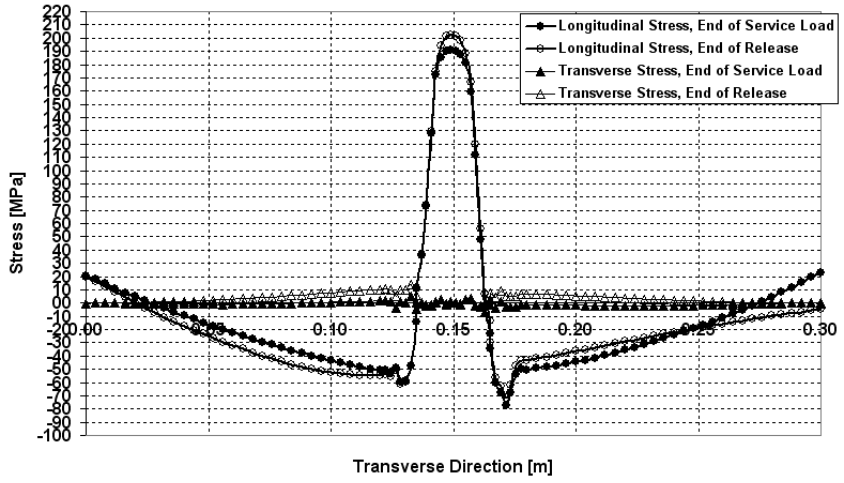


Figure 2.48: Stress distribution along the transverse direction of the plate for the temperature dependent yield stress case in Model-2 [ref. `PAPERs-II-IV`] (Tutum et al., 2008a; Hattel et al., 2008).

2.3.2 Integrated modelling of residual stresses and damage evolution during in-service conditions of FSW joints

In this section an application of coupling a residual stress model with a damage evolution model during loading is presented. The entire procedure is shown schematically in Fig. 2.49. First, the residual stress field is computed using a 3D quasi-static de-coupled thermo-mechanical implicit model implemented in ANSYS, in which the TPM model has been applied as the transient heat source. Following this, a smaller section in the middle of the plate is cut out while still having appropriate boundary conditions and internal variables, i.e. displacement and stress-strain fields, as coming from the residual stress model. Then, a "typical" yield stress profile observed in FSW of Al alloys is assigned to the cut out section in a subsequent static analysis and redistribution of the stress-strain field which satisfies the dynamic force equlibrium (due to the explicit formulation), Eq. (2.13) and the yield condition, Eq. (2.20), is obtained. Finally, the cut out section is sub-

ject to uni-axial loading during which a damage analysis is performed with an in-house 3D explicit FE-code.

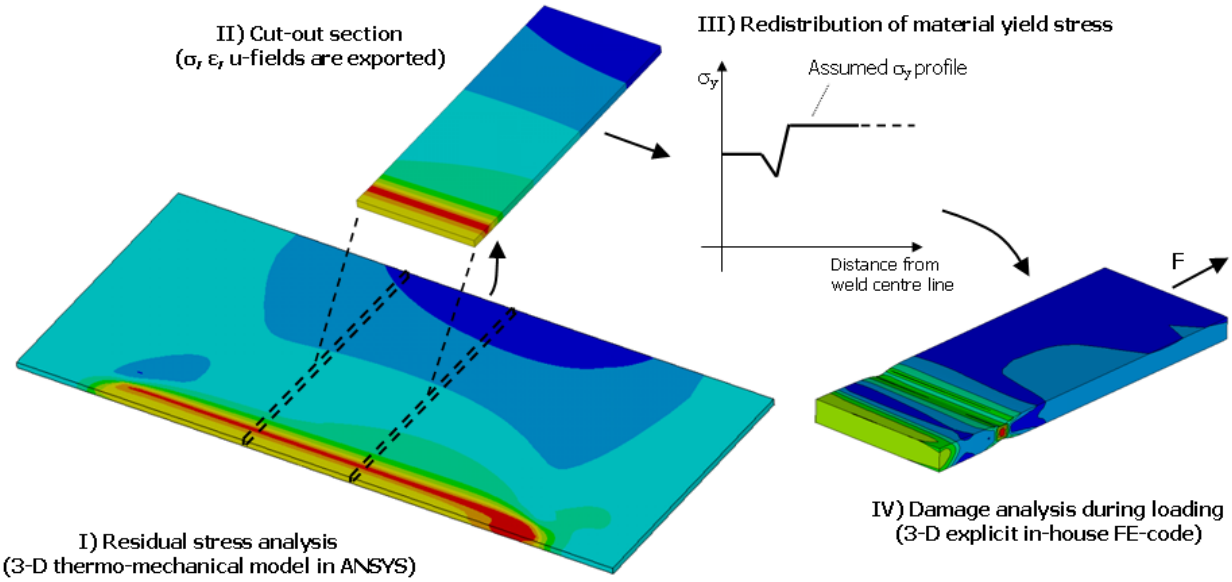


Figure 2.49: Coupling procedure of residual stress calculations with damage evolution during loading.

As the first step of this integrated approach, a sequentially coupled thermo-mechanical model is prepared. A 300 mm x 125 mm x 3 mm plate has been used as the workpiece geometry. The symmetry assumption made in the previous sections, i.e. considering only one of the plates due to neglection of the convection term induced by material flow under the tool, is repeated here. The workpiece material mainly represents the benchmark Al alloy of which properties are given in Table 2.4. The temperature dependent yield stress curve and the flow stress vs. equivalent plastic strain curves are shown in Fig. 2.34 and Fig. 2.50, respectively. Multi-linear isotropic hardening is used together with a strain rate independent plasticity model to mimic a power-law hardening material.

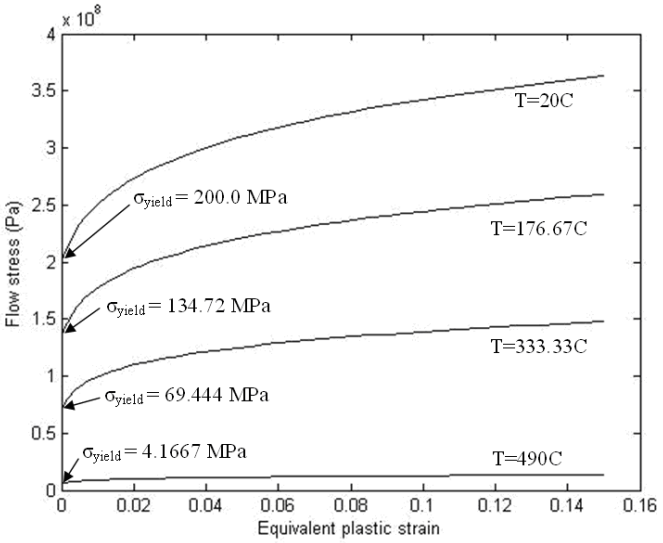
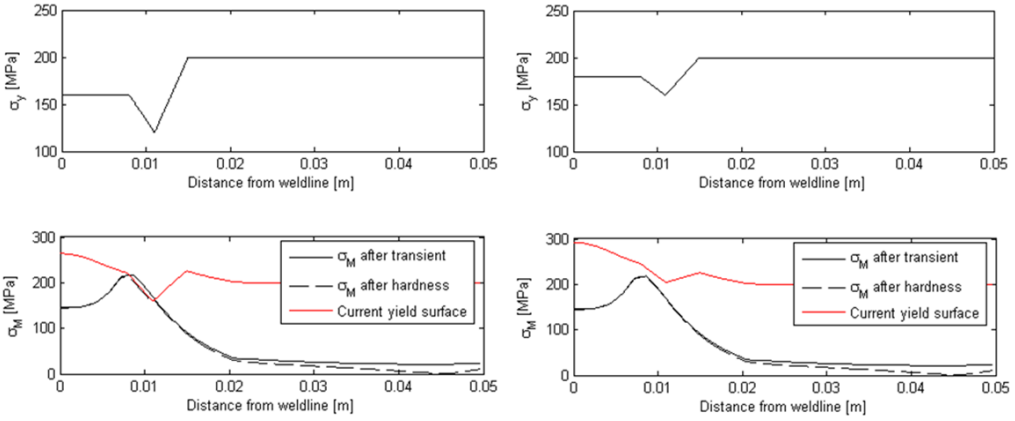


Figure 2.50: Temperature dependent flow stress vs. equivalent plastic strain curve.

The TPM heat source model described in Section 2.1.4 is used for the simulation of the transient evolution of the thermal field. The tool shoulder radius is 10 mm and the heat generation is assumed to be dominated by the friction at the shoulder and modelled as a surface heat flux acting on top surface of the workpiece. A welding speed of 3.25 mm/s and a rotational speed of 500 rpm is used as an input for the moving heat source. Following the transient thermal analysis, a quasi-static standard mechanical analysis, as presented in detail in Section 2.2.1, is performed in order to simulate transient stresses and finally obtain the residual stresses to pass on to the subsequent stages of the integrated procedure mentioned briefly above.

Two different yield stress profiles are assumed, i.e. Weld-1 and Weld-2, as shown in the top-row of Fig. 2.51. In both profiles, the base material yield stress is 200 MPa, but in Weld-1, it is reduced to 160 MPa (80 %) in the center, whereas in Weld-2 it is reduced to 180 MPa (90 %).



a) Weld-1 b) Weld-2

Figure 2.51: a) Assumed yield stress profile for Weld-1 (top) and redistribution of the stresses along the transverse direction (bottom), b) Assumed yield stress profile for Weld-2 (top) and redistribution of the stresses along the transverse direction (bottom).

The ductile damage development is modelled using the Gurson-Tvergaard-Needleman (GTN) model (Gurson, 1977; Tvergaard and Needleman, 1984) accounting for void nucleation, void growth and coalescence during plastic flow localization. Void nucleation is here taken to be plastic strain controlled (Chu and Needleman, 1980), while void coalescence is assumed to occur as a critical void volume fraction is reached. Similar studies can be found in Nielsen (2008); Nielsen and Tvergaard (2009), when omitting the post-weld stress-strain condition.

As seen from Fig. 2.52, Weld-2 shows the highest strength and ductility as compared to Weld-1 when subjected to the uni-axial tensile loading. This is also expected. On the other hand, it seems that taking the residual stresses into account increases the maximal stress a little, whereas the ductility is reduced since the plastic flow localization happens earlier during the loading process. The overall ductility of the specimen is thereby substantially reduced.

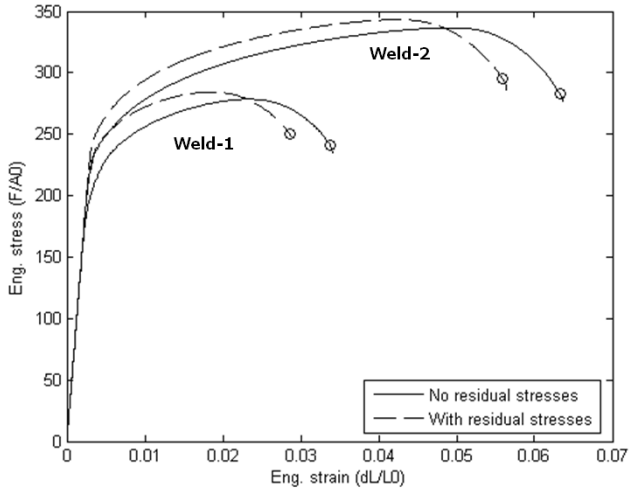


Figure 2.52: The resulting engineering stress vs. engineering strain curves when subject to uni-axial tensile loading

The transient evolution of the damage in the cut-out section is clearly seen in Figs. 2.54a and 2.54b for another assumed yield stress profile in which the softening zone is considerably larger as compared to the assumed profiles in Fig. 2.51. This is new profile is shown in Fig. 2.53.

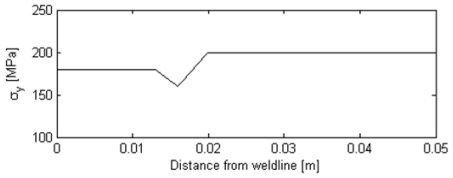
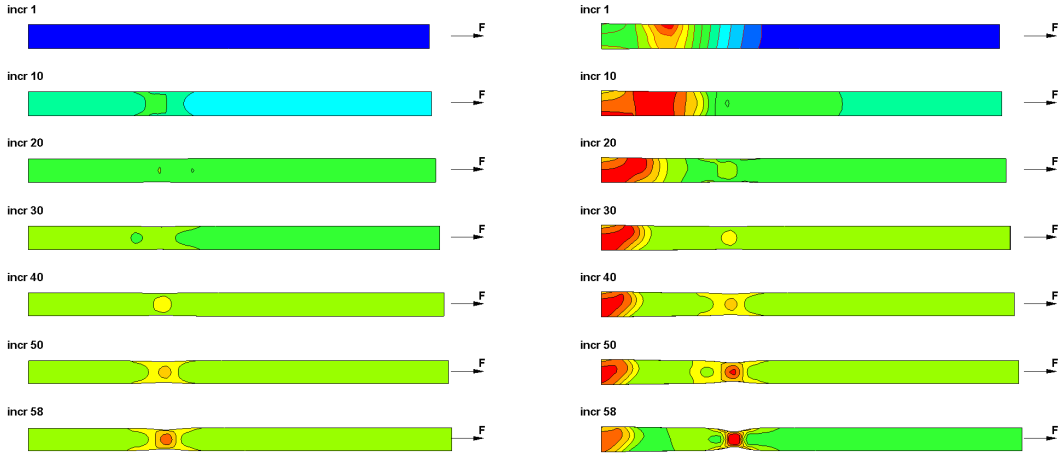


Figure 2.53: Assumed yield stress distribution for Weld-3

Fig. 2.54a shows a side-view of the cut-out section without residual stresses and Fig. 2.54b shows it with residual stresses. A substantial difference between these two cases in distribution of the longitudinal normal stresses is noted for the considered seven points in time. It is obvious that the damage evolves faster in the component on the right hand side (i.e. the one having residual stresses initially) compared to the other and it is initiated at the expected location, i.e. lowest corner of the assumed yield stress profile where the load caryying capacity is the lowest.



(a) (b)

Figure 2.54: a) Evolution of stress component along the weldline - without residual stresses (Weld-3). b) Evolution of stress component along the weldline - with residual stresses (Weld-3).

Chapter 3

Optimization

3.1 An Overview on Optimization

Optimization in general is an attempt of finding the best available solution(s) from a set of goal(s) and limitations whether in a systematic or a trial-and-error approach, arising in the field of mathematics, engineering, operations research or hidden intrinsically in daily life applications or being presented by so many brilliant examples in the nature.

One could mention game artificial intelligence (AI) as a hidden example of optimization arising in the entertainment business, where different imitation techniques are used to generate artificial players (agents) showing human-like behavior, i.e. trying not only to inflict as much damage to their opponents as possible, but also attempting to trick them, playing tactics and strategy (e.g. combat, soccer, car racing), and all that is based on an evolutionary learning approach (Priesterjahn, 2007; Tanev and Shimohara, 2007). Another popular and addictive routine "endeavor" is solving a Sudoku number puzzle which is issued in almost all daily newspapers. Although most of the readers are not aware of it, it is actually a constraint satisfaction problem that they are involved in and moreover, it can be solved, rated and generated by combinatorial genetic algorithms with different difficulty levels (Mantere and Koljonen, 2008; Delahaye, 2006). Similar to Sudoku ("singular number" in Japanese) problems, which is an example of the more general combinatorial optimization problems, Rubik's Cube being a three-dimensional logic game invented by E. Rubik in 1974 became one of the most popular puzzle games (it even won the German Game of the Year special award for Best Puzzle in 1980 (Wikipedia, 2009b)) and turned into a challange for the application of state-of-the-art evolutionary computational techniques, e.g. the Evolution Strategies (Herdy, 1991) and the Harmony Search (Geem et al., 2001) which is a music-inspired evolutionary algorithm mimicking the improvisation process of music players (Geem, 2007).

As mentioned before, the nature is a brilliant teacher in optimization for the scientist, the engineer, or whoever interested in improving any kind of design, process or schedule, etc. More specifically, e.g. in structural optimization, a number of efficient optimization algorithms mimicking natural phenomena and physical processes have been developed. Genetic algorithms (GAs), simulated biological growth (SBG) and simulated annealing (SA) as well as particle swarm optimization (PSO) are some of those successful and popular techniques inspired by the nature. The SBG mimics phenomena that have been observed in the mechanism of tree growth. This involves the self-optimization of

living trees which always try to grow into shapes of constant surface stress (Mattheck and Burkhardt, 1990; Mattheck, 1997). Another example is the phenomenon that has been observed in animal and bone tissue (Huiskes et al., 1987). This involves the addition of bone material in regions of high stress and conversely, the reduction of material in regions of low stress. Simulated annealing (Metropolis et al., 1953), which is very suitable for large-scale optimization problems (e.g. the popular traveling salesman problem (TSP)), is based on statistical thermodynamics and is used to simulate the behavior of the atomic arrangements in solid material during an annealing process. The method has also been used successfully for designing complex integrated circuits: The arrangement of several hundred thousand circuit elements on a tiny silicon substrate is optimized so as to minimize interference among their connecting wires (Vecchi and Kirkpatrick, 1983). The PSO (Kennedy and Eberhart, 1995) is based on the simulation of a simplified social model which evolves (socializes) by the interaction of the individuals (particles) of the population (swarm). Some behavioral patterns that inspired scientists were the underlying rules that enabled large numbers of birds to flock (or fish crowds to school) syncronously, often changing direction suddenly, scattering and regrouping, etc. Thus also in the PSO, as an imitation of this behavioral pattern, the fitness (success criterion) of each particle in the population is iteratively updated, and they will not only remember where they had been successful, but they will also inform the neighbors as well, which eventually will be leading to the convergence towards the optimum solution.

A very well-known application of optimization in engineering is the search for optimal structures with respect to load carrying capacity. Or in other words, aiming at the lightest structure that would be able to carry a prescribed load within certain limits (material's yield or deformation limit, amount of material to be used, cost, etc.). This challenging goal of engineering study calls for some specialized branches trying to find answers to these somewhat practical problems, such as size, shape or topology optimization problems. For instance, the first class of these optimization problems could be exemplified by truss structure optimization, in which the optimal cross-sectional areas of the members are sought in order to achieve a desired static or dynamic response satisfying some performance limits. A recent study made by the author (Tutum, 2006, 2005) investigates the optimum truss design, in which the member areas are provided as the design variables, and simultaneously subjected to stress, displacement, buckling and natural frequency constraints. The members with the minimum cross-sectional areas are removed from the system, meanwhile satisfying the stability of the global structure. Thus, it is actually a topology optimization problem, where an optimum lay-out of a given amout of material in a limited space is sought, for the discrete systems. As expected, this application can also be generalized into continuum structures as well as interaction with multi-physics phenomena, and further details about the theory and wide variety of applications on topology optimization is given by Bendsøe and Sigmund (2003).

A large number of optimization applications have already been given in order to control different aspects of welding applications. For example, Mishra and DebRoy (2007) used a real-coded genetic algorithm (GA) and a neural network (NN) that was trained with the simulation results of a heat transfer and fluid flow model in order to obtain 2 desired target values of weld pool penetration and width for the gas tungsten arc welding. A numerical optimization study, i.e. application of the steepest descent algorithm together with an analytical sensitivity analysis, has been performed by Michaleris et al. (1999) to design the thermal tensioning process which utilizes transient temperature gradients

by localized heating in order to minimize welding residual stress and distortion. Finally, Bogomolny and Bendsøe (2007) applied shape optimization for the resistance welding process, incorporating surrogate modeling based on a Kriging approximation, in order to improve the fatigue performance of the electrodes.

Optimization studies in literature regarding the FSW process consist of both numerical and experimental work. Some of these techniques have been used to investigate uncertain process parameters that enable model behaviour to be fitted to available experimental data. Application of the Differential Evaluation (DE) technique (Storn and Price, 1997) for reducing the uncertainty associated with specific process parameters, i.e. the friction coefficient, the extent of slip between the tool and the workpiece, the heat transfer coefficient at the bottom of the workpiece, the mechanical efficiency, and the extent of viscous dissipation converted to heat, is studied by Nandan et al. (2008b). Richards et al. (2008a) present detailed numerical studies with experimental validation of the mechanical tensioning during FSW and post-welding as well as an investigation of the effectiveness of dynamic cooling in order to control the residual stresses in the FSW process (Richards et al., 2008b). Maximization of the lap joint strength per unit length is examined by Fratini and Corona (2007) using a gradient technique (the steepest descent method) together with an experimental procedure resulting in a mechanical performance of the joint equal to 85% of the parent material resistance. Improvement of the tensile strength and thereby the fatigue life of a FSW butt-weld by an experimental study and application of the Taguchi DOE technique on the process parameters, i.e. the tool rotational speed and the traverse welding speed, was obtained in Lombard et al. (2008) and Lakshminarayanan and Balasubramanian (2008) respectively. Finally, Investigations of process parameters that can control the residual stresses have also been considered, for example using the multi objective genetic algorithm (MOGA-II by Poloni and Pediroda (1997)) applied to a two-dimensional transient thermo-mechanical model (Tutum et al., 2008b) [ref. `PAPER-III`].

George E. P. Box, a world famous industrial statistician from UK, is credited with the quote "all models are wrong, but some are useful". The practical question to be asked is "How wrong do they have to be, not to be useful?" (Wikiquote, 2008). The following optimization studies given in Sections 3.3 and 3.4 are presented keeping this very basic question in mind.

In the next three subsections, it is intended to give an overview of the optimization techniques in use, and these are mainly categorized into classical methods, evolutionary algorithms and hybrid techniques.

3.1.1 Classical Methods

Classical or traditional optimization methods depend on deterministic rules in the search for an optimal solution. This means that a specific set of rules, whether using a gradient information, curvature or an approximation of these as well as an imitation of the gradient information (direct methods), are followed for making a move in the direction, resulting in a minimum or maximum value of the function. These methods consist of the direct search methods, in which only the function values are used (e.g. Random Search, Powell's Method (Powell, 1964), the Simplex Search Method (Nelder and Mead, 1965), Hooke-Jeeves Pattern Search Method (Hooke and Jeeves, 1961), etc.), and the gradient

based algorithms which mainly construct their search directions, as mentioned above, with first-order (gradient) or second-order (curvature) derivatives of the given objective and/or constraint functions, e.g. the Steepest Descent Method, the Conjugate Gradient Method, the Variable Metric Methods (e.g. the Davidon-Fletcher-Powell (DFP) method, the Broydon-Fletcher-Goldfarb-Shanno (BFGS) method which approximate the inverse of the Hessian matrix and for this reason they are also called as quasi-Newton Methods) as first order methods, and the Newton's Method as the second-order method which uses the Hessian directly. A single solution or design point is updated in an iterative manner mostly in which a search direction (e.g. the steepest descent) is found first, and then a step size which minimizes the function value is determined along this direction using a line-search technique, e.g. the Bisection algorithm, the Golden-Section algorithm, until the convergence criterion, i.e. the change in slope or the difference in the successive iterations, is met. Although most of the aforementioned algorithms are mainly developed for unconstrained problems, they have also been used for constrained problems with slight modifications such as converting the problem into an unconstrained one by using (Internal or External) Penalty Methods or a Lagrangian function in which the objective function and the constraints are augmented and solved sequentially, as in e.g. the Sequential Linear Programming (SLP) and the Sequential Quadratic Programming (SQP). In these problems the Karush-Kuhn-Tucker (KKT) conditions serve as the basis for the necessary optimality conditions. If the design space is convex and the Hessian matrix is positive definite, then the KKT conditions also correspond to the sufficiency conditions, which result in that the optimum solution is the global one (Luenberger, 2003; Arora, 2004; Vanderplaats, 2005; Haftka and Gürdal, 1992). One of the difficulties using gradient based algorithms arises in the application on noisy, multi-modal or non-convex landscapes. Then, the optimum solution found, would most likely be a local one depending on the starting (guess) solution which brings the criticisms on robustness. In order for a gradient based algorithm to be capable of finding the global optimum, one should run the optimization simulation with a set of different starting solutions or having a search direction not only aiming at the steepest descent direction, but also being able to consider the steepest ascending direction in order to get out of a local optimum (Deb, 2006). Another concern about the application of the classical algorithms is the multi-objective optimization (MOO). Since the classical algorithms are developed to deal with only a single objective, the MOO problem should be scalirized by some techniques, e.g. the Weighted Sum method (potential weaknesses arise in case of having a non-convex Pareto-front or a non-even distribution of solutions accross the Pareto front (Das and Dennis, 1997)), the Weighted Metric methods (Miettinen, 1999; Deb, 2001), in order to handle multiple criteria. In some cases one of the objectives is chosen to be optimized and the others are taken as design constraints as considered in the ε-Constraint method (Miettinen, 1999; Deb, 2001). As a final application of classical gradient-based algorithms in MOO problems, the NBI-NLPQLP (Das and Dennis, 1998), Normal Boundary Intersection method coupled with the NLPQLP single-objective solver (Schittkowski, 1985) can be mentioned. This method reduces the MOO problem to many single-objective constrained subproblems (the so-called NBI subproblems), which indicates the limitation involved with smooth functions. It is a fast and accurate algorithm, but not as robust as an MOEA.

3.1.2 Evolutionary Algorithms

Evolutionary algorithms (EAs) are non-deterministic (stochastic) methods that mimic evolutionary principles, e.g. natural selection and the survival of the fittest, to constitute their optimization strategy. They work with a set of solutions (population) instead of a single point as in traditional (classical) methods and this gives an opportunity to attack a complex problem (discontinuous, noisy, multi-modal, etc.) in different directions allowing the algorithm to explore as well as exploit the search space. This capability gives an advantage for having a more robust search strategy compared to traditional algorithms. Since they don't need any gradient information, they are very suitable for black-box (e.g. commercial software) optimization applications. Besides their relatively easy computational implementation, they are also proper for distributed computing applications since all individuals (designs) can be computed independently. Due to their population based search strategy, they have been more popular for the MOO problems, often having conflicting objectives resulting not only in a single optimum solution, but in a set of trade-off solutions (Pareto-optimal set), for the last last two decades. Many multi-objective evolutionary algorithms have been developed in order to solve multiple conflicting goals in an ideal way, that is without using weightings between objectives or any scalarization techniques. Besides convergence, a well-spread distribution of these solutions across the Pareto-front has been considered to be an important challenge for MOEAs. Some of the difficulties that a MOEA may have to tackle have been pointed out by some researchers (Zitzler et al., 2000). `TECHNICAL REPORT-I` and `TECHNICAL REPORT-II` also give an overview of MOEAs and some of the well-known benchmark problems. The importance of the elitism concept, which allows the survival of the best solutions ever found during the generations, in MOO research was recognized and experimentally supported by different studies (Parks and Miller, 1998; Zitzler et al., 2000). After some implementations of non-elitist MOEAs (e.g. the Multi Objective Genetic Algorithm (MOGA) by Fonseca and Fleming (1993), the Niched Pareto Genetic Algorithm (NPGA) by Horn et al. (1994) and the Non-dominated Sorting Genetic Algorithm (NSGA) by Srinivas and Deb (1994)), many new ones (e.g. the Strength Pareto Evolutionary Algorithm (SPEA and SPEA2) by Zitzler and Thiele (1998, 1999); Zitzler et al. (2001), the Pareto Archived Evolution Strategy (PAES) by Knowles and Corne (1998), the NSGA-II (the successor of the NSGA) by Deb et al. (2002) and the Pareto Envelope-based Selection Algorithm (PESA) by Corne et al. (2000)) appeared incorporating elitism in different ways. Evolutionary algorithms can be categorized into four major areas, i.e. Genetic Algorithms (GAs), Evolution Strategies (ESs), evolutionary programming (EP) and genetic programming (GP). Although their working principles are similar, there are still specific differences that allow them to be specialized in some applications. More detailed information is given by Goldberg (1989); Deb (2001); Michalewicz (1992) and Mitchell (1998) on GAs, by Rechenberg (1973, 1997) and Schwefel (1995) on ESs, by Fogel (1999) on EP, by Koza (1993) and Poli et al. (2008) on GP, by De Jong (2006) on the general theory of evolutionary computation and finally by Deb (2001); Miettinen (1999) and Branke et al. (2008) on the evolutionary and classical multiobjective optimization.

3.1.3 Hybrid Techniques

Although evolutionary computation has become an important problem solving methodology with its population-based collective learning process, self-adaptation, and robustness, the performance of the algorithms still depend on proper selection of various parameters

(e.g. probabilities, selection and mutation schemes, etc.), namely the proper relationship between the exploration and the exploitation capabilities avoiding premature convergence, as mentioned above. Moreover, the computational speed is relatively slow as compared to the classical (deterministic) algorithms. Therefore, as expected, the need for hybrid algorithms, which combine an evolutionary algorithm with e.g. a local search method, emerges, aiming at both robust and accurate solutions with (if at all possible) less computational cost. Local search methods may be incorporated within the population members (parents) or among the offsprings. The architectures of hybrid evolutionary algorithms have been summarized by Grosan and Abraham (2007) as follows: hybridization between two different EAs (a GP asisted GA), an EA with a neural network, a particle swarm optimization (PSO) or an ant colony optimization (ACO) as well as hybridization between EA and other heuristics (such as local search, tabu search, simulated annealing, hill climbing, etc.). However, as the *No Free Lunch Theorem* proposes, on average, all black-box algorithms have identical behavior, thus there is not a definite answer for which local search procedure to use for any sort of problems. Deb and Goel (2001) presented some test cases on shape optimization of mechanical components, i.e. a simply-supported plate design, a cantilever plate design, a hoister design and a bicycle frame design, using the NSGA-II together with a hill-climbing local search method in order to move closer to the true Pareto-optimal front. In this approach, the results obtained from the MOEA (the NSGA-II) have been clustered for reducing the size of the non-dominated solution set. Following this, the local search method is applied on each member of the reduced set by using a pseudo-weight vector for each solution to be used for the weighted objective metric which scalarizes the multiple conflicting objectives, i.e. minimization of the weight and the deflection of the structure. In another work by Ishibuchi and Murata (1998), a different approach has been followed. The local search method is used during a GA run and applied on each solution before including it in the population, which is reported to have less computational cost as compared to the posteriori approach that Deb and Goel (2001) used. Olympio (2007) hybridized the ε-NSGA-II, which uses ε-dominance for population archiving and dynamic population sizing, by Kollat and Reed (2005) and the weighting sum method for the local search as done by Deb and Goel (2001). First, this hybrid method (the HMOGA) was applied to three test problems (ZDT1, DTLZ3 and DTLZ6), and then to a multi-objective topology optimization problem (related to the design of flexible skins for a morphing aircraft) aiming at minimization of the tip deflection and the mass of the beam. The performance of the algorithm on the three test cases was found to be better compared to the single MOGA case (without local search), but the results for the topology optimization problem have been reported to be not satisfactory due to unrealistic solutions, i.e. related to chekerboard effects (mesh independency filtering was needed). Rigoni (2004a) applied another hybridization technique on some benchmark cases, i.e. TNK, POL (these are also used in the `TECHNICAL REPORT-I` and `II`) and hole problems (Rigoni, 2004b), combining the NBI-NLPQLP with the MOGA-II. This methodology consists of starting with the MOGA-II computation then each single portion of the Pareto-front (since it is discontinuous) is isolated as an independent problem which is handled by an independent NBI-NLPQLP. Recently, Muller et al. (2009) succesfully combined the Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) with the Particle Swarm Optimization (PSO) in order to solve highly noisy and multi-funnel (valley) single objective optimization (SOO) problems where there are many local minima.

3.2 Description of the Algorithms

In the next two subsections, two algorithms, i.e. the Sequential Quadratic Programming (SQP) and the Non-dominated Sorting Genetic Algorithm (NSGA-II), will be described in detail. The first one is used in `PAPER-I` (Tutum et al., 2007) for the single-objective optimization problem, whereas the second one is used in `PAPER-III` (Tutum et al., 2009), `TECHNICAL REPORT-I` and `II` for the multi-objective optimization cases.

3.2.1 Sequential Quadratic Programming (SQP)

The Sequential Quadratic Programming (SQP) method, which is a gradient-based optimization technique, represents the state of the art among nonlinear programming (NLP) methods. Schittkowski (1985) has implemented and tested a version that outperforms every other tested methods in terms of efficiency, accuracy and percentage of successful solutions over a large number of test problems. Due to its success, it was also implemented in both commercial softwares (e.g. NAG-SQP by NAG Ltd. (1988), DOT-SQP by Vanderplaats (1994), `fmincon` by The Math Works Inc. (2006), part of *Active Set Algorithms* in KNITRO by Waltz and Plantenga (2009), NLPQLP/MIPSQP/NBI-NLPQLP in modeFRONTIER by ESTECO s.r.l. (2007)) and open source environments (e.g. OpenOpt (2006), DONLP2 by Spellucci (1999)).

Based on the work of Powell (1978), the method is able to closely mimic the Newton's method (which is a second-order derivative method) for constrained optimization similar to an unconstrained optimization. At each major iteration, an approximation of the Hessian of the Lagrangian function is made using a quasi-Newton updating method, i.e. the BFGS (as Powell (1978) suggests) or the DFP, as mentioned in Section 3.1.1. This is then used to generate a Quadratic Problem (QP) sub-problem, composed of a quadratic objective function and linear constraints, whose solution is used to form a search direction for a line search procedure (Luenberger, 2003; Arora, 2004; Vanderplaats, 2005). Despite the high complexity of the QP-direction seeking algorithm, it is rewarded by faster convergence (Haftka and Gürdal, 1992). This method is also known as Iterative Quadratic Programming, Recursive Quadratic Programming, or Constrained Variable Metric methods.

For a given nonlinear constrained problem, as shown in Eq. (3.1),

$$\begin{aligned} &Minimize: f(\mathbf{X}) \\ &subject\ to: g_j(\mathbf{X}) \leq 0, \quad j = 1, m \\ &\qquad\qquad h_k(\mathbf{X}) = 0, \quad k = 1, l \end{aligned} \tag{3.1}$$

where $f(\mathbf{X})$ is the objective function, $\mathbf{X}$ is the design point (vector) while $g_j(\mathbf{X})$ and $h_k(\mathbf{X})$ represent the inequality and the equality constraints, respectively, the QP-subproblem, which will provide a search direction $\mathbf{S}$, is formulated as follows,

$$\begin{aligned} &Minimize: q(\mathbf{S}) = f(\mathbf{X}) + \nabla f(\mathbf{X})^T \mathbf{S} + \frac{1}{2}\mathbf{S}^T \mathbf{B}(\mathbf{X}, \lambda)\mathbf{S} \\ &subject\ to: \nabla \mathbf{g_j}(\mathbf{X})^T \mathbf{S} + \mathbf{g_j}(\mathbf{X}) \leq 0, \quad j = 1, m \\ &\qquad\qquad \nabla \mathbf{h_k}(\mathbf{X})^T \mathbf{S} + \mathbf{h_k}(\mathbf{X}) = 0, \quad k = 1, l \end{aligned} \tag{3.2}$$

where $\mathbf{B}$ is a positive definite approximation to the Hessian of the Lagrangian function which is given below and λ is the vector of the Lagrange multipliers associated with the constraints. The Lagrange multipliers in general are not only useful for checking optimality (i.e. KKT conditions, as discussed in Section 3.1.1), but they also provide information about the sensitivity of the optimal solution to problem parameters, which can further be interpreted as a measure of the effect of a change in the constraints on the objective function (Haftka and Gürdal, 1992).

After the search direction ($\mathbf{S}$) is found from the QP-subproblem, the design point will be updated by,

$$\mathbf{X}^{i+1} = \mathbf{X}^i + \alpha \mathbf{S} \tag{3.3}$$

where the superscript-i denotes the iteration number, α is the unknown (scalar) step size to be moved along the search direction ($\mathbf{S}$), which will be found by minimizing the Lagrangian function ($\mathcal{L}$),

$$\mathcal{L}(\mathbf{X}, \lambda) = f(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j g_j(\mathbf{X}) + \sum_{k=1}^{l} \lambda_{m+k} |h_k(\mathbf{X})| \tag{3.4}$$

The matrix $\mathbf{B}$ is initialized to some positive definite matrix (e.g. the identity matrix, $\mathbf{I}$ as represented in the flowchart in Fig. 3.1) and then updated ($\mathbf{B}^*$) using the BFGS algorithm,

$$\mathbf{B}^* = \mathbf{B} - \frac{\mathbf{B}\mathbf{p}\mathbf{p}^T\mathbf{B}}{\mathbf{p}^T\mathbf{B}\mathbf{p}} + \frac{\eta\eta^T}{\mathbf{p}^T\eta} \tag{3.5}$$

where,

$$\mathbf{p} = \mathbf{X}^{i+1} - \mathbf{X}^i \tag{3.6}$$

$$\eta = \theta\mathbf{y} + (1-\theta)\mathbf{B}\mathbf{p} \tag{3.7}$$

$$\mathbf{y} = \nabla_x \mathcal{L}(\mathbf{X}^{i+1}, \lambda^{i+1}) - \nabla_x \mathcal{L}(\mathbf{X}^i, \lambda^i) \tag{3.8}$$

Here ∇_x denotes the gradient of the Lagrangian ($\mathcal{L}$) with respect to the design vector, $\mathbf{X}$. In order to guarantee the positive definiteness of $\mathbf{B}$, η is updated as shown above and θ is determined by Eq. (3.9).

$$\theta = \begin{cases} 1.0 & \text{if } \mathbf{p}^T\mathbf{y} \geq 0.2\mathbf{p}^T\mathbf{B}\mathbf{p} \\ \frac{0.8\mathbf{p}^T\mathbf{B}\mathbf{p}}{\mathbf{p}^T\mathbf{B}\mathbf{p} - \mathbf{p}^T\mathbf{y}} & \text{if } \mathbf{p}^T\mathbf{y} < 0.2\mathbf{p}^T\mathbf{B}\mathbf{p} \end{cases} \tag{3.9}$$

The flowchart of the SQP algorithm is shown in Fig. 3.1.

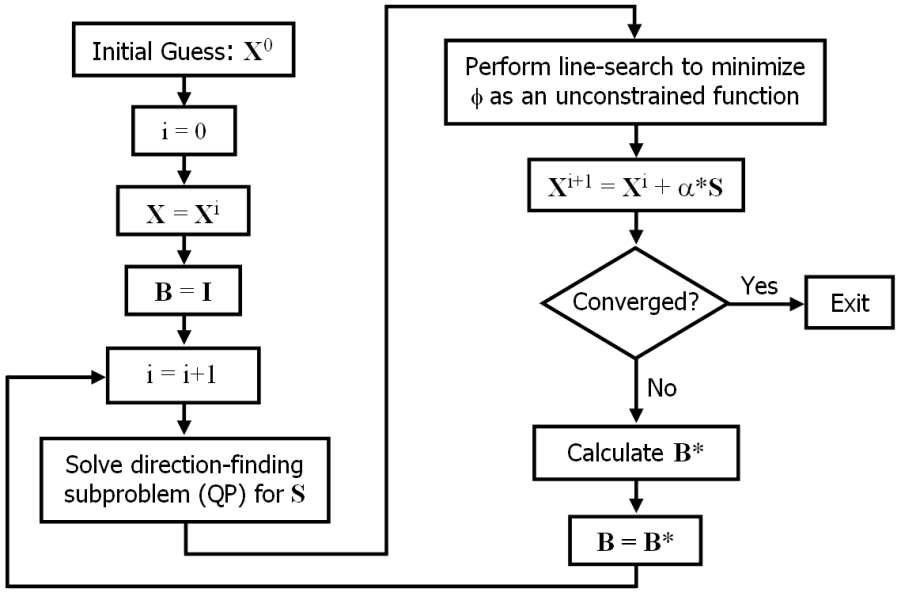


Figure 3.1: The flowchart of the SQP algorithm (Vanderplaats, 2005).

3.2.2 Non-dominated Sorting Genetic Algorithm (NSGA-II)

NSGA-II is a multi-objective evolutionary algorithm (MOEA) proposed by Deb et al. (2000a,b, 2002), which enables finding well-spread multiple Pareto-optimal solutions for an MOO problem by incorporating three substantial features, i.e. non-dominated sorting, an explicit diversity preserving mechanism and elitism. As the name indicates, it is the next generation of its predecessor, i.e. the non-dominated sorting genetic algorithm (NSGA) (Srinivas and Deb, 1994), having substantial improvements with respect to computational complexity and robustness in the sense of having no need to use a sharing parameter (Goldberg and Richardson, 1987) to enforce diversity between the members.

In the case of SOOP (excluding multi-modal problems), it is easier to identify the optimal solution compared to the case of MOOP where there are more than one criterion to satisfy, i.e. hence more than one best (elite) solution to consider. As it was put forward by Goldberg (1989), "if we refuse to compare apples to oranges, then we must come up with a different definition of optimality, one that respects the integrity of each of our seperate criteria" which points out the concept of *Pareto Optimality*, correlated to the fundamental concept of *domination*.

Since the concept of domination allows a way to compare solutions with multiple objectives, most MOO algorithms use this domination concept to search for non-dominated solutions, i.e. the ones that constitute the Pareto-front which will be described in Eq. (3.10). This concept can be explained with a simple example in 2D objective space, in which the minimization of both objective functions (i.e. f_1 and f_2) is considered, as shown in Fig. 3.2. In this figure, point-1 representing a solution or a design, e.g. an individual or a member in evolutionary computation parlance, splits the 2D objective space into two zones: the first zone (the shaded region); the set of the points dominated by point-1, and the second zone; the set of points not dominated by point-1 (the remaining three quadrants), i.e. the *non-dominated* or $noninferior$ set of points. Thus, the two

conditions stated below need to be satisfied if point-1 dominates point-i, where point-i is any other design point in the objective space (Deb, 2001):

- point-1 is no worse than point-i in all objectives.
- point-1 is strictly better than point-i in at least one objective.

These two statements give a general description for the comparison of the two solutions in terms of "betterness", hence emphasizing the applicability of this comparison for other types or combinations of conflicting objectives besides the minimization - minimization type as shown in Fig. 3.2, i.e. maximization - maximization, maximization - minimization as well as minimization - maximization. In case of the aforementioned minimization-minimization problem, it is formulated as follows (Goldberg, 1989; Zitzler et al., 2000):

$$(\forall j \ : f_j(x_1) \leq f_j\,(x_i)) \cap (\exists k \ : f_k(x_1) < f_k(x_i)) \tag{3.10}$$

where x_1 represents point-1 in Fig. 3.2 and x_i represents any design point-i in the same 2D objective space. Eq. (3.10) can be put in words as; at least in one objective, design-1 is "less" than design-i, while in others they might be equal.

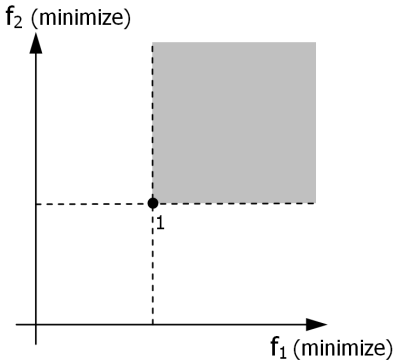


Figure 3.2: The two seperate sets of solutions defined with respect to point-1: i) the set of dominated solutions (the shaded region), ii) the set of non-dominated solutions.

The incremental scheme of an evolutionary algorithm will update the location of point-1 and bring it to a new location in 2D objective space, e.g. point-2, point-3, point-4 or point-5, as shown in Fig. 3.3. If point-1 is a solution from the previous generation and the current generation contains point-4, then the new position is very a favorable one, since point-4 is not only non-dominated by point-1, but also point-4 dominates point-1. This kind of evolution is always desirable since it promotes convergence, thus this transition has certainly to be preserved. On the other hand, if the evolutions bring point-1 to point-3 or point-5, the new point is also a non-dominated one, even though it is not dominating point-1, eventually this transition should also be preserved in order to provide a well-spread distribution of the points along the Pareto frontier.

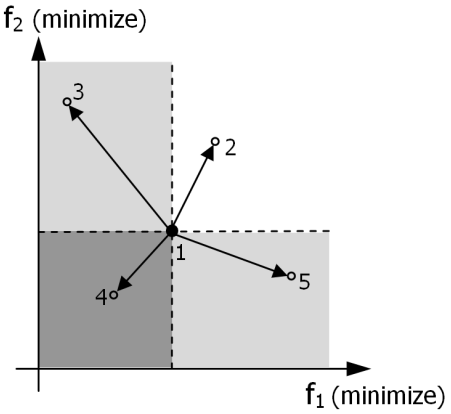


Figure 3.3: Possible search directions for point-1 along the increments (generations) of an evolutionary algorithm (adapted from Poles (2003)).

In case of having a population composed of these five points, a need for an overall relation of dominance arises, i.e. the one-to-one comparison made for point-1 should be generalized for each point. Fig. 3.4a shows the distribution of these five points together with their domination regions which are distinguished with different tones of colors and it is clearly seen that point-4 dominates points-1, 2 and 5, while point-3 does not dominate any other points for the given distribution. Since point-3 and 4 are not dominated by any other points, i.e. the non-dominated solutions, they constitute the non-dominated front which is called the Pareto-front (see Fig. 3.4b) in an MOO context. A similar comparison will reveal that point-1 and point-5 constitute the next domination front which in itself is dominated by the Pareto front and shown as the closest dashed line to the Pareto front in Fig. 3.4b.

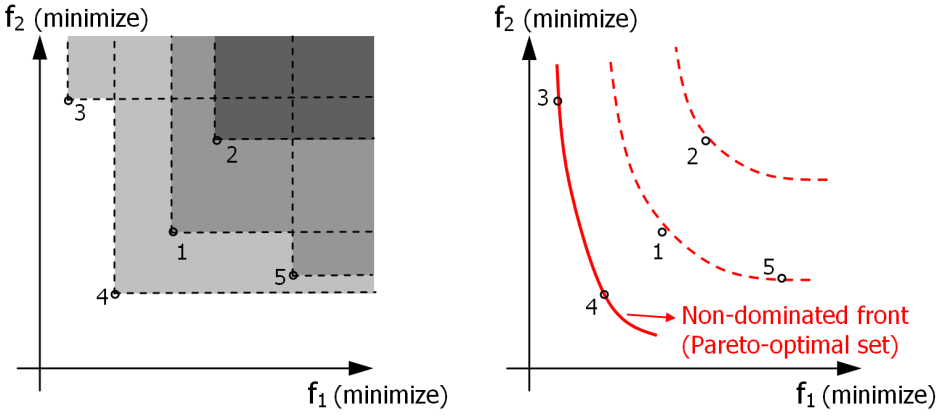


Figure 3.4: a) Each solution in a population has its own sets of dominated and non-dominated solutions (inspired by Zitzler (1999)). b) The dominated and the non-dominated (Pareto-optimal) fronts.

There are different approaches, having different computational complexities (i.e. total number of function evaluations), to apply this procedure of sorting the population into different domination levels for a given set of solutions. Eventually, the efficiency of

the sorting procedure is a critical issue in the case of a relatively high population size (e.g. N=100) and a high total number of generations (e.g. gen_{max}=100) for real world applications, particularly depending on the computational cost of each design. Further details regarding the different sorting approaches are given in Deb (2001). The NSGA-II uses a non-dominated sorting algorithm requiring O(MN^2) overall computational complexity, where M is the number of objectives and N is the population size. It starts with calculation of the two entities for each member: a-) n_i, the *domination counter*, which is the number of solutions which dominate the member-i, b) S_i, which is a set of members which the member-i dominates. It results in all the solutions across the first non-dominated front having a domination counter as zero (n_i=0). Following the identification of these non-dominated solutions, each member in their S_i sets is visited and its domination counter is reduced by one. This procedure is repeated until all solutions in the population are visited and their n_i and S_i entities are updated. In Fig. 3.5, five solutions previously shown in Fig. 3.4 are presented with their final domination information, i.e. solutions-3 and 4 having a domination counter of 1, solutions-1 and 5 having a domination counter of 2, and finally solution-2 having a domination counter of 3 as being on the third and the last front.

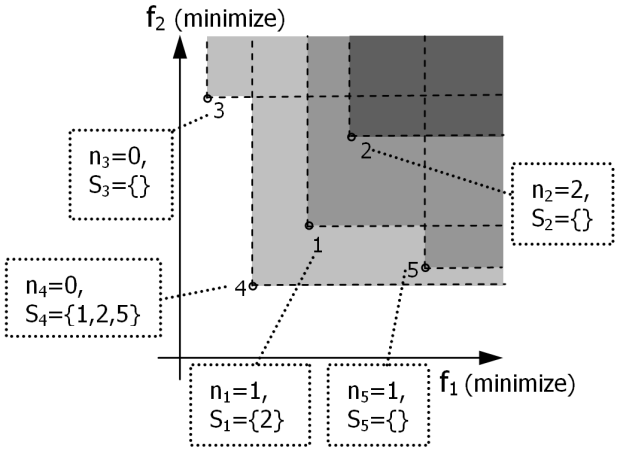


Figure 3.5: Non-dominated sorting in the NSGA-II algorithm.

The preservation of well-distributed (diverse) nondominated solutions is maintained by the crowding distance operator. The crowding distance, which is calculated frontwise, is a measure of how close an individual is to its neighbors. In order to get an estimate of the density of solutions surrounding a particular solution in the population, the average distance between two points on either side of the solution along each of the objectives is calculated. This quantity d_i serves as an estimate of the perimeter of the cuboid formed by using the nearest neighbors as the vertices. A schematic view of the procedure for the computation of the crowding distance for an i^{th} individual is shown in Fig. 3.6, and formulated as (Deb et al., 2002),

$$d_{I_i^m} = d_{I_i^m} + \frac{f_m^{(I_{i+1}^m)} - f_m^{(I_{i-1}^m)}}{f_m^{max} - f_m^{min}}, \qquad for\ i = 2, ..., l-1 \tag{3.11}$$

where I(i) is the solution index of the i^{th} member in the (worse order) sorted list, m is the number of objectives and l is the number of the individuals in that particular front.

The first and the last individuals are assigned an infinite distance in order to emphasize their fitness, hence preserving them for the next generation. A large average crowding distance will result in a better diversity in the population (Deb et al., 2002).

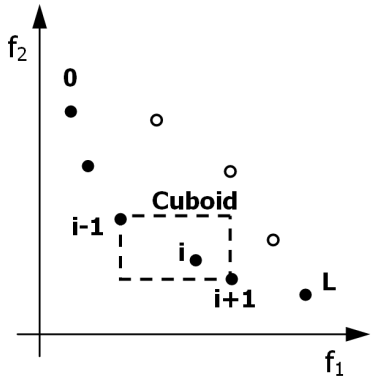


Figure 3.6: The crowding distance calculation (Deb et al., 2002).

After each member is sorted with respect to its own non-domination level (rank), thus is assigned a fittness equal to its rank as well as equiped with the crowding distance information, the next step is to select the good solutions, i.e. to emphasize the non-dominated members, for the next generation. The *binary tournament selection* with a *crowded tournament operator* is applied for this purpose. Shortly, the crowded tournament selection operator compares two solutions, i.e. with respect to their rank and crowding distance, and returns the winner of the tournament (comparison). The solution with the lower rank wins the tournament, but in case of having equal ranks, the solution with a higher crowding distance, i.e. the member placed in a less crowded region, resolves the tie. The working principle of this operator can be shown in a simple case based on the distribution of solutions as presented in Fig. 3.7. There are ten solutions in total and they are sorted with respect to the non-domination criterion for the min.-min. problem, as a result there are three fronts, i.e. the first front is presented by the solutions with circle marker, the second front with square marker and the third one is with diamond marker. These solutions constitute the set of parents. Children (offspring) are determined from this set by the binary tournament selection, meaning that, in particular for this case, five pairs will be chosen twice and each solution (parent) will participate in exactly two tournaments. A random set of pairs is assummed as follows: (3,8), (10,9), (1,5), (2,4), (6,7), (3,10), (1,2), (4,6), (5,7), (8,9). In the first tournament, solution 3 belongs to the first front and solution 8 belongs to the second front, thus solution 3 wins the tournament since it has a lower rank (it belongs to a better front). In the next tournament, both solutions 10 and 9 belong to the first front (i.e. they have the same rank), hence in oder to resolve this tie, their crowding distance values will be checked. Since solution 9 is the last member of the same front, i.e. the crowding distance is infinite, and solution 10 has a crowding distance of a finite value, solution 9 is the winner of this tournament. Performing other tournaments will result in the following mating pool: {3, 9, 5, 4, 7, 3, 1, 4, 7, 9}, which shows that solutions 3, 4, 7 and 9 are selected twice, meaning that members with lower rank are emphasized by the selection operator promoting convergence towards the Pareto-front.

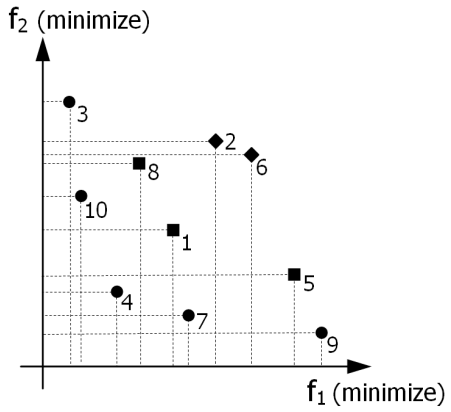


Figure 3.7: Ten solutions, i.e. non-dominated members presented with circles, the second front with squares and the third front with diamond, involved in the crowded tournament selection.

In classical genetic algorithms (Goldberg, 1989), the internal encoding is based on binary coding that each design variable is represented as a binary string, i.e. 0100, 1101101, etc., where the length of the string depends on the number of allowed values for the variable. The classical (single-point) cross-over is a well-known genetic algorithm operator in which two parents (individuals from the mating pool) are randomly chosen and some portions of these binary strings are exchanged between each other. In single-point cross-over, the crossing site is randomly chosen and all bits on the right side of the crossing site are swapped as shown "schematically" in Fig. 3.8.

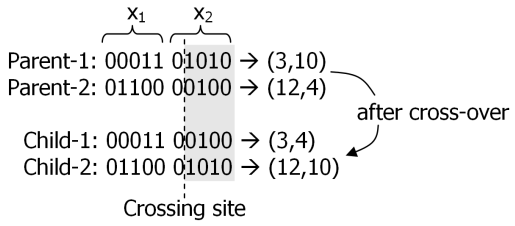


Figure 3.8: Schematic procedure of single-point crossover.

This procedure is also shown for the same example given above (Fig. 3.8) in the 2D parameter space in Fig. 3.9, assuming a representation of each variable with 5 bits (i.e. total string length of 10) considering an interval of [0,31] (the change in each bit corresponds to an increment of 1).

The *mutation* operator is responsible for perturbing the binary string, i.e. changing 1 to 0 or vice versa, with a given *DNA String Mutation Ratio* (probability). The mutation operator can be said to be responsible of the local search as well as maintaining the diversity in the population (Deb, 2006).

However, some potential difficulties arise in case of using binary coded GAs in the continuous parameter space, such as Hamming cliffs as well as the demand for a higher string length depending on the prefered or required accuracy (which is unknown beforehand for the latter case). Hamming cliffs are observed while evolving from one solution

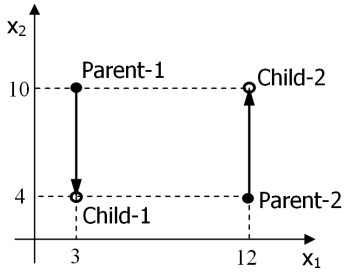


Figure 3.9: Representation of single-point crossover in 2D parameter space.

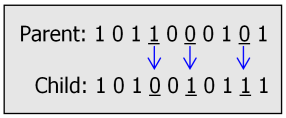


Figure 3.10: Schematic procedure of mutation for a given DNA string mutation ratio, p_m=30%.

(e.g. encoded as string 0111111) to a neighboring solution (e.g. string 1000000) in real space, which requires change of many bits (the longer the string, the more artificial hindrance is experienced) (Deb, 2001). The increase in the required precision (larger string length) calls for a larger population size (Goldberg et al., 1992) which in turn results in higher computational complexity. For these reasons, real-coded GAs (which do not require any encoding) are implemented in order to satify the need for better performance demands on cross-over and mutation operators in the continuous search space. In the NSGA-II, simulated binary cross-over (SBX) (Deb and Agrawal, 1995; Deb and Kumar, 1995) and polynomial mutation is used. The SBX operator uses two parent solutions as inputs and produces two children as output. As the name indicates, the SBX operator simulates the working principle of the single-point cross-over operator on binary strings. The original work is recommended for further details (Deb and Agrawal, 1995; Deb, 2001).

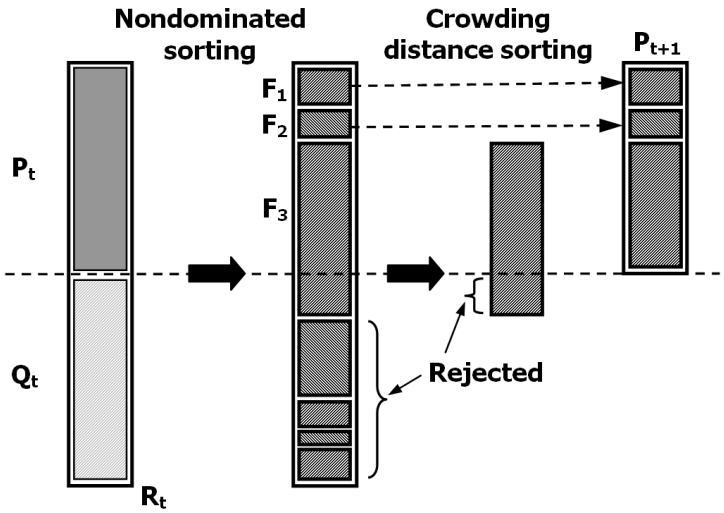


Figure 3.11: Schematic view of the NSGA-II procedure (Deb et al., 2002).

The schematic procedure of the NSGA-II is shown in 3.11. The procedure starts with an initialization of the parent population, P_0, followed by sorting the members of the population based on nondomination into each front, F_i. Each individual is assigned two fitness values, i.e. the one related to its nondomination level (front or rank, i.e. F_i refers to front with rank i in Fig. 3.6) and the crowding distance. Parents are selected by using the binary tournament selection (with crowded tournament operator) based on these two measures. An individual is selected if its rank is less than the other is, since the minimization is assumed, or if its crowding distance is greater than the other is. The selected population generates off-springs, Q_t, by the application of usual genetic operators, i.e. crossover, mutation, etc. (Goldberg, 1989). Since all the parent and the offspring individuals are included in the combined population, R_t, elitism is ensured. In order to reduce the size of the population of the next generation from $2N$ to N, nondomination sorting is applied and the individuals with the relatively lowest possible ranks are accomodated until the size of the population exceeds N. Then the individuals in front F_i are selected based on their crowding distance. This cycle is repeated until the stopping criterion, i.e. the total number of generations, is met. Further details can be found in the original study (Deb et al., 2002).

3.3 Optimization of the thermal models of FSW

3.3.1 Optimization of the analytical thermal model

The optimization case presented here was the first one considered in the Ph.D. project, see `PAPER-I`. The problem is stated as the goal of finding the FSW parameters, i.e. heat input and longitudinal welding speed, which provide a desired average temperature under some process limitations. The problem can be formulated as in Eq. (3.12),

$$\begin{aligned} Minimize : f_1(x) &= (T_{average} - 500)^2 = F^2, \\ subject\ to : 0 &< Q_{tot} \leq 10000\,W, \\ 0 &\leq u_{weld} < 100\,mm/s. \end{aligned} \tag{3.12}$$

where the objective function $\left(F^2\right)$ is the error between the analytically or numerically computed value and the desired average temperature under the tool shoulder. It would be advantageous to obtain an analytical expression for the sensitivity computation based on Eq. (2.6); however a semi-analytical procedure is presented. The necessary Finite Difference calculations were performed for the COMSOL thermal model. Standard MATLAB Optimization toolbox routines are used for the two equivalent optimization problems.

The finite element method gives the user the possibility to define complicated models, including complex geometries and heat source models, etc., while analytical models have some limitations in some aspects of the engineering applications, e.g. FSW modeling. In order to extend the applicability of the analytical formulation, a novel procedure is introduced. The power of the moving line-type heat source is controlled by evaluating the temperatures in a circular region surrounding the heat source and this is used as a process control parameter. The corresponding bound constrained problem is solved using a subspace trust region method based on the reflective Newton method (The Math Works Inc., 2006).

3.3.1.1 Derivation of the Closed Form Solution

A moving heat source representing the tool shoulder is described by a circle in an infinitely extended plate, as mentioned before and shown in Fig. 3.12. In order to obtain a closed form solution for the average temperature under the tool shoulder, the temperature field defined in Eq. (2.6) should be integrated over the circular area in polar coordinates, as formulated in Eq. (3.13).

$$\begin{aligned} T_{average} &= \frac{\int_A T\, dA}{\int_A dA} = \frac{\int_{r=0}^{R_{sh}} \int_{\theta=0}^{2\pi} T(r,\theta) r d\theta\, dr}{A_{circle}}, \\ &= \frac{Q_{tot}}{2\pi^2 k d R_{sh}^2} \int_{r=0}^{R_{sh}} \int_{\theta=0}^{2\pi} \frac{r K_0(wr)}{e^{wr\cos(\theta)}} d\theta\, dr + T_0 \end{aligned} \tag{3.13}$$

The integral in Eq. (3.13) cannot be evaluated analytically due to the singularity of the K_0 function in the formulation of the line-type heat source. Thus, a numerical integration scheme will be defined in the following section.

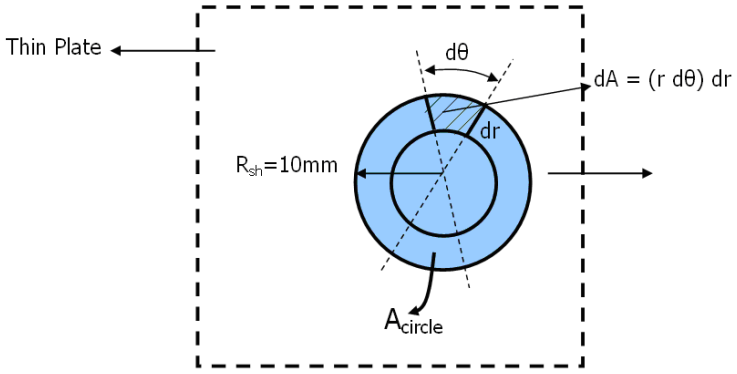


Figure 3.12: A moving heat source in an infinite plate defined in polar coordinates [ref. PAPER-I] (Tutum et al., 2007).

3.3.1.2 Numerical Computation of the Average Temperature

In this section, the numerical integration scheme for evaluating the average temperature under the tool shoulder is formulated. The line-type heat source is positioned in the center of the circle, defined as (0,0) in Cartesian coordinates, in Fig. 3.13.

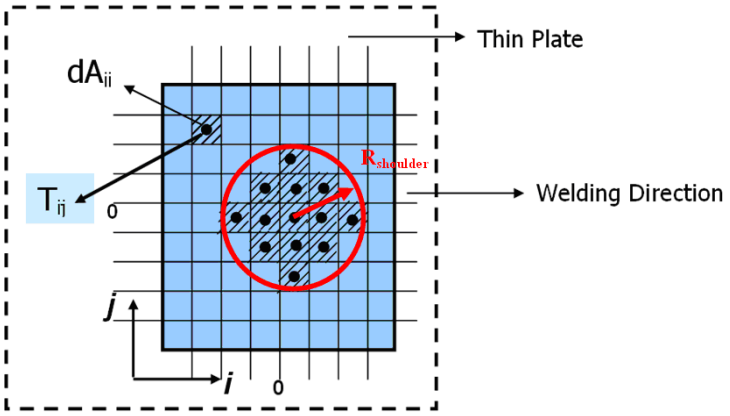


Figure 3.13: A moving heat source in an infinite plate defined in Cartesian coordinates [ref. PAPER-I] (Tutum et al., 2007).

Eq. (3.14) defines the average temperature, similar to Eq. (3.13), in a two dimensional (indices) summation form. Since all cells are equal in size, the denominator of Eq. (3.14) is reduced to the number of cells. The temperature values evaluated in the grid points (centers of the cells) placed in the circular area are summed easily, as defined individually in Eq. (2.6).

$$T_{average} = \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} T_{ij}(x_i, y_j)\, dA_{ij}}{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} dA_{ij}} = \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} T_{ij}}{n_{cell}} \tag{3.14}$$

The grid size obviously influences the accuracy of the average temperature expression defined in Eq. (3.14). Thus, to ensure independence of this, a convergence test is performed by using the same heat input value for changing grid density along the diameter of the circular region representing the tool shoulder. Fig. 3.14 shows that the convergence is achieved for a low number of increments.

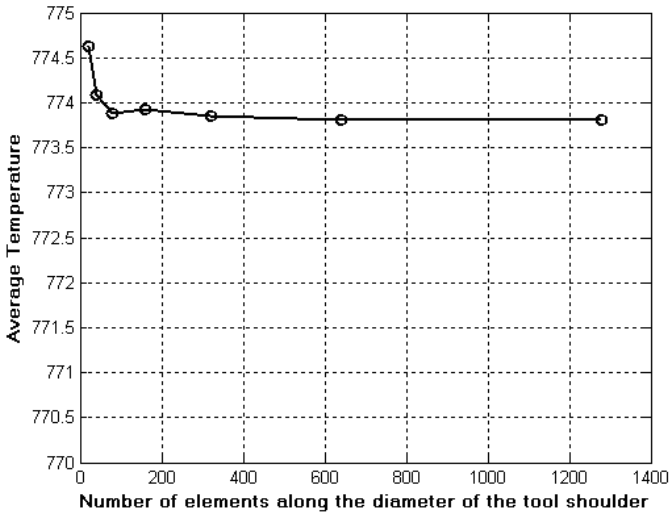


Figure 3.14: Convergence test plot of the average temperature expression for different grid size [ref. PAPER-I] (Tutum et al., 2007).

3.3.1.3 Semi-Analytical Sensitivities

The possibility to use analytical sensitivities provides substantial benefits regarding savings in computational cost. It is found, cf. Section 3.3.1.1, that it is not possible to obtain a closed form solution for the average temperature criterion since integration of the K_0 function causes difficulties. In this section, it is shown that the semi-analytical method can be used to reduce computational cost as much as possible. The sensitivities of the objective function with respect to the welding parameters, i.e. heat input and welding speed, are defined in Eqs. (3.15) and (3.16), respectively.

$$\frac{d(F^2)}{dQ_{tot}} = 2F\left(\frac{C}{\pi R^2}\int_0^R\int_0^{2\pi}\frac{rK_0(wr)}{e^{wr\cos(\theta)}}d\theta dr\right) = \frac{2F(F+T_1)}{q} \tag{3.15}$$

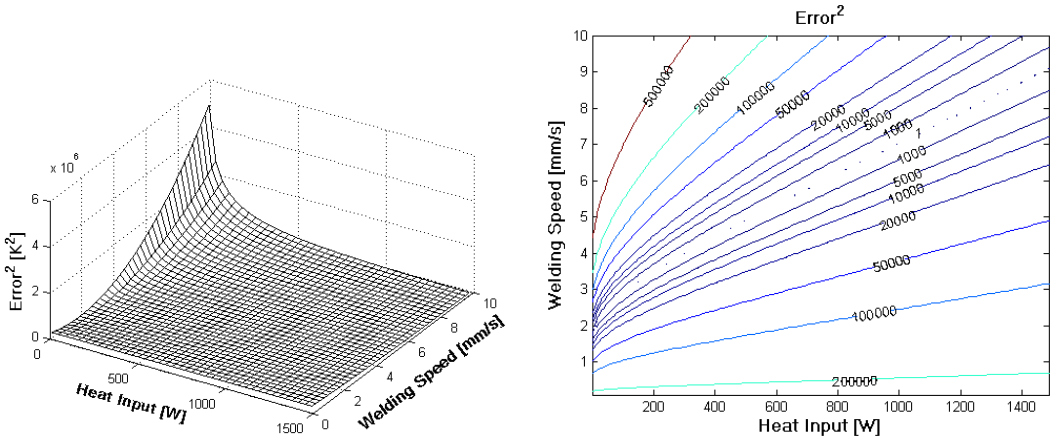
$$\frac{d(F^2)}{du_{weld}} = 2F\,C_2\int_0^R\int_0^{2\pi}\left[\frac{r^2}{e^{wr\cos(\theta)}}\left(K_1(wr)+K_0(wr)\cos(\theta)\right)\right]d\theta dr \tag{3.16}$$

Here F^2 denotes the objective function, T_1=480K, $C=\frac{1}{2\pi kd}$ and $C_2=\frac{-C\,Q_{tot}}{a}$ which are used to simplify the expressions and K_1 is the Modified Bessel Function of the second kind and the first order. The first sensitivity of the objective function with respect to the heat input is relatively easy to compute as compared to the second, since all variables in Eq. (3.15) are already evaluated while computing the objective value as given in Eq. (3.12), using the incremental procedure. The second sensitivity expression is formulated in an analytical way, but again evaluated using the incremental procedure.

3.3.2 Optimization of the numerical (Eulerian) thermal model

In this section, a steady state Eulerian heat flow model implemented in COMSOL is used to control the average temperature under the tool shoulder. In order to find the optimum welding parameters, i.e. the heat input and welding speed for the bounded constrained problem as defined in Eq. (3.12), the Sequential Quadratic Method (SQP) with the BFGS line search algorithm is used (Vanderplaats, 2005). Finite difference calculations are performed in order to obtain the sensitivity information.

The objective function field for the corresponding design variable field is shown in Fig. 3.15a. It is obvious that multiple solutions are found since the optimization problem has only simple boundaries without any equality or inequality constraints. The resulting optimum curve which is labeled as the contour line equal to 1 is lying close to the zero plane, as can be seen in Fig. 3.15b.



(a) (b)

Figure 3.15: a) Objective function field ($Error^2$) vs. design variable field, i.e. heat input and welding speed [ref. PAPER-I] (Tutum et al., 2007). b) Contour plot of the corresponding objective function field [ref. PAPER-I] (Tutum et al., 2007).

In order to evaluate the results from a practical viewpoint, the design variables, i.e. the heat input and the welding speed, are considered seperately. Tables 3.1 and 3.2 show the results regarding the specified optimum design variable for the other fixed welding parameter, respectively. Aluminum is used as the workpiece material which is a 3 mm-thick plate. Notice that the analytical heat input solutions are scalable for different plate thicknesses due to the linearity in terms of the heat generation density as given in Eq. (2.6).

Notice that the higher the welding speed is, the higher the error due to the boundary effect. The analytical heat flow model does not see the effects of the boundary conditions because of the infinitely large domain description, while it is not the case for the numerical model having a finite domain. However, it can be concluded that the deviations between the analytical and the numerical models are reasonable expecially for the lower welding speeds, which is also considerable for the FSW process.

Table 3.1: Comparison of the optimum heat input results for given welding speed and relative error in relation to the analytical heat input [ref. PAPER-I] (Tutum et al., 2007)

u_{weld} (mm/s)	$Q_{tot,Analytical}$ (W)	$Q_{tot,COMSOL}$ (W)	$\frac{\lvert\Delta Q_{tot}\rvert}{Q_{tot,Analytical}}$(%)
2	575	585	2
4	780	890	14
6	971	1153	18
8	1160	1404	21
10	1350	1653	22

Table 3.2: Comparison of the optimum welding speed results for given heat input and relative error in relation to the analytical welding speed [ref. PAPER-I] (Tutum et al., 2007)

Q_{tot} (W)	$u_{weld,Analytical}$ (mm/s)	$u_{weld,COMSOL}$ (mm/s)	$\frac{\lvert\Delta u_{weld}\rvert}{u_{weld,Analytical}}$(%)
500	1.36	1.50	10
750	3.69	3.00	18
1000	6.32	4.80	24
1250	8.97	6.80	24
1500	11.56	8.80	24

This section presented the first optimization case considered in the Ph.D. work. Special about this was its relative simplicity being an SOO problem with some analytical information available, thus calling for the probability of using a classical gradient based optimization algorithm. However in the following, some more complex MOO cases based on thermo-mechanical models of FSW, applying genetic algorithms will be presented.

3.4 Optimization of the thermo-mechanical models of FSW

3.4.1 Optimization of the residual stresses and production rate in FSW

In the following sections we present a study of two multi-objective optimization problems in the FSW process. These consider the minimization of the residual stresses in the workpiece together with the maximization of the production efficiency (reducing welding time and wear), respectively. These two objectives are conflicting and techniques to deal with this issue are also considered. The optimization procedure, which includes process integration of the ANSYS software and the optimization algorithm (NSGA-II), is handled by applying modeFRONTIER. The optimization cycle is initiated by creating an initial population of 16 pre-chosen Design-of -Computational-Experiments (DOE), for the considered process variables, i.e. the tool rotational speed (revolutions per minute), n, and the traverse welding speed, uweld. The FSW thermal and mechanical simulations, which are built by using APDL in ANSYS, are coupled in a sequential way by execution of parametric input files in a batch mode. The design variables are updated by the optimization algorithm and are read by the thermal analysis. Then the peak temperature obtained at the end of the welding session is saved in order to be used as a thermal constraint together with the transient temperature field results to yield the thermal strains for the mechanical analysis. This mechanical analysis gives as a result the maximum longitudinal stress value at the middle of the plate in the transverse direction, and this is used as an objective to be minimized. This optimization cycle runs until the stopping criterion, i.e. the total number of generations, is reached. The flow chart of the optimization procedure is shown in Fig. 3.16.

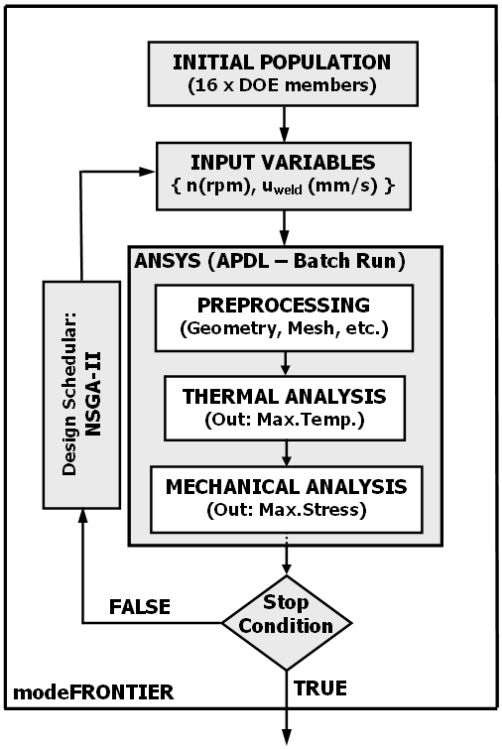


Figure 3.16: Flow chart of the optimization problem [ref. PAPER-III] (Tutum et al., 2009).

Most engineering design activities require a solution of multi-disciplinary and MOO problems which usually deal with conflicting design objectives. As discussed in Section 3.2.2, the consideration of multiple conflicting objectives results in a number of alternative trade-offs, called Pareto-optimal solutions, of which none can be said to be better than any other without any additional information about the problem under consideration. This concept can be formulated as follows

$$(\forall i : f_i(x) \leq f_i(y)) \cap (\exists j : f_j(x) < f_j(y)) \tag{3.17}$$

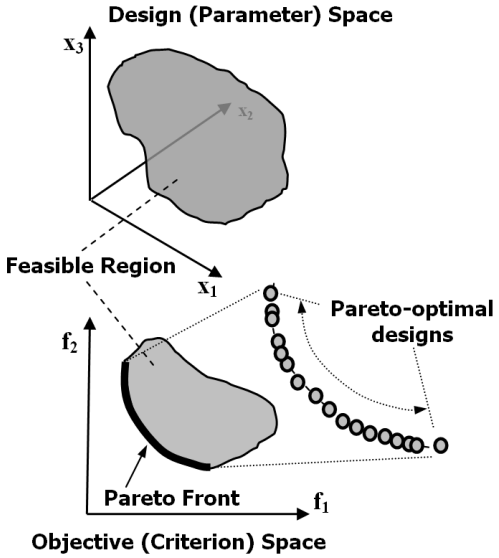


Figure 3.17: Design and objective space associated with multi-objective optimization problem [ref. PAPER-III] (Tutum et al., 2009).

where f_i is the $i-th$ objective function in the feasible region and the minimization of the two-objective problem, which is schematically shown in Figure 3.17, is considered for convenience. In this context, Eq. (3.17) describes that at least in one objective, design x is better than design y, while in others they might be equal; this can also be expressed as the solution x dominates the solution y in the parlance of MOO. Since the concept of domination allows a way to compare solutions with multiple objectives, most MOO algorithms use this domination concept to search for non-dominated solutions, i.e. the ones that constitute the Pareto-front (Deb, 2001, 1999, 2003). Population-based evolutionary algorithms have in the last decade become more and more popular as they effectively can utilize distributed computing facilities. Also, they are typically able to provide a larger spectrum of Pareto-optimal solutions without any additional problem specification. The availability of trade-off solutions, representing varying preference levels between chosen objectives, makes it easier for a decision-maker to choose a particular solution for implementation (Deb, 2001, 1999, 2003). This has been discussed in more detail in Section 3.2.2. In the following sections the two optimization problems for FSW based on thermo-mechanical modeling are presented in detail.

3.4.1.1 Problem Satement-1

The first optimization problem here is stated as the goal of finding the FSW parameters, i.e. tool rotational speed and traverse welding speed, which provide a set of trade-off solutions for the two conflicting objectives. These are minimization of the residual stresses, which are measured at the middle of the plate along the transverse direction, and minimization of the welding time that can also be stated equally as the maximization of the traverse welding speed. The optimization problem is constrained by process-specific thermal constraints, which are given as upper and lower bounds on the peak temperatures in the workpiece. The lower bound of 420 °C on the peak temperature represents the need for easy traversing of the tool, i.e. to limit the tool loads along the weld line by contributing to thermal softening of the workpiece material. The upper bound of 480

°C is defined in order to consider the tool life and the workpiece properties which are affected by hot welding conditions. This constrained multi-objective problem can then be expressed in mathematical terms as

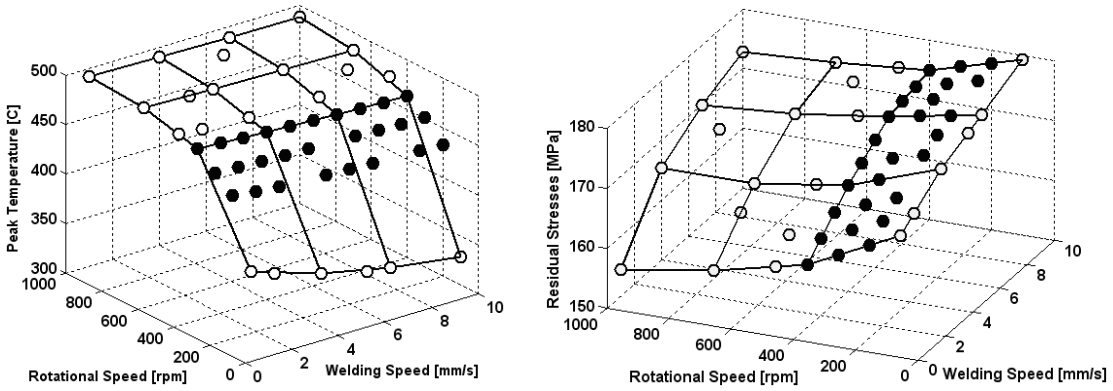
$$\begin{aligned} Minimize : f_1(\mathbf{x}) &= \sigma_{x,max}, \\ Maximize : f_2(\mathbf{x}) &= u_{weld}, \\ subject\ to : g_1(\mathbf{x}) &= 420\,^oC \leq T_{peak}, \\ g_2(\mathbf{x}) &= T_{peak} \leq 480\,^oC, \\ where \quad \mathbf{x} = \begin{Bmatrix} u_{weld} \\ n \end{Bmatrix}, & \begin{cases} u_{weld} = 1, 2, ..., 10\,mm/s, \\ n = 100, 200, ..., 1000\,rpm. \end{cases} \end{aligned} \tag{3.18}$$

where x represents the design variable vector, i.e. u_{weld}, the traverse welding speed that is changing from 1 mm/s to 10 mm/s in 1 mm/s increments and n, the tool rotational speed which varies from 100 rpm to 1000 rpm in 100 rpm increments (this results in 10 discrete values in each design variable), $\sigma_{x,max}$ defines the peak longitudinal stress, and T_{peak} is the peak temperature in the workpiece.

3.4.1.2 Results of Problem Satement-1

The solution of the first optimization case, which is defined in section 3.4.1.1, is presented in both the design and the criterion space in the following figures. Some of the designs out of a total of 320 designs (16 designs x 20 generations) are overlapping due to relatively coarse discretization of the chosen design variables and also the selection operator. This lies in the nature of the genetic algorithm that implies the survival of some designs without evolution. Fig. 3.18a and 3.18b represent feasible and unfeasible designs with dark and fair colors, respectively; constituting 49 different design points out of 100 (we use a 10x10 discretization). The surface in each figure, i.e. the peak temperature and the peak residual stress, is constructed by 16 DOE points which are evaluated as an initial population for the NSGA-II. It can be clearly seen from these figures that the feasible region, which can also be called the robust process parameter region in this case, is defined by n-values in the interval between 200 and 400 rpms. Due to the random evolution strategy of EAs, some of the feasible solutions (4, 6, 8 mm/s with 200, 300 and 400 rpms), are missing but the quality of these can be estimated from the surrounding solutions which are positioned on a linearly varying region. Although the coarse discretization results in a few missing solutions, it is advantageous for having an overall idea regarding the optimal feasible process frame with a moderate computational cost. In addition to this, the purpose of this study is to focus on the optimization methodology to find and discuss alternative trade-off solutions for minimization of welding residual stresses, i.e. not to conclude precise values. Having such solutions provides the engineer or manufacturer practical insight about the relationship among process variables corresponding to the Pareto-optimal solutions.

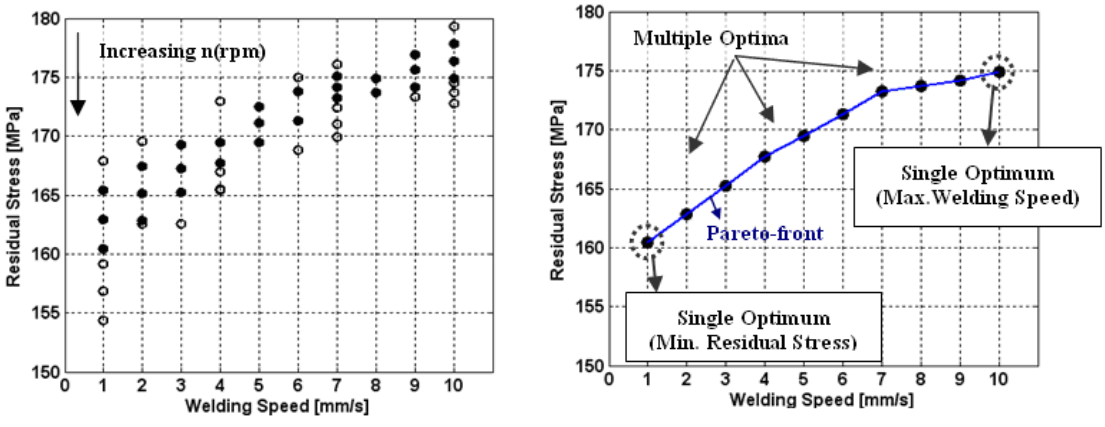
The objective space that is constructed by minimization of the residual stresses and maximization of the welding speed is shown in Fig. 3.19a. Most of the designs lie close to the Pareto-front, which is shown in Figure 10(b), due to the low sensitivity of the n parameter defined in the underlying thermal model, i.e. the TPM model for a given welding speed on the residual stresses. If the minimum-residual stress solutions is considered to be of $100 - 0\%$ importance for product performance and process efficiency,



(a) (b)

Figure 3.18: (a)Peak Temperature vs. design variables [ref. PAPER-III] (Tutum et al., 2009). (b)Residual stress vs. design variables [ref. PAPER-III] (Tutum et al., 2009).

respectively, then the maximum process efficiency, i.e. welding speed or in other words the production rate solution is of $0-100\%$ importance, thus the combination of 1 mm/s and 400 rpm would be chosen. If the other extreme solution, i.e. maximum welding speed, is considered, the combination of 10 mm/s and 400 rpm would be preferred. In case of looking for the $70-30\%$ trade-off solution for the same objectives, it is not clear to estimate the optimal combination of the process variables. Due to this desired preference in objectives, the solution would be expected to be more similar to the minimum residual stress solution than the maximum production rate. It is important to note that there are some different solutions satisfying such trade-off, but there is only one which is the optimum, i.e. 4 mm/s and 400 rpm in this case. In other words, that solution makes the optimum trade-off, meaning that for a particular amount of sacrifice in one objective, the optimum solution will correspond to the maximum gain in the other objective (Deb, 2003).



(a) (b)

Figure 3.19: (a)Objective Space of the solution [ref. PAPER-III] (Tutum et al., 2009). (b)Pareto Set of the solution [ref. PAPER-III] (Tutum et al., 2009).

All Pareto-optimal solutions correspond to the maximum feasible rotational speed

of 400 rpm while having different welding speeds varying from 1 mm/s to 10 mm/s. The Pareto-front shown in Fig. 3.19b gives an idea of ranking the alternative trade-off solutions depending on the available working conditions. If a manufacturer is able to use a standard milling machine instead of an advanced FSW machine and can afford using simple tool designs with low welding speed, he would probably not dare to go from 1 to 7 mm/s in welding speed because the residual stresses yielded per unit increment in welding speed would cost higher comparing to those at higher welding speeds. The amount of sacrifice of the manufacturer relatively depends on the welding speed while one can keep the rotation speed between 200 and 400 rpms.

3.4.1.3 Problem Satement-2

The second optimization problem, which is a slightly modified version of the first case, Eq. (3.18), considers the minimization of the so-called wear path of an arbitrary point on the tool shoulder besides taking the minimization of the residual stresses evolved on the workpiece into account, simultaneously, Eq. (3.19).

$$\begin{aligned} &Minimize: f_1(\mathbf{x}) = \sigma_{x,max}, \\ &Minimize: f_2(\mathbf{x}) = \omega/u_{weld}, \\ &subject\ to: g_1(\mathbf{x}) = 420\,^oC \leq T_{peak}, \\ &\qquad\qquad g_2(\mathbf{x}) = T_{peak} \leq 480\,^oC, \\ &where \quad \mathbf{x} = \begin{Bmatrix} u_{weld} \\ n \end{Bmatrix}, \begin{cases} u_{weld} = 1, 2, ..., 10\, mm/s, \\ n = 100, 200, ..., 1000\, rpm. \end{cases} \end{aligned} \tag{3.19}$$

This wear path at the radius r at any point on the tool/workpiece interface can be approximated by the following expression,

$$L_{path} = v_c\, t_{weld} = \omega r t_{weld} = \omega r \frac{L_{weld}}{u_{weld}} = (rL_{weld})\, \frac{\omega}{u_{weld}} = C \frac{\omega}{u_{weld}} \tag{3.20}$$

where subscript c denotes circumference and C is a parameter that will remain constant for the point under consideration. Minimization of the ratio of rotational speed divided by welding speed corresponds to maximizing the tool advance per revolution. Other studies as explained below, have found weld property variations as a function of either welding speed divided by rotational speed, i.e. the weld pitch, rotational speed divided by welding speed or pseudo heat index. However none of these are based on an argumentation addressing the wear criterion as presented in this paper, neither are they used for optimization purposes, but more for process characterization and understanding.

It has been found out that, in general, the specific weld energy (Joules/meter) correlates with the weld pitch, i.e. the energy per unit length of weld increases as the weld pitch decreases (Seidel and Reynolds, 2001). The peak temperature in the weld zone increases with increasing the ratio of tool rotation rate/traverse speed (Hashimoto et al., 1999). The recrystallized grain size is reduced by decreasing the ratio of tool rotation rate/traverse speed which results in an increase in both degree of deformation and peak temperature of the thermal cycle (Ma et al., 2002). The pseudo-heat index (ω^2/u_{weld}) was used as part of a generalized relationship in order to find the peak temperature for

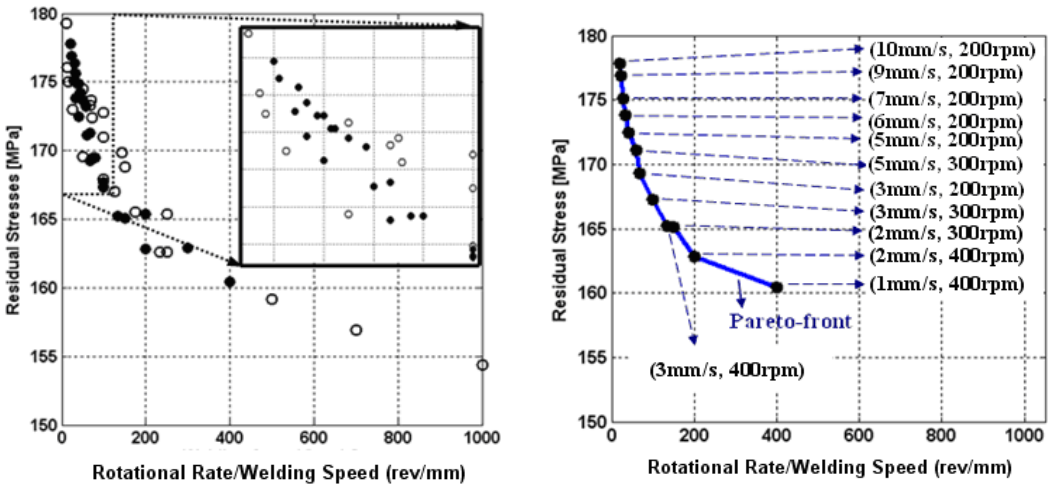
several aluminum alloys (Arbegast and Hartley, 1998). It was found that increasing the ratio of rotational speed/traverse speed (ω/u_{weld}) leads to a slight decrease in yield and ultimate strength of the stir zone and transitional zone. It was also observed that increasing this ratio increases the weld nugget size and decreases the incomplete root penetration (Gharacheh et al., 2006).

3.4.1.4 Results of Problem Satement-2

The second optimization case defined in section 3.4.1.3 is presented below in objective space only since the design domain is kept constant while the solutions of the first optimization case are rearranged by taking the tool rotational speed into account as shown in Fig. 3.20a and 3.20b.

From the Pareto front in Fig. 3.20b it is seen that at high welding speeds the second objective, i.e. minimization of the wear path becomes dominant, whereas for low welding speeds the first objective of minimum residual stress is dominant.

It is also interesting to note that the wear path criterion divides the parameter space into two intervals based on whether the welding speed is higher or lower than 5 mm/s. Thus, it is seen on Fig. 3.20b that for welding speeds above this value, the Pareto solutions all have a rotational speed of 200 rpm, whereas for welding speeds below this value the rotational speed varies between 200 and 400 rpm. So, as compared to the previously found solutions, which are listed in section 3.4.1.4, the rotational speed is reduced from 400 rpm to 200 rpm at high welding speeds resulting in a slightly higher sacrifice in the first objective, i.e. minimization of the residual stresses.



(a) (b)

Figure 3.20: (a)Objective Space of the solution [ref. `PAPER-III`] (Tutum et al., 2009). (b)Pareto Set of the solution [ref. `PAPER-III`] (Tutum et al., 2009).

The discussions made for the first problem case can be generalized for this case as well. The single optimum solution for the lowest welding speed is kept the same, i.e. being combined with a rotational speed of 400 rpm again, while the single optimum solution for the highest welding speed is rearranged in a way that it is combined with the lowest feasible rotational speed, i.e. 200 rpm.

Chapter 4

Summary of Appended Papers

4.1 PAPER-I

C. C. Tutum, H. N. B. Schmidt, J. H. Hattel and M. P. Bendsøe, *"Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization"*, 7th World Congress on Structural and Multidisciplinary Optimization, COEX, Seoul, May 2007, Korea.

In this paper, the classical analytical solutions to the heat conduction equation in a Eulerian frame as given by Rosenthal is programmed together with the corresponding numerical thermal model in a Eulerian frame in the FE-software COMSOL Multiphysics and good agreement is achieved as expected. This is then combined with an optimization procedure in which the objective is to achieve an average temperature of 500 °C under the tool. The 500 °C is chosen as an indication of a suitable temperature level for the welding process to succeed. The design variables for this optimization problem were the welding speed and the heat input and the applied optimization algorithm was Sequential Quadratic Programming (SQP). Closed form expressions for the analytical sensitivities were derived.

4.2 PAPER-II

C. C. Tutum, H. N. B. Schmidt and J. H. Hattel, *"Assessment of Benchmark Cases for Modeling of Residual Stresses and Distortions in Friction Stir Welding"*, 7th International Symposium on Friction Stir Welding, TWI, 2008, Awaji Island, Japan.

This paper presents a thermomechanical model for FSW, which is implemented in ABAQUS and used to analyze several benchmark cases for the formation of residual stresses in FSW in which the effects of the clamping conditions and a temperature dependent yield stress are investigated. The ABAQUS model is also used for a more realistic case in which the FSW of a stringer to a plate is modeled. The thermally induced residual stresses are calculated and subsequently the releasing and loading steps of the stringer structure are modeled also. The results show that for the considered case, the loading of the structure actually means a small reduction of the maximal longitudinal stresses as compared to the residual stresses themselves.

4.3 PAPER-III

C. C. Tutum, H. N. B. Schmidt J. H. Hattel and M. P. Bendsøe, *"A preliminary study on optimization of residual stresses in friction stir welding"*, Materials and Manufacturing Processes, 2009 (In press).

This paper presents a thermomechanical model for FSW, which is implemented in ANSYS and used for a multiobjective optimization case in which the objectives are minimizing the maximal residual stresses as well as maximizing the advancement per revolution which is given by the welding speed divided by the rotational speed. The design variables are the welding speed and the rotational speed. This represents an example of a multi-objective case in which the design variables are also part of one of the objectives. The results first of all show that the NSGA-II algorithm works well together with the thermomechanical model in ANSYS and more specifically they also reveal that increasing the welding speed results in higher residual stresses for the case at hand. This might however, not be the case for other constraining conditions.

4.4 PAPER-IV

J. H. Hattel, H. N. B. Schmidt and C. C. Tutum, *"Thermomechanical Modelling of Friction Stir Welding"*, Proc. 8th International Conference on Trends in Welding Research, Pine Mountain, Atlanta, USA, 2008 (Invited keynote lecture).

In this paper, an overview of the thermomechanical modelling of the FSW process is given. The different types of models are classified based on the model discipline applied, i.e. numerical heat transfer models, computational solid mechanics models and computational fluid dynamics models rather than from their purpose, i.e. thermal models, flow models and residual stress models. The formulation of all three mentioned modelling approaches are presented and discussed together with selected modelling results including prediction of material flow during welding, prediction of heat generation with the thermal-pseudo mechanical model as well as residual stress and deformation analysis combined with in-service loads.

4.5 TECHNICAL REPORT-I

C. C. Tutum and J. H. Hattel, *"A benchmark study for unconstrained multi-objective problems"*, September 2009.

In this report several well-known unconstrained Multi Objective Optimization (MOO) benchmark problems are investigated in order to analyze some of the difficulties that a Multi Objective Evolutionary Algorithm (MOEA) may have to tackle. Three elitist algorithms, i.e. the MOGA-II, the NSGA-II and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB with two versions including an extra Pareto-optimal set achieved for post-processing purposes), are employed for this purpose.

4.6 TECHNICAL REPORT-II

C. C. Tutum and J. H. Hattel, *"A benchmark study for constrained multi-objective problems"*, October 2009.

Similar to the `TECHNICAL REPORT-I` in its structure, however focusing on constrained multi-objective optimization benchmark problems. The results (especially for the constrained case) show that the cNSGA-II shows a good performance for having both a converged and a well-spread distribution of the Pareto-optimal set with a short computational cost.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The work presented in this thesis focuses on two subjects; first, the thermo-mechanical modeling of the FSW process, and secondly, the optimization of the thermal and thermo-mechanical aspects of the process by means of mathematical programming and evolutionary computation techniques. The main conclusions and research achievements are briefly summarized below.

As a natural thing, the task of modeling the thermo-mechanical aspects of this complex process is initially divided into two steps, i.e. thermal and mechanical modeling. The thermal models are addressed first since they in essence constitute the basis of all the mechanical models, i.e. residual stress models, softening models or integrated models which take the service load into account and damage evolution models.

In the first optimization case, an optimal set of two process parameters, i.e. the welding speed and the heat input, which aims for the process specific goal of obtaining an average temperature distribution under the tool is sought by using both analytical and numerical Eulerian thermal models. These models are combined with a gradient based optimization algorithm, i.e. Sequential Quadratic Programming (SQP) in order to investigate the optimal welding parameters as well as the applicability of these varying-complexity models. It is found that even a relatively simple analytical model can be used to predict the temperature distribution fine enough under some assumptions and limits. Alongside with the optimization study, analytical sensitivities for the analytical model are also obtained together with performing the similar finite difference computations for sensitivities in the numerical model.

Following this, are the thermo-mechanical models implemented in both ABAQUS and ANSYS. Initially, these are used for the analysis of the transient temperature and stress evolutions during welding and subsequent cooling, eventually leading to the residual stress. It is well-known that some heat treated metals, e.g. aerospace aluminum alloys loose hardness and strength in the HAZ when subjected to thermal cycles during welding. This softening behavior was modeled and investigated for different (hot and cold) welding conditions using ANSYS. The characteristics of the hardness profiles for these conditions were in good agreement with the experimental results published by other researchers.

Not common to literature, the subsequent loading situation of a real FSW structure was also taken into account, thus making way for an integrated analysis of the welding process and the loading situation during service of the welded part. This was also the theme for another case, where the predicted stresses were combined with a subsequent uni-axial loading situation in which a damage evolution analysis was carried out in order to predict the final weld's load carrying capacity when subject to tension perpendicular to the weld line. It is found that taking the residual stresses into account increases the maximal stress a little, whereas the ductility of the specimen is reduced since the plastic flow localization happens earlier during the loading process.

The thermo-mechanical models predicting residual stresses are subsequently combined with an evolutionary optimization algorithm (NSGA-II) in the search for the optimal combination of the process parameters that essentially are controllable in practice, namely the welding speed and the rotational speed, which would minimize residual stresses and maximize the welding speed divided by rotational speed (known as advancement per revolution and expressing a desired feature of the process). Being the first result of its kind where optimization of FSW is based on residual stress analyses, the results first of all show that the NSGA-II algorithm (in modeFRONTIER) works well together with the thermo-mechanical model in ANSYS and more specifically they also reveal that increasing the welding speed results in higher residual stresses for the case at hand. This might however, not be the case for other constraining conditions.

This work is backed by some theoretical investigations regarding several well-known unconstrained and constrained multi-objective optimization (MOO) benchmark problems. This was done in order to investigate some of the difficulties that a multi-objective evolutionary algorithm may have to tackle. Specifically, the three elitist algorithms, i.e. the MOGA-II, the NSGA-II (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB) were employed for this purpose. The results (especially for the constrained case) show that the cNSGA-II shows a good performance for having both a converged and a well-spread distribution of the Pareto-optimal set in a short computational cost.

5.2 Future Work

The thermo-mechanical models (more specifically the residual stress models) of the FSW process presented here have performed well for the proposed problems under practical limitations. It would however be interesting to apply more detailed residual stress models such as one combining both a local (for a detailed thermal analysis which takes the material flow into account) and a global (which aimes at capturing the overall behavior of the welded structure, i.e. evolution of residual stresses and distortion) approach on the same model. The more advanced residual stress models could involve more complex material modelling approaches in which the yield stress is not only a function of the temperature, but also a function of the softening model which simulates the evolution of the material hardness.

Although the performance of the cNSGA-II algorithm is promising for the chosen benchmark problems, it could be nice to test its performance on the numerical FSW models (e.g. residual stress models) where the exact Pareto-optimal front is not known.

Moreover, in order to improve the performance (the convergence towards the Pareto-optimal region) of the algorithm, a hybrid evolutionary approach which combines the cNSGA-II with a local search technique, e.g. a gradient-based optimization algorithm (SQP), could be very interesting. Besides FSW applications, a wide variety of manufacturing processes, such as casting or glass moulding are other potential application areas for the developed algorithm.

In order to make the number of function (design) evaluations as few as possible, especially for these highly complex thermo-mechanical models of FSW, there is a substantial need for the *meta-modeling* (approximation) techniques. These techniques are used to build relatively simple ("coarse"), thus computationally inexpensive models (surrogate models) instead of more-complicated, high-fidelity models, at the same time satisfying a good correspondance between these two. For example, response surface models (RSMs) or Design-of-Experiments (DOE) could be used to overcome those difficulties related to computational resources. However, application of these methods could be challenging in multi-objective optimization problems as compared to single-objective optimization problems.

Besides those techniques mentioned above, distributed computing capabilities could be investigated. Due to their implicit or inherent parallelism, evolutionary algorithms are very suitable for different types of parallel programming environments, such as massage passing (MPI) or shared memory (e.g. OpenMP) applications. Due to its high scaling, distributed computing using MPI would be a natural choice. However, today, since multi-core architectures are being served almost as a standard with laptops, the shared memory applications (which allow the programmer update his/her original sequential-program to parallel-version without any substantial affords) are quite tempting as well.

Bibliography

Ansys Inc. (2007). `http://www.ansys.com`. (version 11.0).

Arbegast, W. and Hartley, P. (1998). Friction Stir Weld Technology Development at Lockheed Martin MichoudSpace System - An Overview. USA. 5th International Conference on Trends in Welding Research.

Arora, J. (2004). *Introduction to Optimum Design*. Elsevier Inc.

Bellini, A., Hattel, J., and Thorborg, J. (2006). Thermo-mechanical modelling of aluminium cast parts during solution treatment. *Modelling and Simulation in Materials Science and Engineering*, 14:677–688.

Belytschko, T., Liu, W., and Moran, B. (2000). *Nonlinear Finite Elements for Continua and Structures*. John Wiley and Sons, Ltd.

Bendsøe, M. and Sigmund, O. (2003). *Topology Optimization - Theory, Methods and Application*. Springer-Verlag.

Betounes, D. (1998). *Partial differential equations for computational science: with Maple and vector analysis*. Springer-Verlag New York, Inc.

Bhadeshia, H. (2003). Friction Stir Welding. `http://www.msm.cam.ac.uk/phase-trans/2003/FSW/aaa.html`.

Bhide, B., Michaleris, P., Posoda, M., and DeLoach, J. (2006). Comparison of buckling distortion propensity for SAW, GMAW, and FSW. *Welding Journal*, pages 189–195.

Bogomolny, M. and Bendsøe, M. (2007). Some Optimal Design Aspects of Resistance Welding. pages 2687–2695. 7th World Congress on Structural and Multidisciplinary Optimization.

Boley, B. and Weiner, J. (1960). *Theory of Thermal Stresses*. Dover Publications, Inc.

Branke, J., Deb, K., Miettinen, K., and Slowiński, R. (2008). *Multiobjective Optimization - Interactive and Evolutionary Approaches*. Springer-Verlag.

Bussu, G. and Irving, P. (2003). The role of residual stress and heat affected zone properties on fatigue crack propagation in friction stir welded 2024-T351 aluminum joints. *International Journal of Fatigue*, 25:77–88.

Carslaw, H. and Jaeger, J. (1959). *Conduction of heat in solids*. Oxford University Press, USA.

Chao, Y. and Qi, X. (1998). Thermal and Thermo-Mechanical Modeling of Friction Stir Welding of Aluminum Alloy 6061-T6. *Computers and Structures*, 7:215–233.

Chu, C. and Needleman, A. (1980). Void nucleation effects in biaxially stretched sheets. *Journal of Engineering Materials and Technology*, 102:249–256.

COMSOL AB (2007). `http://www.comsol.com`. (version 3.3a).

Cook, R., Malkus, D., Plesha, M., and Witt, R. (2001). *Concepts and Applications of Finite Element Analysis*. John Wiley and Sons, Ltd.

Corne, D., Knowles, J., and Oates, M. (2000). The pareto envelope-based selection algorithm for multiobjective optimization. pages 839–848. 6th International Conference on Parallel Problem Solving from Nature.

Crisfield, M. (1991). *Non-linear Finite Element Analysis of Solids and Structures - Volume 1: Essentials*. John Wiley and Sons, Ltd.

Crumbach, M., Goerdeler, M., Gottstein, G., Neumann, L., Aretz, H., and Kopp, R. (2004). Through-process texture modelling of aluminium alloys. *Modelling and Simulation in Materials Science and Engineering*, 12:S1–S18.

Das, I. and Dennis, J. (1997). A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69.

Das, I. and Dennis, J. (1998). Normal-Boundary Intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8:631–657.

Dassault Systèmes (2007). `http://www.simulia.com`. (version 6.7-1).

De Jong, K. (2006). *Evolutionary Computation: A Unified Approach*. The MIT Press.

Deb, K. (1999). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7:205–230.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons, Ltd.

Deb, K. (2003). Unveiling innovative design principles by means of multiple conflicting objectives. *Engineering Optimization*, 35:445–470.

Deb, K. (2006). *Optimization for Engineering Design: Algorithms and Examples*. Prentice-Hall of India Pvt. Ltd.

Deb, K. and Agrawal, R. (1995). Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000a). A fast and elitist multiobjective genetic algorithm: NSGA-II. Technical report, Indian Institute of Technology, Kanpur: Kanpur Genetic Algorithms Laboratory (KanGAL).

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000b). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. pages 849–858. Parallel Problem Solving from Nature VI (PPSN-VI).

Deb, K. and Goel, T. (2001). A hybrid multi-objective evolutionary approach to engineering shape design. pages 385–399. Evolutionary Multi-Criterion Optimization. First International Conference, EMO 2001.

Deb, K. and Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multi-modal and multi-objective problems. *Complex Systems*, 9:431–454.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197.

Delahaye, J.-P. (2006). The science behind Sudoku. *Scientific American*, pages 80–87.

Domkin, K. (2005). *Constitutive models based on dislocation density - Formulation and implementation into finite element codes.* PhD thesis, Luleå University of Technology.

Dong, P., Hong, J., and Rogers, P. (1998). Analysis of residual stresses in Al-Li repair welds and mitigation techniques. *Welding Journal*, 77:439–445.

ESTECO s.r.l. (2007). `http://www.esteco.com`. (version v4.0).

Feng, Z. (2005). *Processes and mechanisms of welding residual stress and distortion.* Woodhead Publishing, Ltd.

Fogel, L.-J. (1999). *Intelligence Through Simulated Evolution: Forty Years of Evolutionary Programming.* John Wiley and Sons, Inc.

Fonda, R. and Lambrakos, S. (2002). Analysis of friction stir welds using an inverse problem approach. *Science and Technology of Welding and Joining*, 7:177–181.

Fonseca, C. and Fleming, P. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. 5th International Conference on Genetic Algorithms, USA.

Fratini, L. and Corona, V. (2007). Friction stir welding lap joint resistance optimization through gradient techniques. *Journal of Manufacturing Science and Engineering*, 129:985–990.

Frigaard, O. and Grong Ø, M. O. (2001). A process model for friction stir welding of age hardening aluminum alloys. *Metallurgical and Materials Transactions - Series A*, 32:1189–1200.

Gandin, C. A., Bréchet, Y., Rappaz, M., Canova, G., Ashby, M., and Shercliff, H. (2002). Modelling of solidification and heat treatment for the prediction of yield stress of cast alloys. *Acta Materialia*, 50:901–927.

Geem, Z. (2007). Harmony Search Algorithm for Solving Sudoku. *Lecture Notes in Artificial Intelligence*, 4692:371–378.

Geem, Z., Kim, J., and Loganathan, G. (2001). A New Heuristic Optimization Algorithm: Harmony Search. *Simulation*, 76:60–68.

Gharacheh, M., Kokabi, A., Daneshi, G., Shalchi, B., and Sarrafi, R. (2006). The influence of the ratio of "rotational speed/traverse speed" (ω/v) on mechanical properties of AZ31 friction stir welds. *Journal of Machine Tools & Manufacture*, 46:1983–1987.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning.* Addison Wesley Longman, Inc.

Goldberg, D., Deb, K., and Clark, J. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, pages 333–362.

Goldberg, D. and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. pages 41–49. First International Conference on Genetic Algorithms and Their Applications.

Gould, J. and Feng, Z. (1998). Heat flow model for friction stir welding of aluminum alloys. *Journal of Materials Processing & Manufacturing Science*, 7:185–194.

Grosan, C. and Abraham, A. (2007). Hybrid Evolutionary Algorithms: Methodologies, Architectures, and Reviews. *Studies in Computational Intelligence*, 75:1–17.

Gurson, A. (1977). Continuum theory of ductile rupture by void nucleation and growth. I. Yield criteria and flow rules for porous ductile media. *Transactions of the ASME, Journal of Engineering Materials and Technology*, 99:2–15.

Haftka, R. and Gürdal (1992). *Elements of Structural Optimization.* Kluwer Academic Publishers.

Hashimoto, T., Jyogan, S., Nakata, K., Kim, Y., and Ushio, M. (1999). FSW joints of high strength aluminum alloy. USA. 1st International Symposium on Friction Stir Welding.

Hattel, J. (2008). Integrated modelling in materials and process technology. *Materials Science and Technology*, 24:137–148.

Hattel, J., Schmidt, H., and Tutum, C. (2008). Thermomechanical modelling of Friction Stir Welding. USA. 8th International Conference on Trends in Welding Research Conference, ASM.

Herdy, M. (1991). Application of the evolutionsstrategie to discrete optimization problems. Parallel Problem Solving from Nature. 1st Workshop, PPSN 1 Proceedings.

Hooke, R. and Jeeves, T. (1961). Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 8:212–229.

Horn, J., Nafpliotis, N., and Goldberg, D. (1994). A niched Pareto genetic algorithm for multiobjective optimization. pages 82–87. First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence.

Huiskes, R., Weinans, H., Grootenboer, H., Dalstra, M., Fudala, B., and Slooff, T. (1987). Adaptive bone-remodeling theory applied to prosthetic-design analysis. *Journal of Biomechanics*, 20:1135–1150.

Ishibuchi, H. and Murata, T. (1998). A multi-objective genetic local search algorithm and its application to flowshop scheduling. pages 392–403. IEEE Transactions on Systems, Man and Cybernetics, Part C.

Jahazi, M., Dubourg, L., and Cao, X. (2008). Friction Stir Welding of Aerospace Alloys. `http://www.smecanada.ca/montreal/presentations/Jahazi-Mohammad-FSW.pdf`. (presentation).

James, M., Hughes, D., Chen, Z., Lombard, H., Hattingh, D., Asquith, D., Yates, J., and Webster, P. (2007). Residual stresses and fatigue performance. *Engineering Failure Analysis*, 14:384–395.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. pages 1942–1948. IEEE International Conference on Neural Networks Proceedings.

Kermanpur, A., Tin, S., Lee, P., and McLean, M. (2004). Integrated modeling for the manufacture of aerospace discs: grain structure evolution. *Minerals, Metals & Mater. Soc, USA*, 56:72–78.

Khan, A. and Huang, S. (1995). *Continuum Theory of Plasticity*. John Wiley and Sons, Ltd.

Knowles, J. and Corne, D. (1998). The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimization. pages 250–259. Parallel Problem Solving from Nature - PPSN V. 5th International Conference.

Kollat, J. and Reed, P. (2005). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, 29:792–807.

Koza, J. (1993). *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press.

Lakshminarayanan, A. and Balasubramanian, V. (2008). Process parameters optimization for friction stir welding of RDE-40 aluminum alloy using Taguchi technique. *Trans. Nonferrous Met. Soc. China*, 18:548–554.

Larsen, A., Bendsoe, M., Hattel, J., and Schmidt, H. (2009). Optimization of friction stir welding using space mapping and manifold mapping an initial study of thermal aspects. *Struct. Multidisc. Optim.*, 38:289–299.

Lindgren, L. (2007). *Computational Welding Mechanics: Thermomechanical and Microstructural Simulations*. Woodhead Publishing, Ltd.

Lombard, H., Hattingh, D., Steuwer, A., and James, M. (2008). Optimising FSW process parameters to minimise defects and maximise fatigue life in 5083-h321 aluminum alloy. *Engineering Fracture Mechanics*, 75:341–354.

Lubliner, J. (1990). *Plasticity Theory*. Macmillan Publishing Company.

Luenberger, D. (2003). *Linear and Nonlinear Programming*. Kluwer Academic Publishers.

Lundbäck, A., Alberg, H., and Henrikson, P. (2005). Simulation and validation of TIG-welding and post weld heat treatment of an inconel 718 plate. Mathematical Modelling of Weld Phenomena 7, Technische Universität, Graz.

Ma, Z., Mishra, R., and Mahoney, M. (2002). Superplastic deformation behaviour of friction stir processed 7075Al alloy. *Acta Materialia*, 50:4419–4430.

Mantere, T. and Koljonen, J. (2008). Solving, rating and generating sudoku puzzles with GA. IEEE Congress on Evolutionary Computation, CEC 2007.

Marie, F. and Allehaux, D. (2006). Development of FSW for optimum run-out performance. 6th International Symposium on Friction Stir Welding, Canada.

Mattheck, C. (1997). *Design in Nature: Learning from Trees*. Springer-Verlag.

Mattheck, C. and Burkhardt, S. (1990). A new method of structural shape optimization based on biological growth. *International Journal of Fatigue*, 12:185–190.

McClure, J., Tang, W., Murr, L., Guo, X., Feng, Z., and Gould, J. (1998). A thermal model of friction stir welding. USA. 5th International Conference on Trends in Welding Research.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Michaleris, P., Dantzig, J. A., and Tortorelli, D. (1999). Minimization of welding residual stress and distortion in large structures. *Welding Journal*, 78:361s–366s.

Michaleris, P. and Sun, X. (1997). Finite Element Analysis of Thermal Tensioning Techniques Mitigating Weld Buckling Distortion. *Welding Journal*, 76:451s–457s.

Michalewicz, Z. (1992). *Genetic Algorithms+Data Structures=Evolution Programs*. Springer-Verlag.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers.

Mishra, R. and Ma, Z. (2005). Friction stir welding and processing. *Materials Science and Engineering R*, 50:1–78.

Mishra, S. and DebRoy, T. (2007). Tailoring gas tungsten arc weld geometry using a GA and a NN trained with convective heat flow calculations. *Materials Science and Engineering A*, 454-455:477–486.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. The MIT Press.

Muller, C., Baumgartner, B., and Sbalzarini, I. (2009). Particle Swarm CMA Evolution Strategy for the optimization of multi-funnel landscapes. pages 2685–2692. IEEE Congress on Evolutionary Computation (CEC 2009).

Murphy, A., McCune, W., Quinn, D., and Price, M. (2007). The characterization of friction stir welding process effects on stiffened panel buckling performance. *Thin-Walled Structures*, 45:339–351.

Murr, L., Liu, G., and McClure, J. (1997). Dynamic recrystallisation in the friction-stir welding of aluminium alloy 1100. *Journal of Materials Science Letters*, 16:1801–1803.

Myhr, O. and Grong, O. (1991a). Process Modelling Applied to 6082-T6 Aluminum Weldments - 1. Reaction Kinetics. *Acta metall. mater*, 39:2693–2702.

Myhr, O. and Grong, O. (1991b). Process Modelling Applied to 6082-T6 Aluminum Weldments - 2. Applications of model. *Acta metall. mater*, 39:2703–2708.

Myhr, O., Grong, O., Fjaer, H., and Marioara, C. (2004). Modelling of the microstructure and strength evolution in Al-Mg-Si alloys during multistage thermal processing. *Acta Materialia*, 52:4997–5008.

NAG Ltd. (1988). *The NAG Fortran library manual*. Numerical Algorithms Group Ltd., Oxford, UK.

Nandan, R., DebRoy, T., and Bhadeshia, H. (2008a). Recent advances in friction stir welding - process, weldment structure and properties. *Progress in Materials Science*, 53:980–1023.

Nandan, R., Lienert, T., and DebRoy, T. (2008b). Toward reliable calculations of heat and plastic flow during friction stir welding of ti-6al-4v alloy. *Int. J. Mat. Res.*, 99:434–444.

Nelder, J. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.

Nielsen, K. (2008). Ductile damage development in friction stir welded aluminum (AA2024)joints. *Engineering Fracture Mechanics*, 75:2795–2811.

Nielsen, K. and Tvergaard, V. (2009). Effect of ashear modified Gurson model on damage development in a fsw tensile specimen. *International Journal of Solids and Structures*, 46:587–601.

Olympio, K. (2007). A Hybrid Multi-Objective Genetic Algorithm for Topology Optimization. `http://water.engr.psu.edu/reed/Education/CE\%20563\%20Projects/ce563_olympio_project\%20\%20Sp\%2007.pdf`.

OpenOpt (2006). `http://openopt.org`. Free Python-written numerical optimization framework. Optimization Department of Cybernetics Institute, National Academy of Sciences of Ukraine.

Parks, G. and Miller, I. (1998). Selective breeding in a multiobjective genetic algorithm. pages 250–259. Parallel Problem Solving from Nature - PPSN V. 5th International Conference.

Peel, M., Steuwer, A., Preuss, M., and Withers, P. (2003). Microstructure, mechanical properties and residual stresses as afunction of welding speed in aluminum AA5083 friction stir welds. *Acta Materialia*, 51:4791–4801.

Poles, S. (2003). MOGA-II: An improved Multi-Objective Genetic Algorithm. Technical report, ESTECO s.r.l.

Poli, R., Langdon, W., and McPhee, N. (2008). A field guide to genetic programming. `http://www.gp-field-guide.org.uk`.

Poloni, C. and Pediroda, V. (1997). *GA coupled with computationally expensive simulations: tools to improve efficiency*, chapter Genetic Algorithms and Evolution Strategies in Engineering and Computer Science, pages 267–288. John Wiley and Sons.

Powell, M. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7:155–162.

Powell, M. (1978). A fast algorithm for nonlinearly constrained optimization calculations. volume 630 of *Lecture Notes in Mathematics*, pages 144–157. Springer Berlin - Heidelberg-New York.

Priesterjahn, S. (2007). Imitation-Based Evolution of Artificial Game Players. SIGEVOlution, newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation. Volume 2, Issue 4, pp. 2-13.

Radaj, D. (1992). *Heat Effects of Welding Temperature Field, Residual Stress and Distortion*. Springer-Verlag Berlin Heidelberg.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann-Holzboog Verlag.

Rechenberg, I. (1997). Bionik & evolutionstechnik. `http://www.bionik.tu-berlin.de`.

Richards, D., Pragnell, P., Williams, S., and Withers, P. (2008a). Global mechanical tensioning for the management of residual stresses in welds. *Materials Science and Engineering A*, 489:351–362.

Richards, D., Prangnell, P., Withers, P., Williams, S., Nagy, T., and Morgan, S. (2008b). Simulation of the effectiveness of dynamic cooling for controlling residual stresses in friction stir welds. 7th International Symposium Friction Stir Welding, TWI.

Rigoni, E. (2004a). Bench-marking NBI-NLPQLP. Technical Report 4, ESTECO s.r.l.

Rigoni, E. (2004b). Hole functions problem. Technical Report 2, ESTECO s.r.l.

Robin, V., Bernauer, G., Akgün, T., and Heubrandtner, T. (2005). Spotweld performance under high strain rate loading conditions. Mathematical Modelling of Weld Phenomena 7, Technische Universität, Graz.

Rosenthal, D. (1946). The theory of moving sources of heat and its application to metal treatments. *Transactions of the American Society of Mechanical Engineers*, 68:849–865.

Russell, M. and Shercliff, H. (1999). Analytical modelling of microstructure development in friction stir welding. USA. 1st International Symposium on Friction Stir Welding.

Satoh, K. (1972). Transient thermal stresses of weld heat-affected zone by-both-ends-fixed bar analogy. *Transactions of the Japan Welding Society*, 3:125–134.

Schittkowski, K. (1985). NLPQL: A fortran subroutine solving constrained nonlinear programming problems. *Annals of Operations Research*, 5:485–500.

Schmidt, H. and Hattel, J. (2004). An analytical model for the heat generation in friction stir welding. *Modeling and Simulation in Materials Science and Engineering*, 12:143–157.

Schmidt, H. and Hattel, J. (2005). Modeling the heat flow around tool probe in friction stir welding in friction stir welding. *Science and Technology of Welding and Joining*, 10:176–186.

Schmidt, H. and Hattel, J. (2008a). Thermal modeling of friction stir welding. *Scripta Materialia*, 58:332–337.

Schmidt, H. and Hattel, J. (2008b). A thermal-pseudo-mechanical model for the heat generation in Friction Stir Welding. Japan. 7th International Symposium on Friction Stir Welding.

Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*. John Wiley and Sons, Inc.

Seidel, T. and Reynolds, A. (2001). Visualization of the material flow in AA2195 Friction-Stir Welds using a marker insert technique. *Metallurgical and Materials Transactions A*, 32A:2879–2884.

Simo, J. and Hughes, T. (1998). *Computational Inelasticity*. Springer-Verlag New York, Inc.

Skillingberg, M. and Green, J. (2007). Aluminum applications in the rail industry. Light Metal Age.

Spellucci, P. (1999). *DONLP2 Short Users Guide*. Technische Universität Darmstadt, Darmstadt, Germany.

Srinivas, N. and Deb, K. (1994). Multi-objective optimization using non-dominated sorting in genetic algorithms. *Evolutionary Computation*, 2:221–248.

Storn, R. and Price, K. (1997). Differential Evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Tanev, I. and Shimohara, K. (2007). Towards human competitive driving of scale model of a car. SIGEVOlution, newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation. Volume 2, Issue 4, pp. 14-26.

The Math Works Inc. (2006). *MATLAB Optimization Toolbox Users Guide*.

The Math Works Inc. (2007). `http://www.mathworks.com`. (release R2007b).

Thorborg, J., Hald, J., and Hattel, J. (2006). Stellite failure on a P91 HP valve - Failure investigation and modelling of residual stresses. *Welding in the World*, 50:40–51.

Tutum, C. (2005). Structural optimization of truss structures subjected to stress, displacement, buckling and natural frequency constraints. Master's thesis, Istanbul Technical University, Istanbul, TR.

Tutum, C. (2006). Structural optimization of truss structures subjected to stress, displacement, buckling and natural frequency constraints. FİGES Users Conference.

Tutum, C. (2007a). Modelling of residual stresses in friction stir welding. Presentaion at the INNOJOINT Meeting.

Tutum, C. (2007b). Computational Welding Mechanics. Danish Center of Applied Mechanics and Mathematics (DCAMM) Ph.D. Course Report.

Tutum, C. and Hattel, J. (2008). Notes on the modeling of welding process course (41741). Technical report, Technical University of Denmark, Department of Mechanical Engineering.

Tutum, C., Schmidt, H., and Hattel, J. (2008a). Assessment of benchmark cases for modeling of residual stresses and distortions in friction stir welding. Japan. 7th International Symposium on Friction Stir Welding.

Tutum, C., Schmidt, H., and Hattel, J. (2008b). Optimization of the process parameters for controlling residual stress and distortion in friction stir welding. International Conference on Simulation Based Engineering and Sciences, TCN CAE.

Tutum, C., Schmidt, H., Hattel, J., and Bendsøe, M. (2007). Estimation of the welding speed and heat input in friction stir welding using thermal models and optimization. 7th World Congress on Structural and Multidisciplinary Optimization, Seoul.

Tutum, C., Schmidt, H., Hattel, J., and Bendsøe, M. (2009). A preliminary study on optimization of residual stresses in friction stir welding. (In press) "Journal of Materials and Manufacturing Processes".

Tvergaard, V. (2001). Plasticity and creep in structural materials. Department of Mechanical Engineering, Solid Mechanics, Technical University of Denmark.

Tvergaard, V. and Needleman, A. (1984). Analysis of the cup-cone fracture in a round tensile bar. *Acta Metallurgica*, 32:157–169.

TWI (1991). Friction stir welding at TWI. `http://www.twi.co.uk/content/fswintro.html`.

Vanderplaats, G. (1994). *DoT User Manual*. VMA Engineering, Colorado Springs, CO, USA.

Vanderplaats, G. (2005). *Numerical Optimization Techniques for Engineering Design*. Vanderplaats Research & Development, Inc.

Vecchi, M. and Kirkpatrick, S. (1983). Global wiring by simulated annealing. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2:215–222.

Vilaça, P., Quintino, L., and dos Santos, J. (2005). Analytical thermal model for friction stir welding. *Journal of Materials Processing Technology*, 169:452–465.

Waltz, R. and Plantenga, T. (2009). *KNITRO 6.0 Users Manual*. Ziena Optimization, Inc., Evanston, IL, USA.

Wikipedia (last modified on 1 November 2009b). Rubik's Cube. `http://en.wikipedia.org/wiki/Rubik's_Cube`.

Wikipedia (last modified on 24 July 2009a). Friction stir welding. `http://en.wikipedia.org/wiki/Friction_stir_welding.html`.

Wikiquote (last modified on 7 June 2008). George e. p. box. `http://en.wikiquote.org/wiki/George_E._P._Box`.

Zhu, X. and Chao, Y. (2002). Effects of temperature-dependent material properties on welding simulation. *Computers and Structures*, 80:967–976.

Zitzler, E. (1999). *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Emprical results. *Journal of Evolutionary Computation*, 8:173–195.

Zitzler, E., Laumanns, M., and Thiele, L. (2001). SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103, Swiss Federal Institute of Technology (ETH) Zurich.

Zitzler, E. and Thiele, L. (1998). An evolutionary algorithm for multiobjective optimization: the Strength Pareto approach. Technical Report 43, Swiss Federal Institute of Technology (ETH) Zurich.

Zitzler, E. and Thiele, L. (1999). Multiobjective evolutionary algorithms: a comparative case study and the Strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3:257–271.

Appendixes

A PAPER-I

C. C. Tutum, H. N. B. Schmidt, J. H. Hattel and M. P. Bendsøe, *"Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization"*, 7th World Congress on Structural and Multidisciplinary Optimization, COEX, Seoul, May 2007, Korea.

7th World Congress on Structural and Multidisciplinary Optimization
COEX Seoul, 21 May – 25 May 2007, Korea

Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization

Cem C. Tutum[1], Henrik Schmidt[1], Jesper Hattel[1], Martin P. Bendsøe[2]

[1] Department of Manufacturing Engineering and Management, Technical University of Denmark, Lyngby, Denmark, cem@ipl.dtu.dk
[2] Department of Mathematics, Technical University of Denmark, Lyngby, Denmark

Abstract
The objective of this work is to study the applicability of a relatively simple analytical thermal model for optimizing the welding parameters in Friction Stir Welding (FSW), i.e. heat input and welding speed, based on the assumptions of constant heat generation under the tool shoulder and symmetric heat flow along the welding line. A procedure to simulate the distributed heat source present in the real FSW process has been developed using a much simpler line source description. The range of applicability of the simplified model has also been investigated using more advanced numerical models. Sequential Quadratic Programming (SQP) is used for solving the optimization problem, in which the objective function is defined as the average temperature under the tool shoulder. Process related constraints are also included in the problem. Analytical and numerical sensitivity analyses are performed for the analytical and numerical models, respectively.
Keywords: Friction Stir Welding; Constrained Optimization; Analytical Thermal Model; Eulerian Finite Element Model

1. Introduction

Friction Stir Welding (FSW) is an efficient solid-state joining technique that is invented especially for aluminum alloys [1]. Improved mechanical properties, reduced distortion and residual stresses, and environment friendliness are some of its advantages [2]. The requirement for lighter structures, especially in aerospace, automotive and railway industries, emphasizes the need for the FSW process to be better understood.

The FSW process consists of several subsequent procedures here denoted as plunging, dwelling, actual welding and pulling the tool out of the work piece. First, the tool is submerged vertically into the joint line with high rotation speed, e.g. 400-1250 rpm, in the plunge period and then the dwell period takes place, where the tool is held steady relative to the work piece while rotating. Heat dissipation due to the friction and material deformation causes material to soften and allows traversing of the tool along the joint line. After initiation of the actual welding period, combination of sliding and sticking conditions under the tool shoulder plays an important role in obtaining preferred temperature ranges [3]. Following the welding period, the tool is pulled out of the work piece leaving a key hole behind.

New analytical and numerical models are being studied by many research groups in order to improve the process efficiency and performance of the components. Various aspects of modeling FSW are possible regarding microstructure evolution, material and heat flow, heat generation, investigation of the influence of process parameters, residual stress computation, predicting mechanical loads (tool forces, torque and adjustment of work piece fixture, etc.) and final mechanical properties of the joint. Thermal models are of great interest because they are often used as input for modeling some of the aspects mentioned above. Although analytical models can not substitute the state-of-the-art numerical models, they introduce a substantial flexibility and effectiveness in industrial design applications, and moreover they also contribute strongly to increasing the basic understanding of the process.

2. Thermal Modeling of FSW

In the FSW process, heat is generated by friction and plastic deformation between the tool and the work piece. The heat flows into the work piece as well as the tool. The amount of heat conducted into the work piece influences the quality of the weld, distortion and residual stress in the work piece. Insufficient heat from the friction could lead to failure of the tool pin since the work piece material is not soft enough. Therefore, understanding the heat aspect of the FSW process is extremely important, not only for understanding physical phenomena but also for improving the process efficiency.

Due to relatively high heat generation contribution from the surface of the tool shoulder, an assumption based only on modeling the tool shoulder is taken into account [3]. Besides this simplification, the shear layer formed around the tool due to high rotation speed is neglected; so a symmetric temperature field along the joint line is obtained in the numerical model.

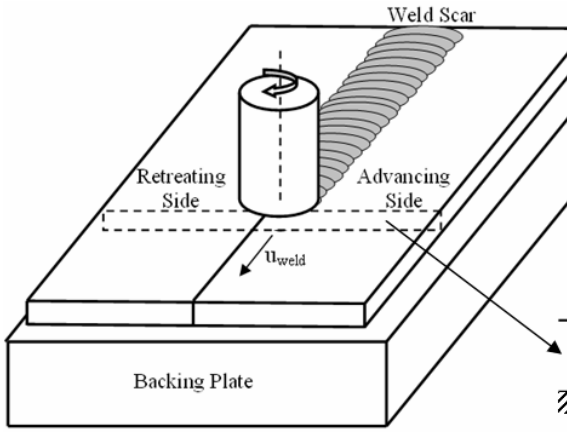


Figure 1a. Schematic view of the friction stir welding process

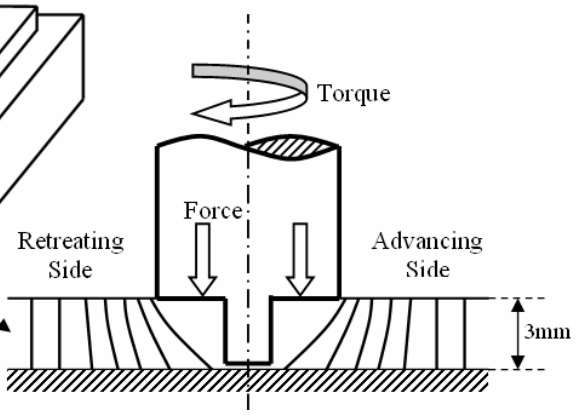


Figure 1b. Cross section view of a simple tool geometry and plate with applied loads

In order to obtain a closed form solution for practical evaluation of the effects of the welding parameters, some simplifications regarding heat input and geometry are considered for the analytical solution. Two-dimensional heat propagation in an infinitely large thin plate is assumed and a constant heat flow density over the plate thickness is suggested due to the low temperature gradients through the plate thickness that is observed in the real FSW process, as seen in Figure 1a-b.

2.1. Numerical Thermal Model

Modeling the whole welding process, i.e. plunge, dwell and pull out periods, holds some notable complexities. In order to reduce the computational cost regarding moving heat source, meanwhile preserving the applicability, only the welding period is taken into account and a moving coordinate system which is located on the heat source is applied. Eq.(1) describes the heat transfer in the plate. The traverse motion of the tool is modeled by prescribing a material flow through the rectangular plate region. Due to this flow prescription, the equation includes a convective term in addition to the conductive term.

$$\rho c_p \frac{\partial T}{\partial t} + \nabla.(\ k\nabla \mathbf{T} + \rho c_p T\mathbf{u}) = q \tag{1}$$

Where ρ denotes material density, c_p is heat capacity, T denotes temperature, k is the coefficient of heat conduction, u is the prescribed flow vector and q denotes the stationary heat source. The present model simulates the steady state conditions during the welding period, thus the time dependent term is neglected. If the heat flow problem involves material flow, the Eulerian formulation is often advantageous. In Figure 2, the boundary conditions are described for the steady state Eulerian model in the simulation tool COMSOL.

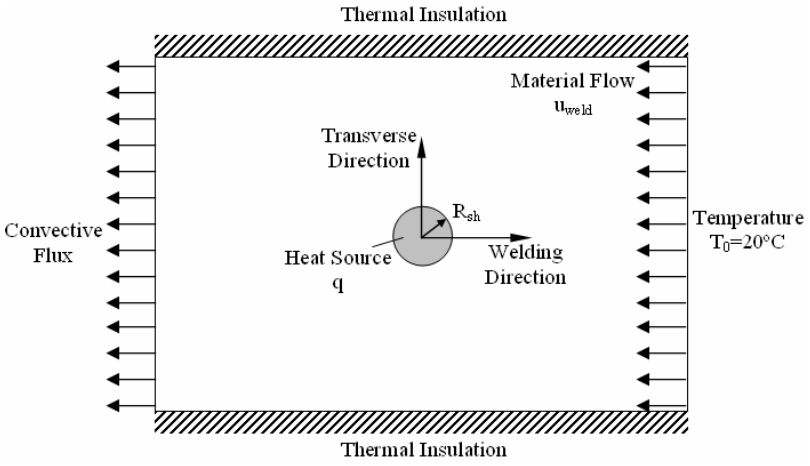


Figure 2. Schematic view of the numerical model with boundary conditions

The circle with a radius of R_{sh}=10mm, in the middle of the rectangular region, represents the tool shoulder where the stationary and a uniform heat source is applied as a volume flux. The tool is stationary while the material flows through it with a velocity of u_{weld} in the opposite welding direction, in Figure 2. The room temperature (20ºC) is defined at the right edge of the rectangular region. The heat flux on the left edge of the plate region, where the aluminum leaves the computational domain, is dominated by convection. On the upper and lower edges of the plate boundaries, thermal insulation is enforced.

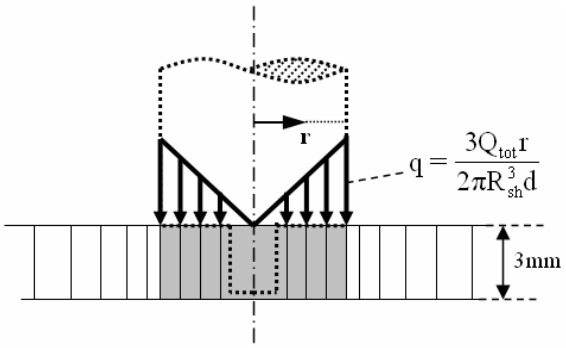


Figure 3. Schematic view of tool geometry with applied linear heat source on tool shoulder

A schematic view of the applied volume heat flux as shown in Figure 3, is given by, (cf., [4])

$$q = \frac{3Q_{tot}r}{2\pi R_{sh}^3 d} \tag{2}$$

where Q_{tot} is the total heat input, d denotes the thickness of the actual shear layer under consideration and r is the radius originating at the center of the tool. This volume heat flux corresponds to a surface heat flux from the tool shoulder (without the tool probe) [4], distributed through the plate thickness under the assumption of 2D heat flow.

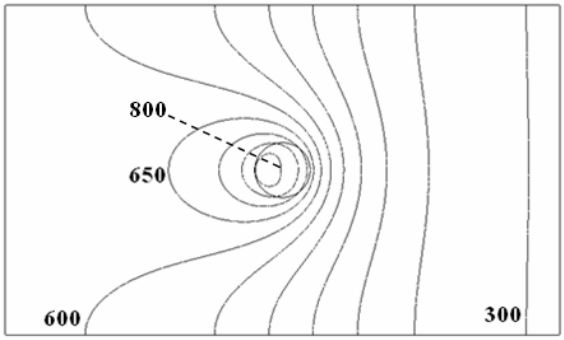


Figure 4. Contour plot of the temperature field. Increment between isotherms is 50K

Figure 4 shows the contour plot of the resulting temperature field of the model (isotherms in 50K increments). The modeling includes aluminum properties for the work piece material for a 3mm-thick plate. The welding speed is 2mm/s. Notice that the maximum temperature is around 800K, which is a result of having higher temperatures than a desired average temperature over the shoulder area of 773.15K. This is not physically possible; however this is a result of the use of a simple heat source model.

2.2. Analytical Thermal Model

A schematic view of a line-type heat source moving with a linear speed in the infinitely extended plate is shown in Figure 5. The temperature field is governed by the same heat conduction partial differential equation, including the convection term, as given in Eq.(1). The closed form solution can be obtained under quasi-steady state assumptions as given by Eq.(3), [5, 6]

$$T = \frac{Q_{tot}}{2\pi kd} e^{u_{weld}\xi/2a} K_0\left(\frac{u_{weld}}{2a} r\right) + T_0 \tag{3}$$

where (Q_{tot}/d) defines the heat input per unit thickness, ξ denotes moving coordinate axis, a is the heat diffusivity, $K_0(\ )$ denotes the Modified Bessel Function of the second kind and zeroth order and T_0 defines the ambient temperature.

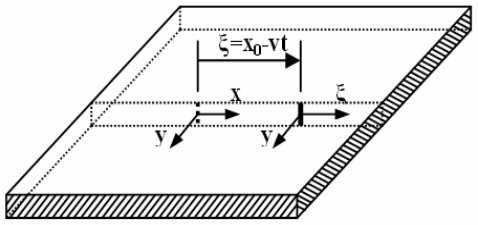


Figure 5. Schematic view of a line-type heat source moving along the joint line of the work piece

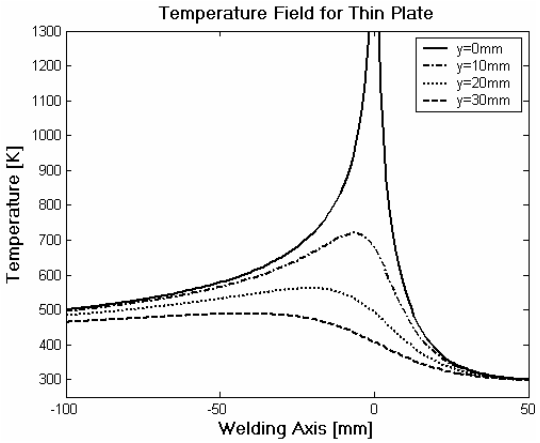


Figure 6. Temperature profile along x-axis at different y-points

Figure 6 shows the resulting temperature field along the welding axis, including variation in the transverse direction. Notice the mathematical singularity of infinite temperature at the location of the heat source, which of course is not possible in real applications.

3. Optimization

The optimization problem here is stated as the goal of finding the Friction Stir Welding parameters, i.e. heat input and longitudinal welding speed, which provides a desired average temperature under some process limitations. The problem can be formulated as in Eq.(4),

$$\begin{aligned} &\textbf{Minimize:}\ (T_{average}\text{-}773.15)^2 = \textbf{Minimize: } \mathbf{F^2} \\ &\textbf{Subject to:}\ 0<Q_{tot}<10000\ \text{W}, \\ &\qquad\qquad 0<u_{weld}<100\ \text{mm/s} \end{aligned} \tag{4}$$

where the objective function (F^2) is the error between the analytically or numerically computed value and the desired average temperature under the tool shoulder.

It would be advantageous to obtain an analytical expression for the sensitivity computation based on Eq.(3); however a semi-analytical procedure is presented in section 3.2.3. The necessary Finite Difference calculations were performed for the COMSOL thermal model. Standard MATLAB Optimization toolbox routines are used for the two equivalent optimization problems.

3.1. Optimization of the Numerical Model

A steady state Eulerian heat flow model implemented in COMSOL is used to control the average temperature under the tool shoulder. In order to find the optimum welding parameters, i.e. the heat input and welding speed for the bounded constrained problem defined as in Eq.(4), the Sequential Quadratic Method with BFGS line search algorithm is used [7]. Finite difference calculations are performed in order to obtain sensitivity information.

The objective function field for the corresponding design variable field is shown in Figure 7. It is obvious to find multiple solutions since the optimization problem has only simple boundaries without any equality or inequality constraints. The resulting optimum curve which is labeled as the contour line equal to 1 is lying close to the zero plane, as can be seen in Figure 8.

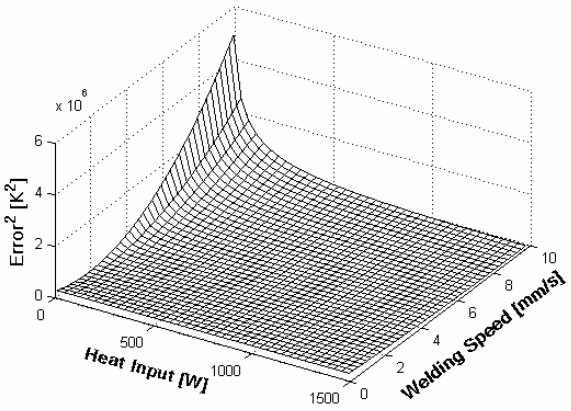


Figure 7. Objective function field (Error^2) vs. design variable field, i.e. heat input and welding speed

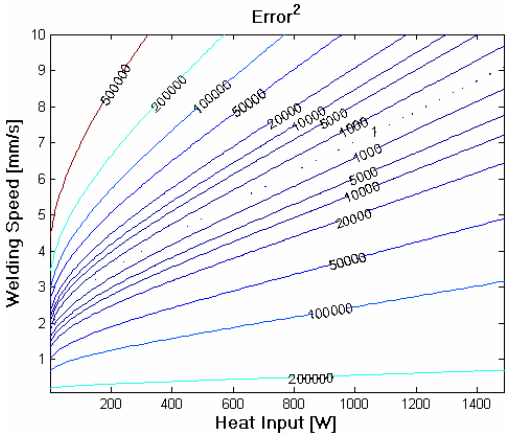


Figure 8. Contour plot of the corresponding objective function field

3.2. Optimization of the Analytical Model

The finite element method gives the user the possibility to define complicated models, including complex geometries and heat source models, etc., while analytical models have some limitations in some aspects of the engineering applications, e.g. FSW modeling. In order to extend the applicability of the analytical formulation, a novel procedure is introduced. The power of the line-type heat source is controlled by evaluating the temperatures in a circular region surrounding the heat source, to be used as a process control parameter. The corresponding bound constrained problem is solved using a subspace trust region method based on reflective Newton method [8].

3.2.1. Derivation of the Closed Form Solution

A Moving heat source representing the tool shoulder is described by a circle in an infinitely extended plate, as mentioned before and shown in Figure 9. In order to obtain a closed form solution for the average temperature under the tool shoulder, the temperature field defined in Eq.(3) should be integrated over the circular area in polar coordinates, as formulated in Eq.(4).

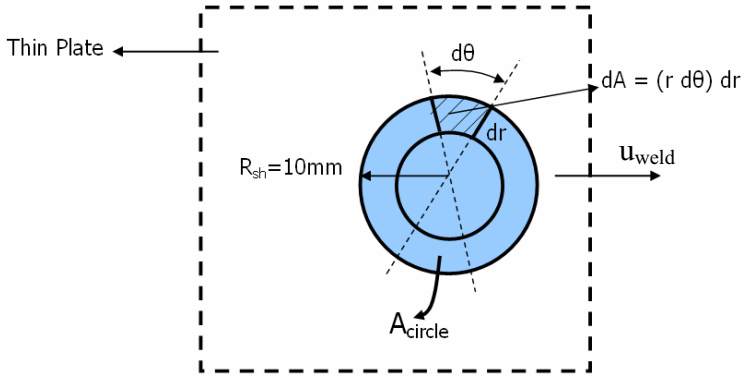


Figure 9. A moving heat source in an infinite plate defined in polar coordinates

$$T_{average} = \frac{\int_A T.dA}{\int_A dA} = \frac{\int_{r=0}^{R_{sh}} \int_{\theta=0}^{2\pi} T(r,\theta)(rd\theta dr)}{A_{circle}} = \frac{Q_{tot}}{2\pi^2 k d R_{sh}^2} \int_{r=0}^{Rsh} \int_{\theta=0}^{2\pi} \frac{r.K_0(wr)}{e^{wr\cos(\theta)}} d\theta dr + T_0 \quad (5)$$

The integral in Eq.(5) can not be evaluated analytically due to the singularity of the K_0 function in the formulation of the line-type heat source. Thus, an incremental scheme will be defined in the following section.

3.2.2. Numerical Computation of the Average Temperature

In this section, the incremental scheme for evaluating the average temperature under the tool shoulder is formulated. The line-type heat source is positioned in the center of the circle, defined as (0,0) in Cartesian coordinates, in Figure 10.

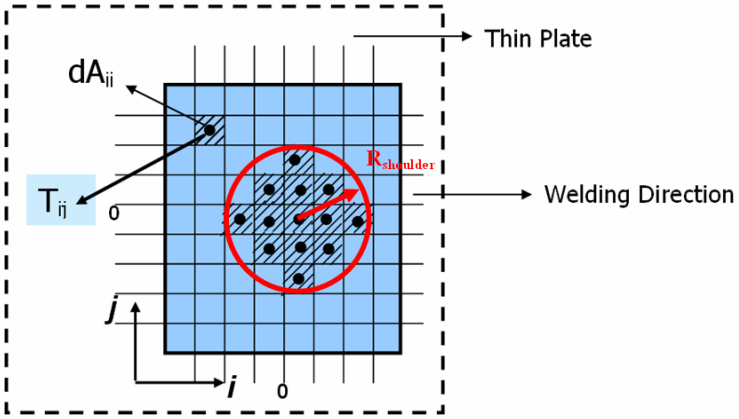


Figure 10. A moving heat source in an infinite plate defined in Cartesian coordinates

Eq.(6) defines the average temperature, similar to Eq.(5), in two dimensional (indices) summation form. Since all cells are equal in size, the denominator of Eq.(6) is reduced to the number of cells. The temperature values evaluated on the grid points (centers of the cells) placed in the circular area are summed easily, as defined individually in Eq.(3).

$$T_{average} = \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} T_{ij} dA_{ij}}{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} dA_{ij}} = \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j} T_{ij}}{n_{cell}} \tag{6}$$

The grid size influences the accuracy of the average temperature expression defined in Eq.(6). Thus, to ensure independence of this, a convergence test is performed by using the same heat input value for changing grid density along the diameter of the circular region representing the tool shoulder. Figure 11 shows that the convergence is achieved for a low number of increments.

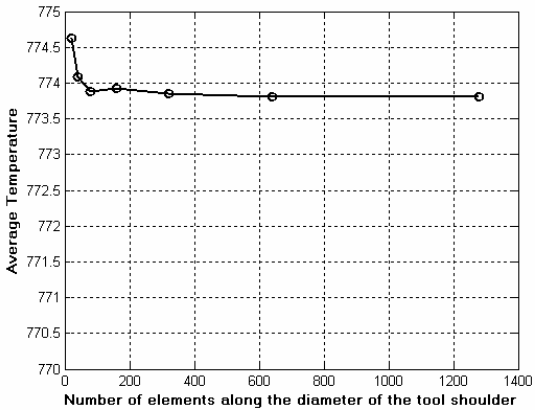


Figure11. Convergence test plot of the average temperature expression for different grid size

3.2.3. Semi-Analytical Sensitivities:

The possibility to use analytical sensitivities provides substantial benefits regarding savings in computational cost. It is found, cf. section 3.2.1, that it is not possible to obtain a closed form solution for the average temperature criteria since integration of the K_0 function causes difficulties. In this section, it is shown that the semi-analytical method can be used to reduce computational cost as far as possible. The sensitivities of the objective function with respect to the welding parameters, i.e. heat input and welding speed, are defined in Eq.(7-8), respectively.

$$\frac{d(F^2)}{dQ_{tot}} = 2F.\left(\frac{C}{\pi R^2}\int_0^{R}\int_0^{2\pi}\frac{r.K_0(wr)}{e^{wr\cos(\theta)}}d\theta dr\right) = \frac{2F(F+T_1)}{q} \tag{7}$$

$$\frac{d(F^2)}{du_{weld}} = 2F.C_2\int_0^{R}\int_0^{2\pi}\left[\frac{r^2}{e^{wr\cos(\theta)}}\{K_1(wr)+K_0(wr).\cos(\theta)\}\right]d\theta dr \tag{8}$$

Here F^2 denotes the objective function, T_1=480K, $C=1/(2\pi kd)$ and $C_2=-C(Q_{tot}/a)$ which are used to simplify the expressions and $K_1(\ )$ is the Modified Bessel Function of the second kind and the first order. The first sensitivity of the objective function with respect to the heat input is relatively easier to compute as compared to the second, since all variables in Eq.(7) are already evaluated while computing the objective value as given in Eq.(4), using the incremental procedure. The second sensitivity expression is formulated in an analytical way, but again evaluated using the incremental procedure.

4. Results and Discussion

The optimization problem defined in Eq.(4) gives multiple solutions as mentioned in section 3.1. In order to evaluate the results from a more practical point of view, the design variables (heat input and welding speed) are considered separately. Table 1 and 2 show the results regarding the specified optimum design variable for the other fixed welding parameter, respectively. Aluminum is used for the work piece material that is a 3 mm-thick plate. Notice that the analytical heat input solutions are scalable for different plate thicknesses due to the linearity in terms of the heat generation density as given in Eq.(3).

Table 1. Comparison of the optimum heat input results for given welding speed and relative error in relation to the analytical heat input

u_{weld} (mm/s)	$Q_{tot,Analytical}$ (W)	$Q_{tot,COMSOL}$ (W)	$\frac{\lvert\Delta Q_{tot}\rvert}{Q_{tot,Analytical}}$ (%)
2	575	585	2
4	780	890	14
6	971	1153	18
8	1160	1404	21
10	1350	1653	22

Notice that the higher the welding speed, the higher the error due to the boundary effect. The analytical heat flow model does not see the effects of the boundary conditions because of the infinitely large domain description, while it is not the case for the numerical model having a finite domain.

Table 2. Comparison of the optimum welding speed results for given heat input and relative error in relation to the analytical welding speed

Q_{tot} (W)	$u_{weld,Analytical}$ (mm/s)	$u_{weld,COMSOL}$ (mm/s)	$\frac{\lvert\Delta u_{weld}\rvert}{u_{weld,Analytical}}$ (%)
500	1.36	1.50	10
750	3.69	3.00	18
1000	6.32	4.80	24
1250	8.97	6.80	24
1500	11.56	8.80	24

5. Conclusions

A procedure to simulate the distributed heat source present in the real FSW process has been developed using a simple line-type heat source description. The range of applicability of the simplified model has also been investigated using more advanced Eulerian finite element model. It is stated that the deviations between analytical and numerical models are reasonable especially for the lower welding speeds, which is also considerable for the FSW process.

Sequential Quadratic Programming (SQP) is used for solving the bound constrained problem, in which the objective function is defined as the average temperature under the tool shoulder. Semi-analytical and numerical sensitivity analyses are performed for the analytical and Eulerian heat flow models, respectively.

6. References

1. Thomas W M et al 1991 Friction Stir Butt Welding International Patent Application No: PCT/GB92/02203
Thomas W M et al 1995 Friction Stir Butt Welding GB Patent Application No 9125978.8
Thomas W M et al 1995 Friction Stir Butt Welding UP Patent 5 460 317
2. R.S. Mishra and Z.Y. Ma, Friction stir welding and processing, Materials Science and Engineering R 50, 2005, 1-78.
3. H Schmidt, J Hattel and J Wert, "An Analytical Model for the Heat Generation in Friction Stir Welding", Modelling and Simulation in Material Science Engineering 12, 2004, 143–157.
4. H Schmidt, J Hattel, "Modelling heat flow around tool probe in friction stir welding", Sci. Technol. Weld. Joining, 2005, 10, 176-186.
5. D. Rosenthal, "The theory of moving sources of heat and its application to metal treatments", Trans. ASME, 1946, 849–866.
6. D., Radaj, Heat Effects of Welding: Temperature Field, Residual Stress, Distortion, Spring-Verlag, 1992.
7. G. N., Vanderplaats, Numerical Optimization Techniques for Engineering Design, 2005.
8. The Math Works Inc, MATLAB Optimization Toolbox User's Guide.

B PAPER-II

C. C. Tutum, H. N. B. Schmidt and J. H. Hattel, *"Assessment of Benchmark Cases for Modeling of Residual Stresses and Distortions in Friction Stir Welding"*, 7th International Symposium on Friction Stir Welding, TWI, 2008, Awaji Island, Japan.

Assessment of Benchmark Cases for Modeling of Residual Stresses and Distortions in Friction Stir Welding

Tutum, Cem C.*, Schmidt, Henrik N. B., Hattel, Jesper H.

Department of Mechanical Engineering, Technical University of Denmark
* Corresponding author: cem@ipl.dtu.dk

Abstract:
In the present paper, the residual stresses and distortions of several friction stir welded (FSW) joints are investigated with 3-dimensional fully-coupled thermo-mechanical models implemented in ABAQUS. A non-linear transient thermal analysis is performed assuming a heat source linearly increasing with radius and distributed over a circular area representing the tool shoulder. The resulting temperature field is simultaneously coupled with an elasto-plastic mechanical model in order to obtain residual stresses and deformations. In addition to this, the residual state of the simulated structure is also used as an input for a subsequent load analysis in order to analyze the performance of the joint/structure during exposure to service loads. Based on the analyses a suggestion for evaluating different geometries as benchmark cases for residual stress and deformation analysis of FSW is presented.

1. Introduction:

Friction stir welding is an efficient solid-state joining technique that is intended to be used for joining of especially the aluminum structures besides dissimilar welds [18]. Although the level of residual stress measurements observed during FSW process is relatively lower comparing to the traditional welding techniques such as fusion welding [8], it has been shown that residual stresses play a major role for the fatigue crack growth [7,9,10], buckling behavior [6,17], etc. In order to understand and control the evolution of thermal and residual stresses in friction stir welding, modeling works have been carried out as given in Ref. [11-16].

The benchmark cases presented hereby could be seen as a frame work for reproducibility, validation and comparison of the results obtained by other modeling groups using different commercial or in-house codes. Besides validation purposes, the general characteristics of the residual stress and deformation due to different kinds of fixture settings which are already used in industrial applications are investigated in a structured way.

These benchmark cases are described in details regarding workpiece geometry, boundary conditions, heat source and service loads together with the corresponding finite element models. Each benchmark case focuses on specific aspects of the thermal and mechanical response, i.e. temperature field, in-plane/out-of-plane stresses and distortions. The in-service loads are chosen with respect to the mechanical response characteristics of the structure, e.g. beam structures exposed to bending loads; at structures for in-plane tensile/compressive loads. Hence, the in-service loads are chosen for the purpose of benchmarking thereby resembling, but not being real experimental in-service load cases.

2. General Description of the Models:

Two geometrical models have been prepared. The first model presents the fundamental benchmark study which resembles bead on the plate. Only one of the plates is considered due to the symmetry conditions in the model, i.e. mechanical and thermal boundary conditions and thermal load which is given as moving heat source on the centre line. Benchmark material properties are defined as given in Table-1.

Table-1. Thermal and Mechanical Material Properties of Benchmark Material

Heat Conduction, k	[W/(m.K)]	160
Heat Capacity, c_p	[J/(kg.K)]	900
Initial Yield Strength, σ_y	[MPa]	200
Young's Modulus, E	[GPa]	70
Tangent Modulus, E_t	[GPa]	7
Thermal Expansion Coefficient, α	[1/K]	2.3×10^{-5}
Density, ρ	[kg/m^3]	2700

The temperature dependent yield strength is defined basically as shown in Figure 1, which is given as a linear function with a negative slope, decreasing from 200 MPa to 10 MPa at 20 °C and 500 °C, respectively. Model-1 is tested with both temperature dependent and constant material properties in order to see the thermal softening effect on residual stress evolution together with different mechanical boundary conditions.

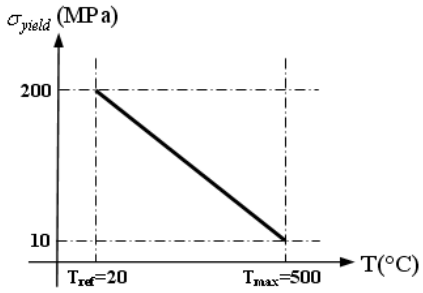


Figure 1. Temperature dependent yield strength

Due to relatively high heat generation contribution from the surface of the tool shoulder, an assumption based only on modeling the tool shoulder is taken into account. Besides this simplification, the shear layer formed around the tool is neglected; so a symmetric temperature field along the joint line is obtained. A schematic view of the applied surface heat flux as shown in Figure 2, is given by, (cf., [1-4])

$$q = \frac{3Q_{tot}r}{2\pi R_{sh}^3} \tag{1}$$

where Q_{tot} is the total heat input, r is the radius originating at the center of the tool. Eq.1 represents the surface heat flux which is increasing linearly along the tool radius due to the increasing peripheral velocity, thus increasing plastic deformation.

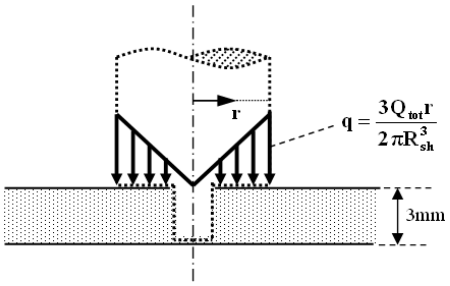


Figure 2. Schematic view of the surface heat flux

The mechanical boundary conditions play a major role in the evolution of the thermal stresses and resulting residual stresses. Common definitions are used in order to keep consistency between different models. In the first benchmark model, three different mechanical boundary conditions are defined, i.e. clamping both edges which are perpendicular to welding line, clamping the side edge which is parallel to the welding line and clamping the left corner of the plate lying on the symmetry line. Besides these different kinds of clamping options, a common displacement boundary condition which is called as rollers, representing the displacement restriction on the bottom of the plate due to backing plate, are applied on all models, i.e. model-1 and 2. The actual joint between the two plates is modeled by a symmetry boundary condition; displacement perpendicular to the symmetry plane is constrained as well as rotation around the axis perpendicular to the workpiece (z-axis).

First benchmark model consists of three steps, e.g. welding, cooling and release, respectively. A service load step is also considered for the second benchmark model in order to evaluate the stress distribution in a case of possible industrial application and to find an optimal way to use the residual state of the friction stir welded structure, i.e. relatively homogeneous stress distribution all over the structure for fatigue assessment, choosing the type of loading depending on the residual stress distribution on the structure, etc.

Welding step is modeled using a moving surface heat flux along the joint line of the workpiece. In the cooling step, the plate is allowed to cool down to the room temperature by keeping the thermal boundary conditions without the application of the heat source and setting the mechanical boundary conditions constant. In the release step, the structure is clamped in the corner to suppress rigid body motions while retaining the symmetry boundary conditions. Following these three steps, a case-dependent static service load is defined in the second model as an additional step.

2.1. Model-1:

This model represents the welding of the two flat plates resembling the bead on the plate. Due to symmetry assumptions, e.g. neglecting the asymmetric thermal source contribution due to formation of the shear layer, symmetric thermal and mechanical boundary conditions, etc., only one of the plates is modeled. The dimension of the workpiece is 300 x 150 x 3 mm and the material properties, which are given in Table-1, does not correspond to any specific commercial Al-alloy but should be understood as a benchmark material. Welding speed is defined as 2 mm/sec. The diameter of the tool shoulder is 20mm. The moving heat source starts and stops at 50-mm from the left and the right edges of the plate, respectively, so the welding step is defined for 100 seconds. Different clamping conditions, corresponding mechanical boundary conditions defined in the following sections from 2.1.1, Case-A to 2.1.3, Case-C, are tested in order to visualize the effect of them on transient and residual stress states.

Model-1 is composed of two regions, i.e. weld region and outer region, that are enmeshed using a structural scheme; weld region: 2 x 2 x 1.5mm (corresponding to 150 x 10 x 2 elements along the edges), outer region: 2 x 3.25 x 1.5mm (corresponding to 150 x 40 x 2 elements), i.e. a total of 18,000 C3D8T linear coupled temperature-displacement elements, see Figure 4.

In order to simulate the moving heat source during the welding step, a position and time dependent distributed surface heat flux is applied on element faces using *DFLUX user subroutine in ABAQUS. The total heat power of 600W and the radius of the tool shoulder of 10mm are used for the corresponding constants in Eq.1. The heat source is turned off during the rest of the steps, i.e. cooling and releasing. The heat transfer from the plate to the environment, which plays an important role in usual applications, is neglected in order to simplify the benchmark study and focus on the effect of the mechanical boundary conditions.

Since all the thermal boundary conditions are same in three cases of the first benchmark model, only the mechanical boundary conditions, which is also same at the third step, are described in detail in the following three sections.

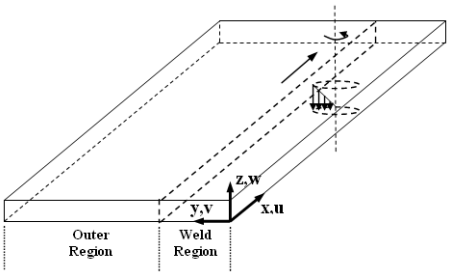


Figure 3. Schematic view of Model-1

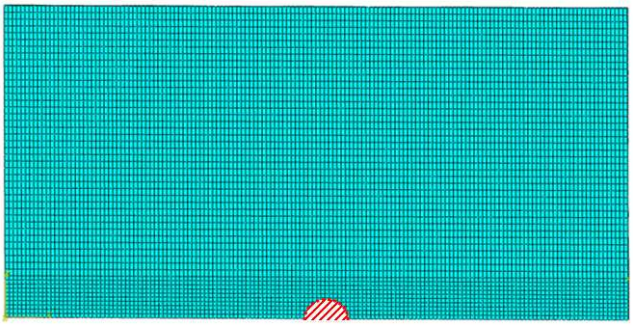


Figure 4. Enmeshment of Model-1

2.1.1. Case-A

Case-A considers the clamping of the both ends of the plate perpendicular to the symmetry line. Rollers, which resemble the vertical displacement restriction due to the backing plate, are applied at the bottom of the plate. These clamping and roller constraints are kept constant also during the cooling step and releasing step is applied as defined in the previous section.

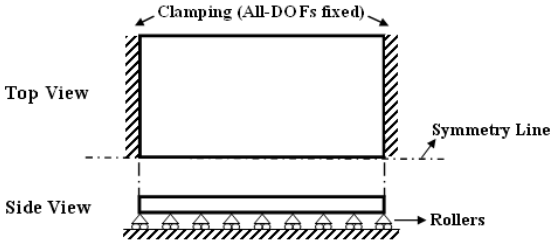


Figure 5. Step-1 and 2 for Case-A

Figure 6 shows the common description of the clamping procedure applied in the third step of each case of the first model, i.e. Case-A, B and C. The down left corner is clamped that all translational and rotational degrees of freedom are fixed to zero in order to simulate releasing condition.

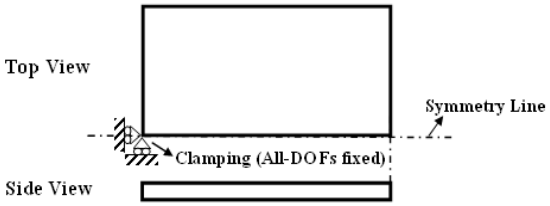


Figure 6. Schematic description of the mechanical boundary conditions at the third step of each case for Model-1

2.1.2. Case-B

Case-B considers the type of clamping mostly used in industrial applications, that the side edge of the plate parallel to the symmetry line are fixed. Rollers are applied at the bottom of the plate.

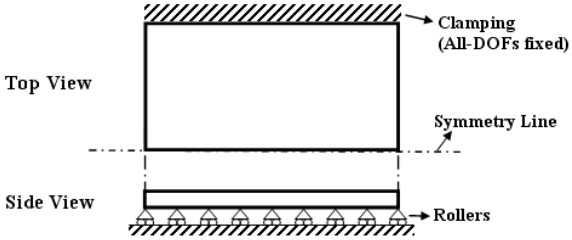


Figure 7. Step-1 and 2 for Case-B

2.1.3. Case-C

Case-C considers a theoretical case with the boundary conditions, shown in Figure 8, mainly applied to account for the investigation of the effect of the mechanical constraints, which corresponds to bead on plate with clamping.

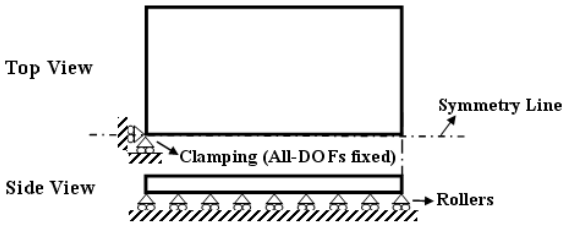


Figure 8. Step-1 and 2 for Case-C

2.2. Model-2:

This model represents the friction stir welding of an L-shaped stringer on a 300 x 300 x 1.6-mm plate as shown schematically in Figure 9, both having the benchmark material properties, see Table-1. The cross-sectional dimensions of the stringer are given in Figure 10, representing an industrial application [5]. In order to define a benchmark model for a complex case as presented here, the interfacial condition between

the plate and the stringer has been assumed to be “perfect”, i.e. continuum properties across and along the plate during the interface. Hence the workpiece is an extrusion of a continuous cross-sectional profile has been modeled. The welding speed is 2 mm/sec and the heat input is 1050 W. The diameter of the tool shoulder is 20mm. The weld starts and stops at 50-mm away from the ends of the plate. The heat transfer from the plate surfaces to environment is taken into account. The heat transfer coefficients are 20 W/(m^2.K) and 1000 W/(m^2.K) at top and bottom surfaces of the plate, respectively. The reference temperature is 20 °C.

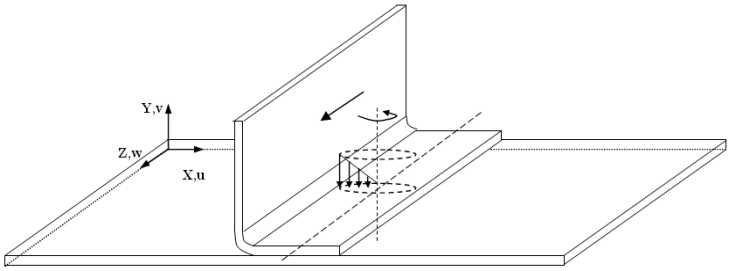


Figure 9. Model-2: the friction stir welding of an L-shaped stringer on a plate

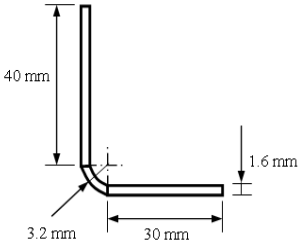


Figure 10. Dimensions of the cross-section of the L-shaped stringer

The plate is clamped at both sides along the weld direction and rollers are applied at the bottom of the plate due to the displacement restriction into the backing plate during the welding and cooling steps. In the releasing step, one of the corners of the plate is constrained and all of the previous mechanical boundary conditions are deactivated. At the end of releasing step, the plate is clamped along one of the edges perpendicular to the welding line and a service load, which is defined as a bending load of 25kg at the other end of the plate, is applied using a traction stress distributed over the cross section area of the “free” end.

3. Results:

Thermo-mechanical results of the two benchmark models are presented in the following sections. Model-1 has three cases consisting of same thermal field vs. three different mechanical fields due to different mechanical boundary conditions. Since the temperature field is the same for all of the cases of the Model-1, thermal results are given first in the following section, and then the mechanical field is considered separately for each case.

3.1. Results of Model-1:

The finite element model of the first benchmark geometry defined in Section 2.1, i.e. having total heat input of 600 W, welding speed of 2 mm/s, plate dimensions of 300 x 150 x 3 mm, results a maximum temperature of 494.2°C during whole welding step and 107.2°C at the end of cooling step of the simulation. In order to check the reliability of the results depending on the referred finite element mesh density, a analytical calculation of the conservation of the energy principle is given in Eq. 2,

$$\rho c_p V (T_{end} - T_{ref}) = t_{weld} \frac{Q_{tot}}{2} \tag{2}$$

where ρ , c_p and V are the material density, heat capacity and volume of the workpiece, respectively. T_{end} shows the temperature obtained at the end of cooling step and T_{ref} is the reference temperature of 20°C. T_{weld} shows the welding time of 100s and Q_{tot} is the total heat input. Due to symmetry, only half of the total heat input is considered. The analytical value of T_{end} is found to be 111.4°C using Eq. 2. The difference between the numerical and analytical values of T_{end} (4.2°C) could be reduced by increasing mesh density relative to the heat source, however the current setup is a trade off between accuracy and computational costs. This difference could be seen as a criterion when comparing this benchmark case of FSW residual stress model using different software.

3.1.1. Case-A:

At the end of the cooling step, the temperature field is stabilized at 107.2°C. Longitudinal and transverse stresses, which are noted at the middle of the plate on a path defined between the points of (0.15, 0.0, 0.0015) and (0.15, 0.15, 0.0015), are shown in Figure 11. The workpiece material, that has already yielded in compression, transpose into tensile stresses by the decrease of thermal strains during the cooling period and approximately 60MPa is obtained on the center line while 175MPa compressive stress is observed at

approximately 60mm away from the center line in the transverse direction. Following the cooling step, the clamps and rollers are released at both ends and bottom surface of the plate, respectively, and the stresses reestablish as shown in Figure 11, and stresses shift approximately 125 MPa in the positive axis. Transverse stresses are quite low compared to longitudinal stresses as expected due to the geometry of the plate, welding direction, mechanical boundary conditions.

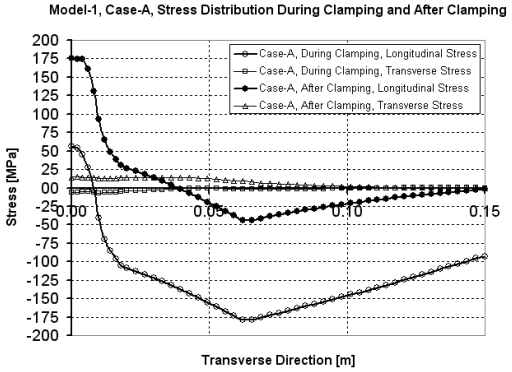


Figure 11. Stress distribution along the transverse direction of the plate for Case-A

Both longitudinal and transverse stresses get higher when the temperature dependent yield stress is taken into account as seen in Figure 12. The maximum tensile stress obtained in the center line shifts from 200 MPa to 275 MPa, due to much higher plastic strain observed, at the end of releasing step. The increase in the maximum value of the transverse stresses along the transverse direction of the plate compared to the ones observed in Figure 11 is approximately 25 MPa. The maximum compressive longitudinal stress obtained close to the middle of the plate lays around 60 MPa at the end of releasing step. It can be seen by comparing Figure 11 and 12 that the general characteristics of both the longitudinal and transverse stress are preserved while changing in magnitude.

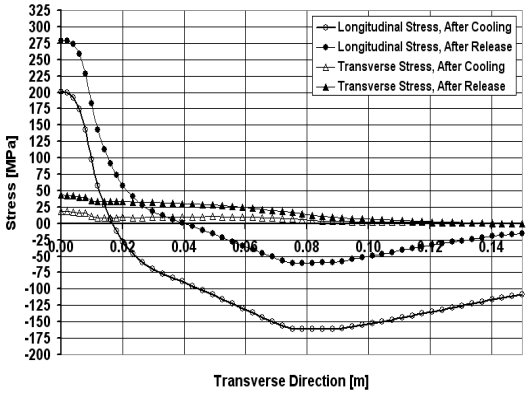


Figure 12. Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-A

3.1.2. Case-B:

Same temperature field is obtained through all three cases, as should be, due to neglecting of adiabatic heating effects and same temperature field related model parameters, i.e. total heat input, thermal expansion coefficient, adiabatic thermal boundary conditions, etc. Longitudinal and transverse stresses are shown in Figure 13, which is graphed using the same path defined in the previous section. In this case, the side of the plate is clamped and stress evolution directly captures this resistance in a way of building a localized compressive zone close to the transverse edge of the plate during the welding step. Maximum compressive longitudinal stresses obtained are approximately 170 MPa, while 135 MPa is observed also for the maximum transverse stresses. In the following releasing step, equilibrium in the longitudinal stresses is observed and the maximum tensile stress reaches approximately to 125 MPa while almost negligible values are obtained for the transverse stresses.

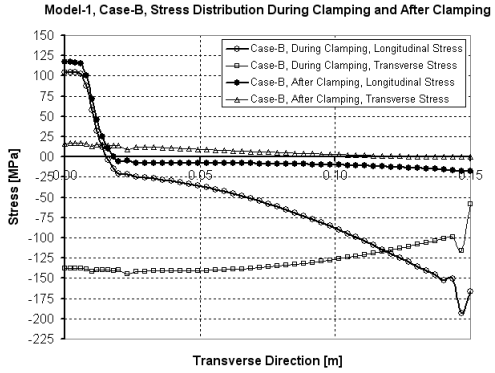


Figure 13. Stress distribution along the transverse direction of the plate for Case-B

Temperature dependence for the yield stress results higher longitudinal stress due to higher plastic strain, as mentioned in the previous section. Figure 14 shows the longitudinal and transverse stress results along the transverse direction for Case-B. Maximum longitudinal stress is 180 MPa on the center line which is loaded in tension while maximum transverse stress is approximately 20 MPa on the centerline which is also loaded in tension at the end of releasing step.

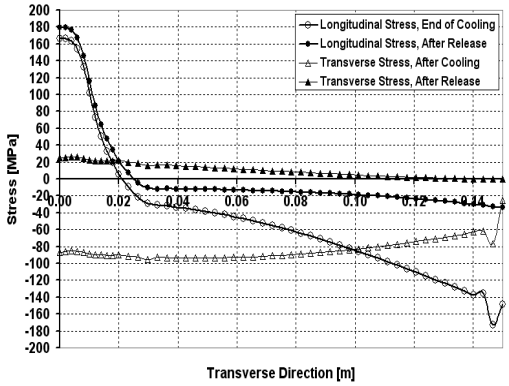


Figure 14. Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-B

3.1.3. Case-C:

This is the simplest case out of three cases due to the simplicity of the mechanical boundary conditions. After both the cooling and releasing steps, the stress evolution almost doesn't change since the force equilibrium remains almost constant in these two steps.

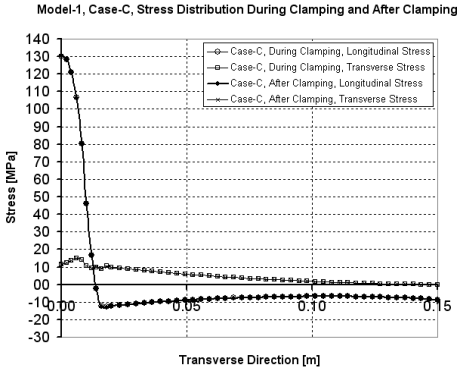


Figure 15. Stress distribution along the transverse direction of the plate for Case-C

Figure 16 shows the stress distribution along the transverse direction of the plate regarding temperature dependent yield stress model of Case-C excluding the results obtained after cooling step since they are same with ones obtained after releasing step. The maximum longitudinal tensile stress observed in the center line is approximately 230 MPa which is 100 MPa higher than the previous results shown in Figure 15.

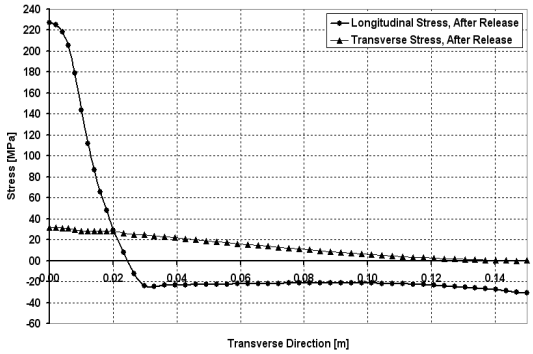


Figure 16. Stress distribution along the transverse direction of the plate for the temperature dependent yield stress model of Case-C

3.2. Results of Model-2:

The second benchmark geometry defined in Section 2.2, i.e. having total heat input of 1050 W, welding speed of 2 mm/s, plate dimensions of 300 x 150 x 1.6 mm, and stringer geometry which is shown in detail at Figure 9, results a maximum temperature of 500.6°C during whole welding step and 20°C, which is the reference temperature, at the end of cooling step of the simulation. Asymmetric longitudinal stress distribution is obtained, as shown in Figure 17, since an asymmetric temperature field is observed due to the addition of the L-shaped stringer on the plate. Longitudinal and transverse stresses are plotted based on the paths given in Table-2.

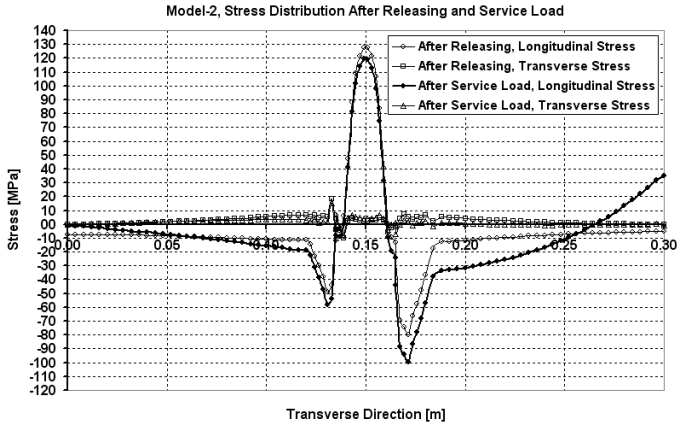


Figure 17. Stress distribution along the transverse direction of the plate for Model-2.

Specific characteristics are noticed in Figure 17, i.e. relatively higher compressive stress just near the right edge of the stringer and tensile stresses along the center line of both the plate and the stringer, after releasing the clamping and the rollers. Similar characteristics are preserved after the application of the service load, which is defined as bending load of 25 kg at the end of the plate while being clamped at the opposite edge. There is 10 MPa decrease in the tension zone which could be considered as a relatively beneficial way of utilization of a welded structure having residual stresses on, while having 20 MPa increase in the compression zone that should also be considered. Although 10 MPa decrease might seem to be low, it should be recognized that for a 300 mm-long plate is quite important since in industrial applications much longer plate-stringer constructions are used; as a result more than 10 MPa decrease in longitudinal stress can be gained by considering the appropriate loading for the welded structure having residual stresses.

Table-2. Description of the Paths for the Stress Distribution in Model-2

PATH	POINTS	X	Y	Z
Path-1	1	0.0000	0.0008	0.1500
	2	0.1350	0.0008	0.1500
Path-2	3	0.1350	0.0024	0.1500
	4	0.1650	0.0024	0.1500
Path-3	5	0.1650	0.0008	0.1500
	6	0.3000	0.0008	0.1500

Addition of temperature dependence for the yield stress results in higher longitudinal stresses as seen in previous sections while preserving the same level of transverse stresses in the transverse direction of the plate. Approximately 10 MPa difference in longitudinal stress, observed after release and service load steps, can also be noticed. It can be noted that the asymmetry in the stress distribution, which is explained as a reason of adding stringer on the thin plate, has been disappeared and same levels of compressive stresses at both sides of the stringer root have been obtained.

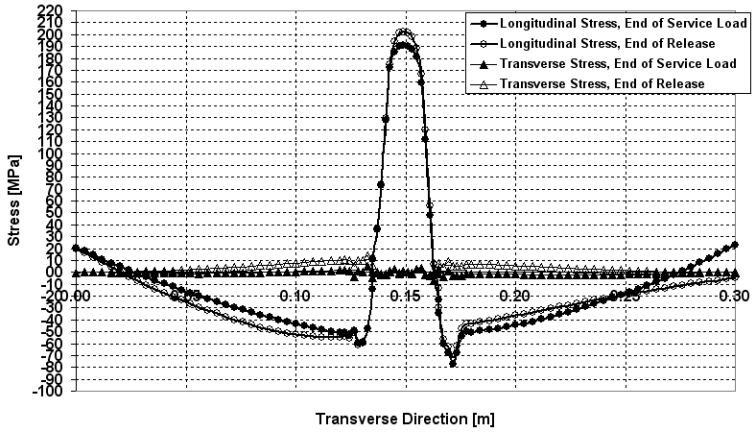


Figure 18. Stress distribution along the transverse direction of the plate for the temperature dependent yield stress case in Model-2

4. Discussion and Conclusion:

The evolution of residual stresses of two friction stir welded (FSW) joints are investigated with 3-dimensional fully-coupled thermo-mechanical models implemented in ABAQUS. In addition to the simulation of the welding and cooling process, the welded structure having residual stress and deformation is also tested for a subsequent load analysis in order to analyze the performance of the joint/structure during exposure to service loads.

These benchmark cases presented hereby could be seen as a frame work for reproducibility, validation and comparison of the results obtained by other modeling groups using different commercial or house-built codes. Besides validation purposes, the general characteristics of the residual stress and deformation due to different kinds of fixture settings which are already used in industrial applications are investigated in a structured way.

Test cases analyzed for the first benchmark model clearly indicates the effect of the fixture settings. In Case-A, higher longitudinal residual stresses are obtained as expected comparing to the other two test cases having different mechanical boundary settings. Case-B also emphasizes the type and placement of the fixtures on the evolution of residual stresses. Although Case-C could be considered to be academic, it is included in order to prepare a complete set of benchmark cases for residual stress validation. Results considering temperature dependent yield stress data have also been included for all cases. General characteristics of the stress distribution are preserved while having higher values due to higher plastic strain.

Second model, which represents rather more complex geometry of an L-shaped stringer, friction stir welded on a thin plate, is used for mainly two reasons; 1) to prepare a benchmark study for residual stresses

in friction stir welding by considering the in-service loads for a relatively complex industrial application, 2) to show that the evaluation of residual stresses can only be done by taking the in-service loads into account.

5. Acknowledgements:
Authors would like to thank EADS for the supplement of an industrial case study which is used as a second benchmark model presented in this paper and CENAERO for the discussion on the need of preparation of this benchmark proposal on the residual stresses in FSW.

6. References:

1. H. Schmidt, J. Hattel, "Modelling heat flow around tool probe in friction stir welding", Sci. Technol. Weld. Joining, 2005, 10, 176-186.
2. H. Schmidt, J. Hattel and J. Wert, "An Analytical Model for the Heat Generation in Friction Stir Welding", Modelling and Simulation in Materials Science Engineering 12, 2004, 143-157.
3. C. Tutum, H. Schmidt, J. Hattel, M. Bendsøe, "Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization", 7th World Congress on Structural and Multidisciplinary Optimization, COEX Seoul, Korea, 21-25 May 2007.
4. H. Schmidt, J. Hattel, "Thermal modeling of friction stir welding", Scripta Materialia 58, 2008, 332–337.
5. F. Marie, D. Allehaux, "Development of FSW for optimum run-out performance", 6th International Symposium on Friction Stir Welding, Saint-Sauveur, Canada, 10-13 October 2006.
6. A. Murphy, W. McCune, D. Quinn, M. Price, "The characterization of friction stir welding process effects on stiffened panel buckling performance", Thin-Walled Structures 45, 2007, 339-351.
7. A-L. Lafly, C. Donne, G. Biallas, D. Allehaux, F.Marie, "Role of residual stresses on Fatigue Crack Propagation of FSW 6056-T78 aluminum joints under various technologies", Materials Science Forum Vols. 519-521, 2006, 1089-1094.
8. M. Prime, T. Gnaupel-Herold, J. Baumann, R. Lederich, D. Bowden, R. Sebring, "Residual stress measurements in a thick, dissimilar aluminum alloy friction stir weld", Acta Materialia 54, 2006, 4013-4021.
9. G. Pouget, A. Reynolds, "Residual stress and microstructure effects on fatigue crack growth in AA2050 friction stir welds", International Journal of Fatigue 30, 2008, 463-472.
10. R. John, K. Jata, K. Sadananda, "Residual stress effects on near-threshold fatigue crack growth in friction stir welds in aerospace alloys", International Journal of Fatigue 25, 2003, 939-948.
11. C. Chen, R. Kovacevic, "Parametric finite element analysis of stress evolution during friction stir welding", Proc. IMechE Vol. 220 Part B: J. Engineering Manufacture, 2006, 1359-1371.
12. Y. Chao, X. Qi, "Thermal and Thermo-Mechanical Modeling of Friction Stir Welding of Aluminum Alloy 6061-T6", Journal of Materials Processing & Manufacturing Science Vol. 7, 1998, 215-233.
13. V. Soundararajan, S. Zekovic, R. Kovacevic, "Thermo-mechanical model with adaptive boundary conditions for friction stir welding of Al 6061", International Journal of Machine Tools & Manufacture 45, 2005, 1577-1587.
14. Z. Feng, X-L. Wang, S. David, P. Sklad, "Modeling of residual stresses and property distributions in friction stir welds of aluminum alloy 6061-T6".
15. D. Richards, P. Prangnell, P. Withers, S. Williams, A. Wescott, E. Oliver, "Geometry Effects When Controlling Residual Stresses in Friction Stir Welds by Mechanical Tensioning", Materials Science Forum Vols. 524-525, 2006, 71-76.
16. D. Richards, P. Prangnell, P. Withers, S. Williams, A. Wescott, E. Oliver, "FE Modelling of Mechanical Tensioning for Controlling Residual Stresses in Friction Stir Welds", Materials Science Forum Vols. 539-543, 2007, 4025-4030.
17. B. Bhide, P. Michaleris, M. Posoda, J. DeLoach, "Comparison of Buckling Distortion Propensity for SAW, GMAW, and FSW", Welding Journal, 2006, 189-195.
18. A. Steuwer, M. Peel, P. Withers, "Dissimilar friction stir welds in AA5083-AA6082: The effect of process parameters on residual stress", Materials Science and Engineering A 441, 2006, 187-196.

C PAPER-III

C. C. Tutum, H. N. B. Schmidt J. H. Hattel and M. P. Bendsøe, *"A preliminary study on optimization of residual stresses in friction stir welding"*, Materials and Manufacturing Processes, 2009 (In press).

A Preliminary Study on Optimization of Residual Stresses in Friction Stir Welding

Cem C. Tutum*[,1], Henrik B. Schmidt[1], Jesper H. Hattel[1], Martin P. Bendsøe[2]

[1] Technical University of Denmark, Department of Mechanical Engineering, Section of Manufacturing Engineering, building 425
[2] Technical University of Denmark, Department of Mathematics, building 303
2800 Kgs. Lyngby, Denmark
*Email: **cctu@mek.dtu.dk**

Abstract

In the present paper, we consider the determination of optimized process parameters in Friction Stir Welding (FSW). That is, the choice of the tool rotational speed and the traverse welding speed have been investigated using genetic algorithms while simultaneously considering thermomechanical issues and production efficiency. The welding process is simulated in a 2-dimensional finite element model which sequentially couples the transient thermo-mechanical behaviour. The thermal-pseudo-mechanical (TPM) heat source model, which is governed by the temperature dependent yield stress, is the main driver for the simulation of the residual stresses in the workpiece. Two optimization scenarios have been investigated. In both of these, the minimization of the residual stresses is the main focus. This is combined with objectives that take welding time and tool wear into consideration, respectively.

Keywords

Friction stir welding, thermal-pseudo-mechanical model, residual stress, constrained multi-objective optimization, nondominated sorting genetic algorithm (NSGA-II)

1. Introduction

Friction Stir Welding (FSW) is an efficient solid-state joining technique that without melting is intended to be used for joining of for example aluminum alloys; also welding of dissimilar materials, which are difficult to weld with traditional welding techniques, can be handled with FSW. Some of the advantages of FSW are improved mechanical properties of the welds, reduced distortion and residual stresses, and benign environmental characteristics [1]. The requirement for lighter and load resistant structures, especially in aerospace and automotive industries, emphasizes the need for investigating the efficient choice of the important process parameters that control the FSW welding procedure.

In the FSW process shown schematically in Figure 1, heat is generated by friction and by the plastic deformation that occurs between the tool and the work piece. The heat flows into the work piece as well as the tool. The amount of heat that is conducted into the work piece influences the quality of the weld, as well as the distortion and residual stresses in the work piece. For example, insufficient heat from the friction could lead to failure of the tool pin since the work piece material is not soft enough. Therefore, understanding the heat flow aspects of the FSW process is extremely important, not only for understanding the physical phenomena but also for improving the process efficiency in order to obtain good weld properties [2-8].

Despite the relatively low heat generation during the FSW process, the rigid clamping used in the process gives rise to high reaction forces acting on the plates so as to avoid shrinkage of the weld center region. As a result, this generates a high amount of yielding in compression at high temperatures, finally resulting in longitudinal and transverse residual stresses. These residual stresses will act as pre-stresses on the final structure and this is critical for the fatigue performance during service [9]. Although the level of residual stresses resulting from the FSW process is somewhat lower as compared to traditional welding techniques

such as fusion welding, it has been shown that the residual stresses play a major role for the fatigue crack growth [10] and buckling behavior [11,12], etc in welded structures obtained using FSW. Thus, in order to understand and control the evolution of thermal and residual stresses arising from the FSW, much work on modeling of the process has been reported in the literature, e.g. [13-22].

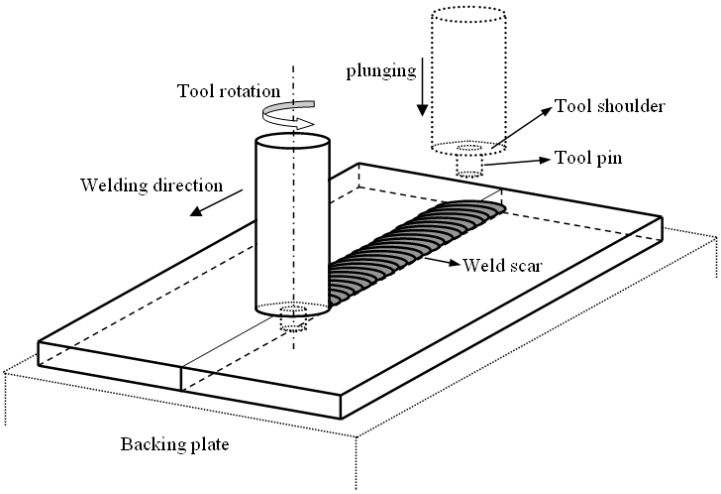


Figure 1. Schematic view of the FSW process

A large number of optimization applications have also been performed in order to control different aspects of the welding process. Murugananth used NSGA-II to obtain an optimal combination of strength and toughness of steel welds where these two objectives are conflicting with each other [23]. Mishra and Debroy used a neural network (NN) that is trained with the simulation results of a heat transfer and fluid flow model together with a real-coded genetic algorithm (GA) in order to obtain desired target values of weld pool penetration and width for a gas tungsten arc welding [24]. An application of the steepest descent algorithm together with an analytical sensitivity analysis has been performed by Michaleris *et al*. to design the thermal tensioning process which utilizes transient temperature gradients by localized heating; the objective was here to minimize welding residual stress and distortion [25]. Also, Bogomolny *et al.* applied shape optimization for the resistance welding process, incorporating surrogate modeling based on a Kriging approximation in order to improve the fatigue performance of the electrodes [26].

The literature relating to optimization of the FSW process consists of both numerical and experimental work. Some of these techniques have also been used to investigate uncertain process parameters that enable model behaviour to be fitted to available experimental data. Application of the Differential Evaluation (DE) technique for reducing the uncertainty associated with specific process parameters, i.e. the friction coefficient, the extent of slip between the tool and the workpiece, the heat transfer coefficient at the bottom of the workpiece, the mechanical efficiency, and the extent of viscous dissipation converted to heat, is studied by Nandan *et al* [27]. Richards *et al.* present detailed numerical studies with experimental validation of the mechanical tensioning during welding and post-welding [13] as well as an investigation of the effectiveness of dynamic cooling in order to control the residual stresses in the FSW process [14]. Investigations of process parameters that can control the residual stresses have also been considered, for example using multi objective genetic algorithm (MOGA-II) applied to a two-dimensional transient thermo-mechanical model [28]. Maximization of the lap joint strength per unit length is examined by Fratini and Corona using a gradient technique (the steepest descent method) together with an experimental procedure resulting in a mechanical performance of the joint equal to the 85.5% of the parent material resistance [29]. Improvement of the tensile strength and thereby the fatigue life of a FSW butt-weld by an experimental study and application of the Taguchi DOE technique on the process parameters, i.e. tool rotational speed and the traverse welding speed, was obtained in ref. [30], [31] respectively. Another application of a gradient-based optimization technique incorporating a simple analytical thermal model is presented by Tutum *et al* [32] in order to obtain a desirable average temperature distribution under the tool shoulder in the FSW process. The same process criterion is studied using space and manifold mapping by Larsen *et al* [33].

In the next section, the applied thermal-pseudo-mechanical (TPM) model, which is the main driver for the coupled thermo-mechanical model used in this study is presented. Following the description of the principles of the TPM model a detailed explanation of the the prediction of the evolution of the residual stresses in the FSW process is emphasized. A brief parameter study has been performed in order to show the main effects of tool rotational speed and traverse welding speed which here are the central process parameters that are considered.. The solution methodology for minimizing the residual stresses together with other complementary issues are then addressed. Next, after expressing the process goals in terms of mathematical formulations, the numerical optimization procedure is solved by modeFRONTIER,; this is a multi-objective optimization and design environment. Results of two optimization scenarios are presented and discussed in detail with emphasis on the consideration of modelling approaches and the exploitation of alternative trade-off solutions.

2. Thermo-mechanical model

The de-coupled thermo-mechanical model of the FSW process used in this work consists of a transient thermal model and a quasi-static elasto-plastic mechanical model, which is simulated by utilizing the commercial finite element software ANSYS. In order to facilitate the automation of the optimization procedure both the models are implemented by means of the Parametric Design Language (APDL) of ANSYS.

The model represents the welding of two flat plates by considering the bead on plate. Due to symmetry assumptions, e.g. neglecting the asymmetric shear layer and the asymmetric heat source, only one of the plates is modeled. The dimension of the workpiece is 300 mm x 100 mm x 3 mm. This means that the thermally induced out-of-plane stresses will be negligible and a plane-stress analysis is reasonable. The temperature independent material properties, which are given in Table-1, do not correspond to any specific commercial aluminum alloy but should be understood as a representative aluminum benchmark material. This assumption is in agreement with the study presented by Zhu and Chao [34], which concludes that the temperature dependent yield stress has a significant effect on the residual stress and distortion, and except for this yield stress, using material properties at room temperature gives reasonable predictions of the transient temperature fields, the residual stresses and distortions. The temperature dependence of the yield stress for both thermal and mechanical analyses is shown in Figure 2 as a linear function of temperature with a negative slope, decreasing from 200 MPa to 0 MPa at 20 °C and 500 °C, respectively. There is an exception for the mechanical analysis that 475 °C is chosen to be the cut-off temperature and the yield stress is kept constant at 10 MPa above this temperature. This engineering simplification provides a substantial convenience for controlling computational cost as it reduces the nonlinearities that do not have a significant effect on the global behavior of the thermo-mechanical model [34].

Table-1. Temperature Independent Material Properties of Benchmark Material

Heat Conduction, k	[W/(m.K)]	160
Heat Capacity, c_p	[J/(kg.K)]	900
Young's Modulus, E	[GPa]	70
Tangent Modulus, E_t	[GPa]	7
Thermal Expansion Coefficient, α	[1/K]	$2.3x10^{-5}$
Density, ρ	[kg/m^3]	2700

Due to the very low contribution to the heat generation coming from the tool pin only the tool shoulder is considered in the heat source. The diameter of the tool shoulder is 20 mm. The mechanical effects of the tool are not included, and thus residual stresses are assumed to be primarily a function of the thermal load history [5,20]. The moving heat source starts and stops at 50 mm away from the left and the right edges of the plate, respectively. The SHELL 131, 4-Node layered thermal shell element is used for the transient thermal analysis while the PLANE 182, 2-D 4-node structural solid element is used for the quasi-static mechanical analysis and the same structured finite element mesh is used in both cases.

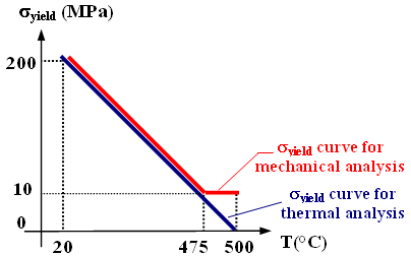


Figure 2. Temperature dependent yield stress curves considered for the thermal (red) and mechanical (blue) analyses, respectively

The accuracy of the thermal and mechanical simulation using shell and plane stress models has been compared with a three-dimensional solid linear 8-node element model. Here, SOLID 70 elements are used for the transient thermal analysis and SOLID 45 for the quasi-static mechanical analysis, respectively. Figure 3(a) shows the thermal profiles obtained at the middle of the plate along the transverse direction while the tool is traversing with 2 mm/s and rotating with 300 rpm. There is at most a 1°C difference between two thermal models.

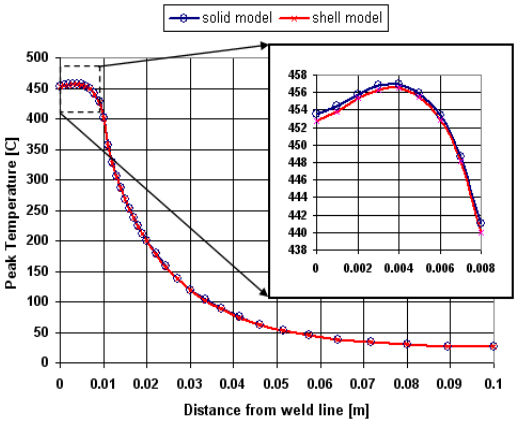


Figure 3. (a) Comparison of temperature profiles obtained by shell (red) and solid (blue) models

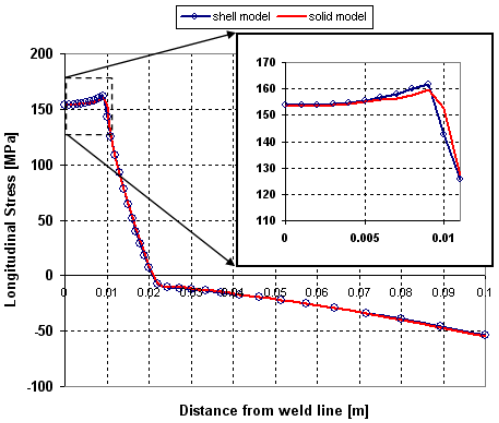


Figure 3. (b) Comparison of longitudinal stress profiles obtained by shell (red) and solid (blue) models

Longitudinal stress profiles obtained with the two models are also in good agreement, e.g. the maximum difference is 3 MPa. Thus, it can be concluded that for the present case the sequentially coupled shell and plane-stress models can be used for the optimization studies while being both accurate and computational efficient.

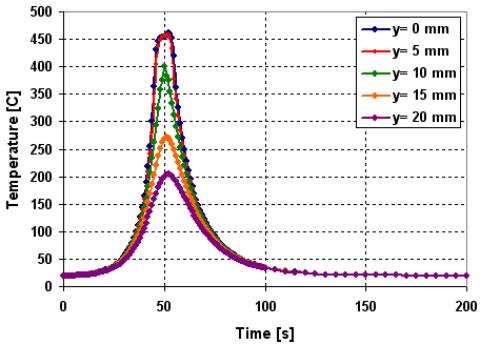


Figure 4. (a) Thermal history

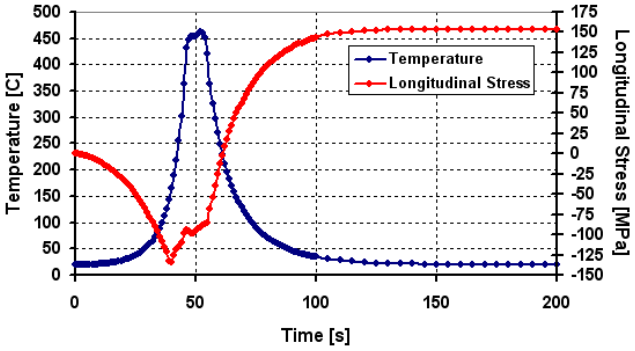


Figure 4. (b) Transient longitudinal stress and temperature profiles as a function of time

Transient thermal profiles at different distances in 5 mm increments from the weld line are shown in Figure 4(a); here the heat source is located at the middle of the plate. Due to the high heat transfer from the workpiece to the backing plate, the thermal profiles are settled at 20 °C at the end of the cooling session.

Figure 4(b) shows the same transient thermal profile along the weld line together with the transient longitudinal stress; this is the main contribution to the residual stresses evolved locally along the centre line, i.e. in the tension zone.

2.1. Thermal-pseudo-mechanical (TPM) model

The TPM model, which is proposed by Schmidt and Hattel [2,35], is in this work applied in order to simulate the thermal effect of the welding process. As the name indicates, although being purely thermal, the model still takes some mechanical effects into account without actually solving for any mechanical field. The main advantage of this is that the obvious simplicity of thermal models is still maintained in the TPM model. The temperature dependent yield stress of the workpiece material is the main driver for the heat source and is part of the numerical solution itself. This makes the model nonlinear. However, the computational cost is still affordable as compared to thermal calculations that are based on analytical heat source models and it is only a fraction of that of a full thermo-mechanical model. The good correlation of the thermal profiles obtained from experiments with those obtained from the TPM model serves as a validation of the model as well [2,35].

For the analysis problem, the heat conduction equation, Eq. (1) below, has to be solved together with an appropriate set of initial and boundary conditions in order to obtain the transient thermal field. The equation is a partial differential equation of the form

$$\rho c_p \frac{\partial T}{\partial t} = \nabla(k \nabla T) + q \tag{1}$$

where ρ denotes the material density, c_p the specific heat capacity, T the temperature, k the thermal conductivity, and q the volumetric heat source term. The latter is in this case zero because the heat generation is modelled as a surface heat flux from the tool shoulder (without the tool probe). The surface heat flux, q, which is a function of the tool radius and the temperature, is given as follows

$$q(r,T) = \omega r \tau(T) = \left(\frac{n\,2\pi}{60}\right) r \frac{\sigma_{yield}(T)}{\sqrt{3}}, \quad \text{for } 0 \le r \le R_{shoulder} \tag{2}$$

where n is the tool revolutions per minute, r is the radial position originating in the tool center, d is the thickness of the plate and $R_{shoulder}$ is the tool shoulder radius. Finally, the temperature dependent yield stress σ_{yield} is defined as

$$\sigma_{yield}(T) = \sigma_{yield,ref}\left(1 - \frac{T - T_{ref}}{T_{melt} - T_{ref}}\right) \tag{3}$$

where $\sigma_{yield,ref}$ is the yield stress at room temperature, T_{ref} is 20 °C and T_{melt} is the solidus temperature (500°C). Once temperatures reach the solidus temperature, i.e. T becomes equal to T_{melt} in Eq. (3), the "self stabilizing effect" causes the heat source to "turn itself off", i.e. the material looses all its resistance, and the heat generation decreases automatically due to thermal softening. The effect of the thermal contact with the backing plate is modelled by an equivalent heat transfer coefficient of 700 W/m^2K at the bottom of the workpiece and with an an ambient temperature of 20 °C.

Figure 5(a) shows the contour plots of the resulting temperature field of the symmetric models (with increments of 47.2 °C) for a chosen welding speed of 1 mm/s and 10 mm/s, respectively, for a rotational speed of 1000 rpm. Figure 5(b) shows a parameter study for the thermal profiles along the transverse direction and at the middle of the plate with two welding speeds, i.e. 1 and 10 mm/s, and two rotational speeds, i.e. 100 and 1000 rpm. These results clearly show the main characteristics of the TPM model:

i) Higher welding speed for a fixed rotational speed yields lower temperatures in general, but the decrease in peak temperature vanishes for rotational speeds that are higher than 500 rpm.

ii) Higher rotational speed for a fixed welding speed results in higher peak temperature and the incease in peak temperature vanishes for higher rotational speeds.

iii) The gradients in thermal profiles along the transverse direction of the plate become higher with increasing welding speed, while more uniform and wider thermal profiles are obtained for the lower welding speeds.

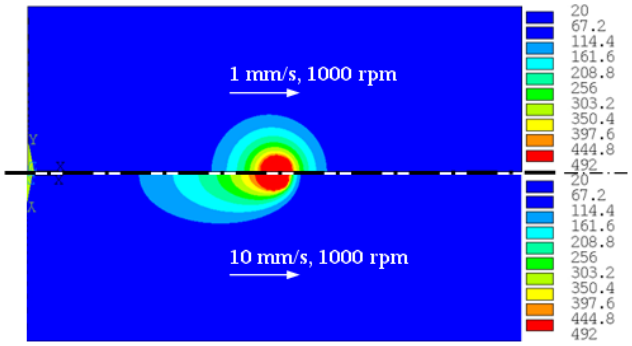


Figure 5. (a) Contour plots of the temperature fields for rotational speed of 1000 rpm for each traverse welding speed of 1mm/s (top) and 10mm/s (bottom), respectively.

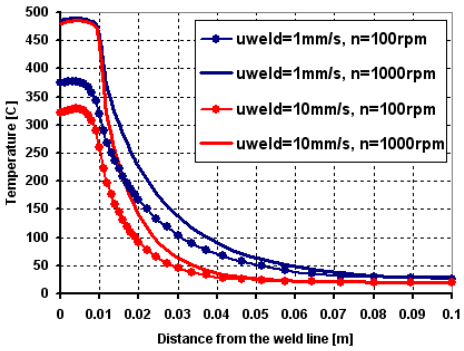


Figure 5. (b) Temperature profiles for different process variables along transverse direction

2.2. Mechanical model

For evaluation of residual stress fields, a standard mechanical model based on the solution of the static force equilibrium equations is used [36-38], i.e., one solves

$$\sigma_{ij,i} + p_j = 0 \tag{4}$$

where p_j is the body force at any point within the plate and σ_{ij} is the stress tensor. The well-known Hooke's law and a linear decomposition of the strain tensor as well as small strain theory are applied:

$$\begin{aligned} \sigma_{ij} &= \frac{E}{1+\nu}\left\{\frac{1}{2}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) + \frac{\nu}{1-2\nu}\delta_{ij}\delta_{kl}\right\}\varepsilon_{kl}^{el} = E_{ijkl}^{el}\varepsilon_{kl}^{el} \\ \varepsilon_{kl} &= \varepsilon_{kl}^{el} + \varepsilon_{kl}^{pl} + \varepsilon_{kl}^{th} \\ \varepsilon_{kl} &= \frac{1}{2}(u_{i,j} + u_{j,i}) \end{aligned} \tag{5}$$

The plastic strain is based on standard J2-flow theory with a temperature dependent yield surface. The conditions during yielding are described by the associated flow rule that consists of the von Mises yield surface and the consistency condition, i.e.

$$\begin{aligned} \dot{\varepsilon}_{ij}^{pl} &= \dot{\lambda} s_{ij} \\ f(\sigma_e, \varepsilon_e^{pl}, T) &= \sigma_e - \sigma_Y(\varepsilon_e^{pl}, T) \le 0 \\ \dot{f} = \dot{f}(s_{ij}, \varepsilon_e^{pl}, T) &= \frac{\partial f}{\partial s_{ij}}\dot{s}_{ij} + \frac{\partial f}{\partial \varepsilon_e^{pl}}\dot{\varepsilon}_e^{pl} + \frac{\partial f}{\partial T}\dot{T} = 0 \end{aligned} \tag{6}$$

This material description is considered to be adequate for the FSW simulation where the first order effect is the yield strength of the material as a function of temperature.

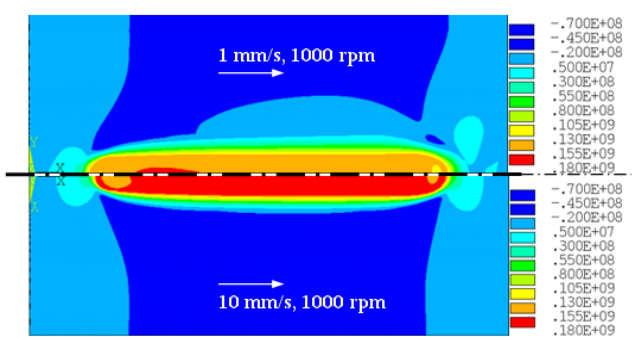


Figure 6. (a) Contour plot of the longitudinal stress field with increments of 22 MPa.

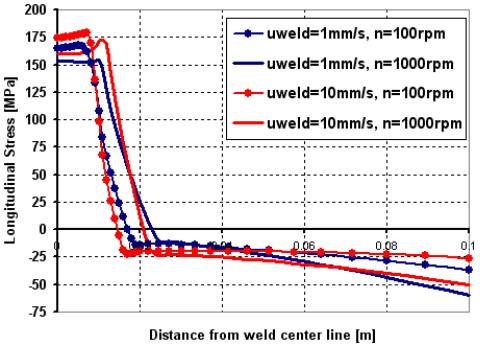


Figure 6. (b) Residual normal stress in longitudinal direction as a function of distance from weld line

Figure 6(a) shows the contour plots of the resulting longitudinal stress field of the symmetric models (isotherms in 25 MPa) for a chosen welding speed of 1 mm/s and 10 mm/s, respectively, for a rotational speed of 1000 rpm. A parameter study regarding longitudinal stress profiles along the transverse direction and at the middle of the plate with two welding speeds, i.e. 1 and 10 mm/s, and two rotational speeds, i.e. 100 and 1000 rpm, is presented in Figure 6(b). These results clearly show the main characteristics of the model of the residual stresses in the FSW process:

i) Higher welding speed for a fixed rotational speed results - in general - in higher stress levels in the tension zone.

ii) Higher rotational speed for a fixed welding speed yields lower peak residual stress.

iii) The gradients in residual stress profiles along the transverse direction of the plate become higher with increasing welding speed, while the model results in wider residual stress profiles for lower welding speeds.

3. Optimization Methodology

In the following sections we present a study of two multi-objective optimization problems in the FSW process. These consider the minimization of the residual stresses in the workpiece together with the maximization of the production efficiency (reducing welding time and wear path), respectively. These two objectives are conflicting and techniques to deal with this issue are also considered.

The optimization procedure, which includes process integration of the ANSYS software and the optimization algorithm (NSGA-II), is handled by applying modeFRONTIER. The optimization cycle is initiated by creating an initial population of 16 pre-chosen Design-of -Computational-Experiments (DOE), for the considered process variables, i.e. the tool rotational speed (revolutions per minute), n, and the traverse welding speed, u_{weld}. The FSW thermal and mechanical simulations, which are built by using APDL in ANSYS, are coupled in a sequential way by execution of parametric input files in a batch mode. The design variables are updated by the optimization algorithm and are read by the thermal analysis. Then the peak temperature obtained at the end of the welding session is saved in order to be used as a thermal constraint together with the transient temperature field results in order to yield the thermal strains for the mechanical analysis. The mechanical analysis gives as a result the maximum longitudinal stress value at the middle of the plate in the transverse direction, and this is used as an objective to be minimized. This optimization cycle run until the stopping criterion, i.e. the total number of generations, is reached. The flow chart of the optimization procedure is shown in Figure 7.

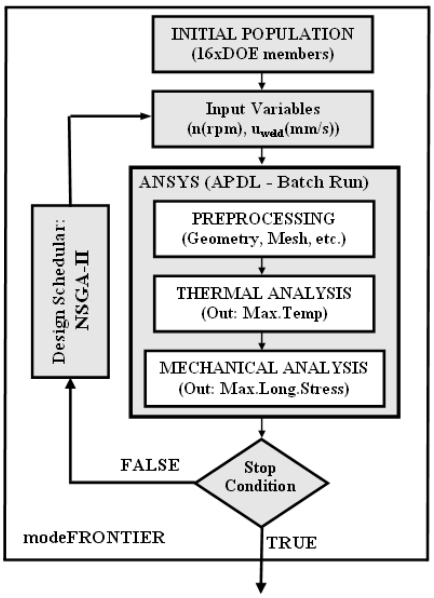


Figure 7. Flow chart of the optimization problem

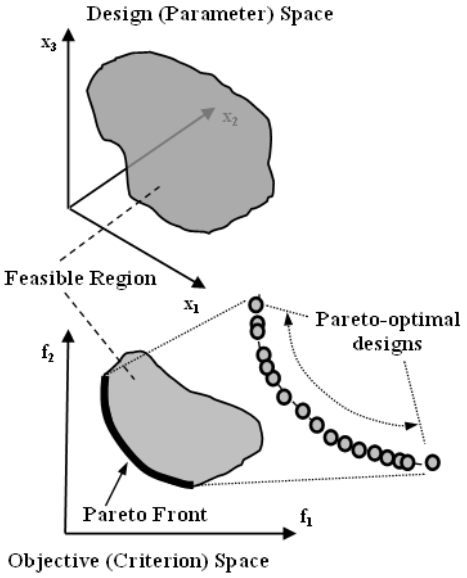


Figure 8. Design and objective space associated with multi-objective optimization problem

3.1. Multiobjective Optimization (MOO) Problem

Most engineering design activities require a solution of multi-disciplinary and multiobjective optimization (MOO) problems which usually deal with conflicting design objectives. The consideration of multiple conflicting objectives results in a number of alternative trade-offs, called Pareto-optimal solutions, of which none can be said to be better than any other without any additional information about the problem under consideration. This concept can be formulated as follows

$$(\forall i\, f_i(x) \leq f_i(y)) \cap (\exists\, j : f_j(x) < f_j(y)) \tag{7}$$

where f_i is the *i-th* objective function in the feasible region and the minimization of the two-objective problem, which is schematically shown in Figure 8, is considered for convenience. In this context, Eq. (7) describes that at least in one objective, design x is better than design y, while in others they might be equal; this can also be expressed as the solution x dominates the solution y in the parlance of MOO. Since the concept of domination allows a way to compare solutions with multiple objectives, most MOO algorithms use this domination concept to search for non-dominated solutions, i.e. the ones that constitute the Pareto-front [39-41]. Population-based evolutionary algorithms have in the last decade become more and more popular as they effectively can utilize distributed computing facilities. Also, they are typically able to provide a larger spectrum of Pareto-optimal solutions without any additional problem specification. The availability of trade-off solutions, representing varying preference levels between chosen objectives, makes it easier for a decision-maker to choose a particular solution for implementation [39-41].

3.2. Nondominated Sorting Genetic Algorithm (NSGA-II)

NSGA-II is a multi-objective evolutionary algorithm (MOEA) proposed by Deb *et al* [42], which enables finding well-spread multiple Pareto-optimal solutions for a MOO problem by incorporating three substantial features, i.e., elitism, non-dominated sorting, and diversity preserving mechanism. Elitism, which is a genetic operator used to preserve the good solutions through the generations, is also shown to increase the convergence speed of a MOEA [43].

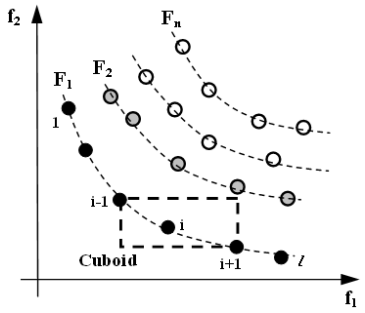


Figure 9. (a) The crowding distance calculation [42]

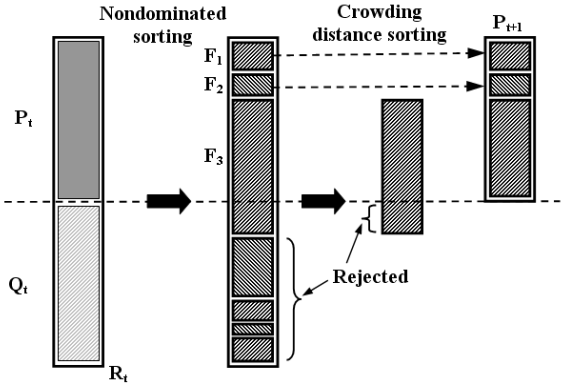


Figure 9. (b) Schematic view of the NSGA-II procedure. (From Deb *et al.*, [42])

The preservation of well-distributed (diverse nondominated) solutions is maintained by the crowding distance operator. The crowding distance, which is calculated frontwise, is a measure of how close an individual is to its neighbors. In order to get an estimate of the density of solutions surrounding a particular solution in the population, the average distance of two points on either side of this point along each of the objectives is calculated. This quantity d_i serves as an estimate of the perimeter of the cuboid formed by using the nearest neighbors as the vertices. A schematic view of the procedure for the computation of the crowding distance for an i^{th} individual is represented in Figure 9a, and formulated as

$$d_i = d_i + \frac{f_m^{(I_{i+1}^m)} - f_m^{(I_{i-1}^m)}}{f_m^{max} - f_m^{min}}, \quad for\ i=2,\ ...,l-1 \tag{8}$$

where $I(i)$ is the solution index of the i^{th} individual in the (worse order) sorted list, m is the number of objectives, l is the number of the individuals in that particular front. The first and the last individual is assigned an infinite distance in order to emphasize their fitness. A large average crowding distance will result in better diversity in the population.

The schematic view of NSGA-II on a flowchart is shown in Figure 9b. The procedure starts with an initialization of the parent population, P_0, followed by sorting the members of the population based on nondomination into each front, F_i. Each individual is assigned two fitness values, i.e. the one related to its nondomination level (front or rank, i.e. F_i refers to front with rank i in Figure 9) and the crowding distance. Parents are selected by using binary tournament selection based on these two measures. An individual is selected if its rank is less than the other, since the minimization is assumed, or if the crowding distance is greater than the other. The selected population generates off-springs, Q_t, by the application of usual genetic operators, i.e. crossover, mutation, etc. [44]. Since all the parent and the offspring individuals are included in the combined population, R_t, elitism is ensured. In order to reduce the size of the population of the next generation from 2N to N, nondomination sorting is applied and the individuals with relatively lowest possible ranks are accomodated until the size of the population exceeds N. Then the individuals in front F_i are selected based on their crowding distance. This cycle is repeated until the stopping criterion, i.e. the total number of generations, is met. For details, please consult the original study [42].

3.3. Problem Statement-1

The first optimization problem here is stated as the goal of finding the friction stir welding process parameters, i.e. tool rotation speed and traverse welding speed, which provide a set of trade-off solutions for the minimization of two conflicting objectives. These are the residual stresses, which are measured at the middle of the plate along the transverse direction, and the welding time that can also be stated equally as the maximization of the traverse welding speed. The optimization problem is constrained by the process-specific thermal constraints, which are given as the upper and the lower bounds on the peak temperatures in

the workpiece. The lower bound of 420°C on the peak temperature represents the need for easy traversing of the tool, i.e. to minimize the tool loads along the weld line by contributing to thermal softening of the workpiece material. The upper bound of 480°C is defined in order to consider the tool life and the workpiece properties which are affected by hot welding conditions. This constrained multi-objective problem can then be expressed in mathematical terms as

$$\begin{gathered} \text{Minimize: } f_1(\mathbf{x}) = \sigma_{x,max} \\ \text{Maximize: } f_2(\mathbf{x}) = u_{weld} \\ \text{subject to: } g_1(\mathbf{x}) = 420°C \le T_{peak} \\ g_2(\mathbf{x}) = T_{peak} \le 480°C \\ x = \begin{Bmatrix} u_{weld} \\ n \end{Bmatrix}, \quad \begin{matrix} u_{weld} = 1, 2, ..., 10 \text{ (mm/s)} \\ n = 100, 200, ..., 1000 \text{ (rpm)} \end{matrix} \end{gathered} \tag{9}$$

where x represents the design variable vector, i.e. u_{weld}, the traverse welding speed that is changing from 1 mm/s to 10 mm/s in 1 mm/s increments and n, the tool rotational speed which varies from 100 rpm to 1000 rpm in 100 rpm increments (this results in 10 discrete values in each design variable), $\sigma_{x,max}$ defines the peak longitudinal stress, and T_{peak} is the peak temperature in the workpiece.

3.4. Problem Statement-2

The second optimization problem, which is a slightly modified version of the first case, considers the minimization of the so-called wear path of an arbitrary point on the tool shoulder, and it also takes the minimization of the residual stresses on the workpiece into account:

$$\begin{gathered} \text{Minimize: } f_1(\mathbf{x}) = \sigma_{x,max} \\ \text{Minimize: } f_2(\mathbf{x}) = \omega/u_{weld} \\ \text{subject to: } g_1(\mathbf{x}) = 420°C \le T_{peak} \\ g_2(\mathbf{x}) = T_{peak} \le 480°C \\ x = \begin{Bmatrix} u_{weld} \\ n \end{Bmatrix}, \quad \begin{matrix} u_{weld} = 1, 2, ..., 10 \text{ (mm/s)} \\ n = 100, 200, ..., 1000 \text{ (rpm)} \end{matrix} \end{gathered} \tag{11}$$

The wear path at the radius r at any point on the tool/workpiece interface can be approximated by the following expression,

$$L_{path} = v_{circumf} t_{weld} = \omega r t_{weld} = \omega r \frac{L_{weld}}{u_{weld}} = (r L_{weld}) \frac{\omega}{u_{weld}} = C \frac{\omega}{u_{weld}} \tag{10}$$

where C is a parameter that will remain constant for the point under consideration. Minimization of the ratio of rotational speed/welding speed corresponds to maximizing the tool advance per revolution. Other studies (described below) have found weld property variations as a function of either welding speed/rotational speed, i.e. the weld pitch, rotational speed/welding speed or pseudo heat index. However, none of these are based on an argumentation addressing the wear criterion as presented in this paper, neither are they used for optimization purposes, but more for process characterization and understanding.

In the literature it has been reported that, in general, the specific weld energy (Joules/meter) correlates with the weld pitch, i.e. the energy per unit length of weld increases as the weld pitch decreases [45]. The peak temperature in the weld zone increases with an increase in the ratio of tool rotation rate/traverse speed [46]. The recrystallized grain size is reduced by decreasing the ratio of tool rotation rate/traverse speed which then results in an increase in both the degree of deformation and the peak temperature of the thermal cycle [47]. In [48], the pseudo-heat index ($w=\omega^2/u_{weld}$) was used as part of a generalized relationship in order to find the peak temperature for several aluminum alloys. It was found that an increase in the ratio of rotational speed/traverse speed (ω/u_{weld}) leads to a slight decrease in yield and ultimate strength of the stir

zone and transitional zone. It was also observed that increasing this ratio increases the weld nugget size and decreases the incomplete root penetration [49].

4. Results

The initial population for the NSGA-II algorithm that is used for both cases is chosen as a modified Full Factorial Design with 4-levels (*n*: 100, 400, 700, 1000 rpm and u_{weld}: 1, 4, 7, 10mm/s) resulting totally in 16 designs. The crossover and the mutation probabilities of 0.9 and 1.0, respectively, are chosen for running a total of 20 generations giving in total a number of 320 solutions.

4.1. Results of the 1st Problem Statement

The solution of the first optimization case, which is defined in section 3.1, is presented in both the design and the criterion space in the following figures. Some of the designs out of a total of 320 designs are overlapping due to relatively coarse discretization of the chosen design variables and also the selection operator. This lies in the nature of the genetic algorithm that implies the survival of some designs without evolution. Figure 10(a) and 10(b) represent feasible and unfeasible designs with dark and fair colors, respectively; constituting 49 different design points out of 100 (we use a 10x10 discretization). The surface in each figure, i.e. the peak temperature and the peak residual stress, is constructed by 16 DOE points which are evaluated as an initial population for NSGA-II. It can be clearly seen from these figures that the feasible region, which can also be called the robust process parameter region in this case, is defined by *n*-values in the interval between 200 and 400 rpms. Due to the random evolution strategy of EA, some of the feasible solutions (4, 6, 8 mm/s with 200, 300 and 400 rpms), are missing but the quality of these can be estimated from the surrounding solutions which are positioned on a linear varying region. Although the coarse discretization results in few missing solutions, it is advantageous for having an overall idea regarding the optimal feasible process frame with a moderate computational cost. In addition to this, the purpose of this study is to focus on the optimization methodology to find and discuss alternative trade-off solutions for minimization of welding residual stresses, otherwise not to conclude precise values. Having such solutions provides the engineer or manufacturer practical insight about the relationship among process variables corresponding to the Pareto-optimal solutions.

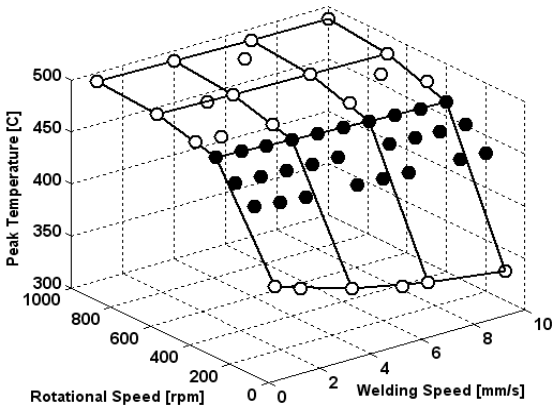


Figure 10. (a) Peak Temperature vs. design variables

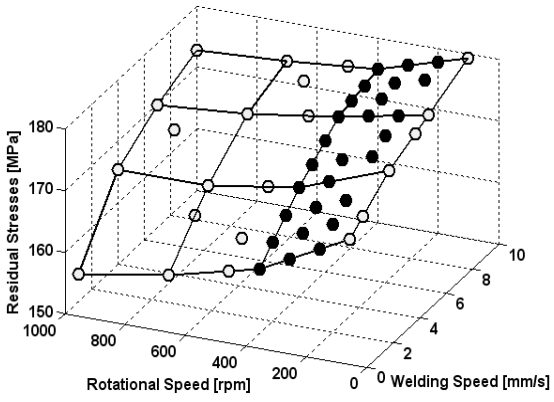


Figure 10. (b) Residual stress vs. design variables

The objective space that is constructed by minimization of the residual stresses and maximization of the welding speed is shown in Figure 11(a). Most of the designs lie close to the Pareto-front, which is shown in Figure 11(b). This is due to the low sensitivity of the *n* parameter on the residual stresses as defined in the underlying thermal model, i.e. the TPM model for a given welding speed. If the minimum-residual stress solution is emphasized for the MOO problem, i.e. choosing weightings of 1.0 for the objective of minimum-residual stress and 0.0 for the objective of maximum-welding speed, the combination of 1mm/s and 400rpm would be chosen. If the other extreme solution, i.e. with opposite weightings, is considered, the combination of 10mm/s and 400rpm would be preferred. In the case where one is looking for a 70-30%

trade-off solution for the same objectives, respectively, it is not clear how to estimate the optimal combination of the process variables. Due to this desired preference in objectives, the solution would be expected to be more similar to the minimum residual stress solution than the maximum welding speed. It is important to note that there are some different solutions satisfying such trade-off, but there is only one which is the optimum, i.e. 4mm/s and 400rpm in this case. In other words, that solution makes the optimum trade-off, meaning that for a particular amount of sacrifice in one objective, the optimum solution will correspond to the maximum gain in the other objective [41].

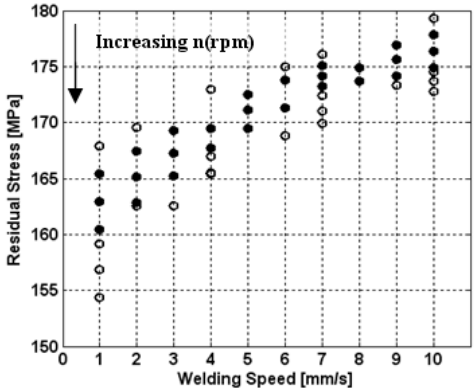


Figure 11. (a) Objective Space of the solution

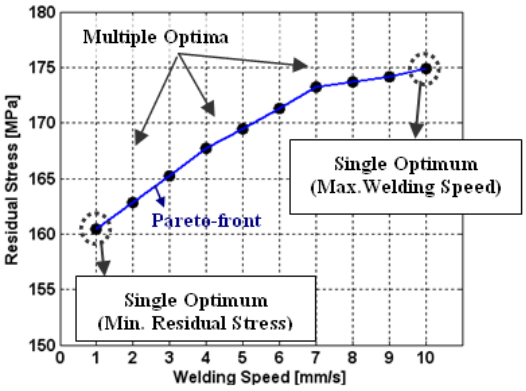


Figure 11. (b) Pareto Set of the solution

All Pareto-optimal solutions correspond to the maximum feasible rotational speed of 400rpm while having different welding speeds varying from 1 mm/s to 10 mm/s. The Pareto-front shown in Figure 11(b) gives an idea of ranking the alternative trade-off solutions depending on the available working conditions. If a manufacturer is able to use a standard milling machine instead of an advanced FSW machine and can afford using simple tool designs with low welding speed, he would probably not dare to go from 1 to 7mm/s in welding speed because the residual stresses yielded per unit increment in welding speed would cost higher comparing to those at higher welding speeds. The amount of sacrifice of the manufacturer relatively depends on the welding speed while one can keep the rotation speed between 200 and 400 rpms.

4.2. Results of the 2nd Problem Statement

The second optimization case defined in section 3.2 is presented below in objective space only since the design domain is kept constant while the solutions of the first optimization case are rearranged by taking the tool rotational speed into account as shown in Figure 12(a, b).

From the Pareto front in Figure 12(b) it is seen that at high welding speeds the second objective, i.e. minimum wear path, becomes dominant, whereas for low welding speeds the first objective of minimum residual stress is dominant.

It is also interesting to note that the wear path criterion divides the parameter space into two intervals based on whether the welding speed is higher or lower than 5 mm/s. Thus, it is seen on Figure 12(b) that for welding speeds above this value the Pareto solutions all have a rotational speed of 200 rpm, whereas for welding speeds below this value the rotational speed varies between 200 and 400 rpm. So, as compared to the previously found solutions, which are listed in section 4.1, the rotational speed is reduced from 400 rpm to 200 rpm at high welding speeds resulting in a slightly higher sacrifice in the first objective, i.e. minimization of the residual stresses.

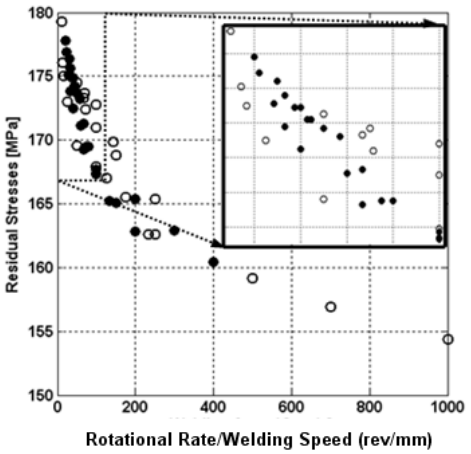


Figure 12. (a) Objective Space of the solution

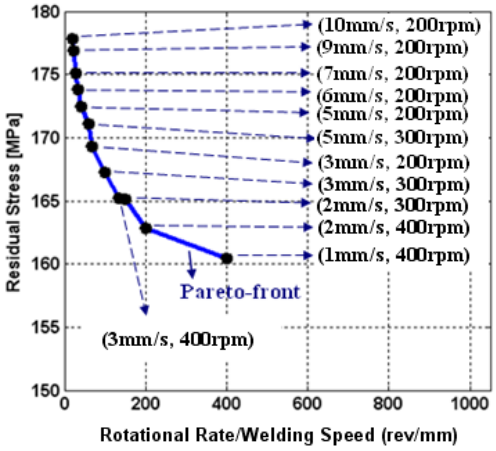


Figure 12. (b) Pareto Set of the solution

The discussion for the first problem case can be seen to hold for this case as well. The single optimum solution for the lowest welding speed is kept the same, i.e. being combined with a rotational speed of 400 rpm again, while the single optimum solution for the highest welding speed is rearranged in a way that it is combined with the lowest feasible rotational speed, i.e. 200 rpm.

5. Conclusion and Discussions

In this work a multi-objective optimization application for the FSW process has been presented. Two cases have been considered in which minimization of the residual stresses have been performed in combination with a maximization of the welding speed and a minimization of the wear path, respectively. The tool rotational speed and the traverse welding speed has been used as design variables. In addition to the description of the process goals, process-specific thermal limitations have been added to the optimization problem in order to take the tool loads and tool life issues into account; this has been formulated in terms of lower and upper bounds on the peak temperature. The results show that a tool rotation speed of 400 rpm is desired for the first problem while 200 rpm is preferred in general for the second case. The second optimization case definitely shows that the problem of minimization of the residual stresses can not be considered independently of the tool traversing speed, which is expected. Depending on the Pareto designs, ranking of the trade-off solution alternatives has been discussed in addition to looking at the optimization problem from a manufacturer point of view.

One could discuss whether it is enough to run 20 generations while having a population size of 16, resulting in a total of 320 design combinations in order to find the true Pareto-optimal solutions. It can be clearly seen that the feasible region is almost fully covered but in case of “what-if” studies it would not be sufficient. Then it could also be argued whether making enumeration makes sense for these highly demanding analyses. In this preliminary optimization study, the design variables are roughly discritized in order to obtain a general overview before going deep into the complex nature of the modeling and resulting optimization studies. The main focus is on finding a methodology to optimize process variables, i.e. welding speed and rotational speed, for multiple objectives under process related constraints.

The heat generation results at high traverse welding speeds are probably not estimated with high enough fidelity as only the thermal effect is taken into account; rather one should consider the highly complex phenomena in the mechanical deformation field as investigated in [50]. At higher welding speeds, e.g. 10mm/s, there might be a need for higher rotational speeds in order to avoid cold weld conditions.

It should be emphasized that since this is a preliminary study of an MOO problem regarding residual stresses in FSW, microstructural changes due to the thermal history have not been included yet. This is of course a limitation of the model since the microstructural evolution determines the final mechanical properties of the weld and in turn the service load performance of the part.

Another issue of the model that could be discussed is the the objective of minimization of the peak residual stress. It might be more relevant to use the "shrinkage force", which can be evaluated as the integration of the area under the residual longitudinal stress profile in tension along the transverse direction of the work piece. This way, the width and the gradients of the stress profile in the tension zone would be taken into account also, resulting in a more relevant optimization problem seen from an industrial view point.

Despite these limitations, the procedure put forward is generic and it would be manageable to include for example the microstructures and the shrinkage force in the analysis.

6. References

[1] Mishra, R.S., Ma, Z.Y., Friction stir welding and processing, Materials Science and Engineering R, Vol. 50, 2005, pp. 1-78

[2] Schmidt, H., Hattel, J., Thermal modelling of friction stir welding, Scripta Materialia, 58, 2008, pp. 332-337

[3] Schmidt, H., Hattel, J., Modelling heat flow around tool probe in friction stir welding, Science and Technology of Welding and Joining, Vol. 10, No. 2, 2005, pp. 176-186

[4] Zhu, X.K., Chao, Numerical simulation of transient temperature and residual stresses in friction stir welding of 304L stainless steel, Journal of Materials Processing Technology, Vol. 146, 2004, pp. 263-272

[5] Chao,Y.J., Qi, X., Thermal and Thermo-Mechanical Modeling of Friction Stir Welding of Aluminum Alloy 6061-T6, Journal of Materials Processing & Manufacturing Science, Vol. 7, 1998, pp. 215-233

[6] Song, M., Kovacevic, R., Thermal modeling of friction stir welding in a moving coordinate system and its validation, International Journal of Machine Tools & Manufacture, Vol. 43, 2003, pp. 605-615

[7] Chen, C.M., Kovacevic, R., Finite element modeling of friction stir welding – thermal and thermomechanical analysis, Journal of Machine Tools & Manufacture, Vol. 43, 2003, pp. 1319-1326

[8] Hattel, J.H., Schmidt, H.B., Tutum, C.C., Thermomechanical modelling of Friction Stir Welding, 8th International Conference on Trends in Welding Research Conference, ASM, June 2008.

[9] James, M.N., Hughes, D.J., Chen, Z., Lombard, H., Hattingh, D.G., Asquith, D., Yates, J.R., Webster, P.J., Residual stresses and fatigue performance, Engineering Failure Analysis, Vol. 14, 2007, pp. 384-395

[10] Bussu, G., Irving, P.E., The role of residual stress and heat affected zone properties on fatigue crack propagation in friction stir welded 2024-T351 aluminum joints, International Journal of Fatigue, Vol. 25, 2003, pp. 77-88.

[11] Murphy, A., McCune, W., Quinn, D., Price, M., The characterization of friction stir welding process effects on stiffened panel buckling performance, Thin-Walled Structures, Vol. 45, 2007, 339-351

[12] Bhide, B., Michaleris, P., Posoda, M., DeLoach, J., "Comparison of Buckling Distortion Propensity for SAW, GMAW, and FSW", Welding Journal, 2006, 189-195.

[13] Richards, D.G., Pragnell, P.B., Williams, S.W., Withers, P.J., Global mechanical tensioning for the management of residual stresses in welds, Materials Science and Engineering A, Vol. 489, 2008, pp. 351-362

[14] Richards, D.G., Prangnell, P.B., Withers, P.J., Williams, S.W., Nagy, T., Morgan, S., Simulation of the Effectiveness of Dynamic Cooling for Controlling Residual Stresses in Friction Stir Welds, 7th International Symposium Friction Stir Welding, TWI, May 2008

[15] Bastier, A., Maitournam, M.H., Roger, F., Dang Van, K., Modelling of the residual state of friction stir welded plates, Journal of Materials Processing Technology, Vol. 200, 2008, pp. 25-37

[16] Feng, Z., Wang, X., David, S.A., Sklad, P., Modeling of residual stresses and property distributions in friction stir welds of aluminum alloy 6061-T6, Science and Technology of Welding & Joining, Vol. 12, No. 4, May 2007 , pp. 348-356(9)

[17] Altenkirch, J., Steuwer, A., Peel, M., Richards, D.G., Withers, P.J., The effect of tensioning and sectioning on residual stresses in aluminum AA7749 friction stir welds, Materials Science and Engineering A, Vol. 488, 2008, pp.16-24

[18] Steuwer, A., Peel, M.J., Withers, P.J., Dissimilar friction stir welds in AA5083-AA6082: The effect of process parameters on residual stress, Materials Science and Engineering A, Vol. 441, 2006, pp. 187-196

[19] Sutton, M.A., Reynolds, A.P., Wang, D.-Q., Hubbard, C.R., A Study of Residual Stresses and Microstructure in 2024-T3 Aluminum Friction Stir Butt Welds, Journal of Engineering Materials and Technology, Vol. 124, 2002, pp. 215-221

[20] Peel, M., Steuwer, A., Preuss, M., Withers, P.J., Microstructure, mechanical properties and residual stresses as a function of welding speed in aluminum AA5083 friction stir welds, Acta Materialia, Vol. 51, 2003, pp. 4791-4801
[21] Khandkar, M.Z.H., Khan, J.A., Reynolds, A.P., Sutton, M.A., Predicting residual thermal stresses in friction stir welded metals, Journal of Materials Processing Technology, Vol. 174, 2006, pp. 195-203
[22] Tutum, C.C., Schmidt H.B. Hattel, J, Assesment of Benchmark Cases for modeling of Residual stresses and Distortions in Friction Stir Welding, 7th International Symposium Friction Stir Welding, TWI, May 2008
[23] Murugananth, M., "Metaheuristic Multiobjective Optimization in Steel Welds", Materials and Manufacturing Processes, Vol. 24, No. 2, 2009, pp. 230-239
[24] Mishra, S., DebRoy, T., Tailoring gas tungsten arc weld geometry using a GA and a NN trained with convective heat flow calculations, Materials Science and Engineering A, Vol. 454–455, 2007, pp. 477–486
[25] Michaleris, P., Dantzig, J., Tortorelli, D., Minimization of welding residual stress and distortion in large structures, Welding Journal, Vol. 78, 1999, pp. 361-366
[26] Bogomolny, M., Bendsøe, M.P., Some Optimal Design Aspects of Resistance Welding, 7th World Congress on Structural and Multidisciplinary Optimization, Seoul, May 2007, pp. 2687-2695
[27] Nandan, R., Lienert, T.J., DebRoy, T., Toward reliable calculations of heat and plastic flow during friction stir welding of Ti-6Al-4V alloy, Int. J. Mat. Res., Vol. 99, No. 4, 2008, pp. 434-444
[28] Tutum, C.C., Schmidt, H.B., Hattel, J.H., Optimization of the Process Parameters for Controlling Residual Stress and Distortion in Friction Stir Welding, International Conference on Simulation Based Engineering and Sciences, TCN CAE 2008
[29] Fratini, L., Corona, V., Friction Stir Welding Lap Joint Resistance Optimization Through Gradient Techniques, Journal of Manufacturing Science and Engineering, Vol. 129, 2007, pp. 985-990
[30] Lombard, H., Hattingh, D.G., Steuwer, A., James, M.N., Optimising FSW process parameters to minimise defects and maximise fatigue life in 5083-H321 aluminum alloy, Engineering Fracture Mechanics, Vol. 75, 2008, pp. 341-354
[31] Lakshminarayanan, A.K., Balasubramanian, V., Process parameters optimization for friction stir welding of RDE-40 aluminum alloy using Taguchi technique, Trans. Nonferrous Met. Soc. China, Vol. 18, 2008, pp. 548-554
[32] Tutum, C.C., Schmidt, H., Hattel, J., Bendsøe, M., Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization, 7th World Congress on Structural and Multidisciplinary Optimization, Seoul, May 2007, pp. 2639-2646
[33] Larsen, A.A., Bendsøe, M.P., Schmidt, H.N.B., Hattel, J.H., Optimization of Thermal Aspects of Friction Stir Welding – Initial Studies Using a Space Mapping Technique, 7th World Congress on Structural and Multidisciplinary Optimization, Seoul, May 2007, pp. 2663-2669
[34] Zhu, X.K., Chao, Y.J., Effects of temperature-dependent material properties on welding simulation, Computers and Structures, Vol. 80, 2002, pp. 967-976
[35] Schmidt, H.B., Hattel, J.H., A thermal-pseudo-mechanical model for the heat generation in Friction Stir Welding, 7th International Symposium Friction Stir Welding, TWI, May 2008
[36] Lindgren, L.-E., "Computational Welding Mechanics: Thermomechanical and Microstructural Simulations", Woodhead Publishing Ltd, 2007
[37] Labeas, G., Tsirkas, S., Diamantakos, J., Kermanidis, A., "Effect of residual stresses due to laser welding on the Stress Intensity Factors of adjacent crack", 11th International Conference on Fracture, March 2005
[38] Nadimi, S., Khoushehmehr, R.J., Rohani, B., Mostafapour, A., "Investigation and Analysis of Weld Induced Residual Stresses in Two Dissimilar Pipes by Finite Element Modeling", Journal of Applied Sciences, Vol. 8, No. 6, pp. 1014-1020
[39] Deb, K., Multi-Objective Optimization Using Evolutionary Algorithms, John Wiley & Sons, 2001
[40] Deb, K., Multi-objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems, Evolutionary Computation, Vol. 7, No. 3, 1999, pp. 205-230
[41] Deb, K., Unveiling innovative design principles by means of multiple conflicting objectives, Engineering Optimization, Vol. 35, No. 5, 2003, pp. 445-470
[42] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, 2002, pp. 182-197
[43] Zitzler, E., Deb, K., Thiele, L., Comparison of multiobjective evolutionary algorithms: Emprical results, Evolutionary Computation, Vol. 8, No. 2, 2000, pp. 173-195

[44] Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989
[45] Seidel, T.U., Reynolds, A.P., Visualization of the Material Flow in AA2195 Friction-Stir Welds Using a Marker Insert Technique, Metallurgical and Materials Transactions A, Vol. 32A, 2001, pp. 2879-2884
[46] T. Hashimoto, S. Jyogan, K. Nakata, Y.G. Kim, M. Ushio, FSW joints of high strength aluminum alloy, Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, USA, June 14–16, 1999
[47] Ma, Z.Y., Mishra, R.S., Mahoney, M.W., Superplastic deformation behaviour of friction stir processed 7075Al alloy, Acta Materialia, Vol. 50, No. 17, 2002, pp. 4419-4430
[48] W.J. Arbegast, P.J. Hartley, Friction Stir Weld Technology Development at Lockheed Martin Michoud Space System-An Overview, Proceedings of the Fifth International Conference on Trends inWelding Research, Pine Mountain, GA, USA, June 1–5, 1998, p. 541
[49] Gharacheh, M.A., Kokabi, A.H., Daneshi, G.H., Shalchi, B., Sarrafi, R., The influence of the ratio of "rotational speed/traverse speed" (w/v) on mechanical properties of AZ31 friction stir welds, International Journal of Machine Tools & Manufacture, Vol. 46, 2006, pp. 1983-1987
[50] Schmidt, H., Hattel, J., A local model for the thermomechanical conditions in friction stir welding, Modelling Simul. Mater. Sci. Eng., Vol. 13, 2005, pp. 77-93

D PAPER-IV

J. H. Hattel, H. N. B. Schmidt and C. C. Tutum, *"Thermomechanical Modelling of Friction Stir Welding"*, Proc. 8th International Conference on Trends in Welding Research, Pine Mountain, Atlanta, USA, 2008 (Invited keynote lecture).

Trends in Welding Research, Proceedings of the 8th International Conference
Stan A. David, Tarasankar DebRoy, John N. DuPont, Toshihiko Koseki, Herschel B. Smartt, editors, p 1-10
DOI: 10.1361/cp2008twr001

Thermomechanical Modelling of Friction Stir Welding

J.H. Hattel, H.N.B. Schmidt and C. Tutum
Department of Mechanical Engineering
Technical University of Denmark
DK-2800 Lyngby, Denmark

Abstract

Friction Stir Welding (FSW) is a fully coupled thermomechanical process and should in general be modelled as such.
Basically, there are two major application areas of thermomechanical models in the investigation of the FSW process: i) Analysis of the thermomechanical conditions such as e.g. heat generation and local material deformation (often referred to as flow) during the welding process itself. ii) Prediction of the residual stresses that will be present in the joint structure post to welding. While the former in general will call for a fully-coupled thermomechanical procedure, however, typically on a local scale, the latter will very often be based on a semi-coupled, global procedure where the transient temperatures drive the stresses but not vice-versa.
However, in the latter, prior knowledge about the heat generation must be obtained somehow, and if experimental data are not available for the FSW process at hand, the heat generation must either be prescribed analytically or based on a fully coupled analysis of the welding process itself. Along this line, a recently proposed thermal-pseudo-mechanical model is presented in which the temperature dependent yield stress of the weld material controls the heat generation. Thereby the heat generation is still numerically predicted but the cumbersome fully coupled analysis avoided.
In the present work the formulation of all three mentioned modelling approaches as well as the very fundamental pure thermal models are briefly presented and discussed together with selected modelling results including prediction of material flow during welding, prediction of heat generation with the thermal-pseudo mechanical model as well as residual stress and deformation analysis combined with in-service loads.

Introduction

During the last decade an increasing amount of contributions have been given in literature in the field of thermomechanical modelling of FSW. These models are generally chategorized by either their area of application, i.e. flow models or residual stress models, or by the continuum mechanics approach they are based upon, i.e. Computational Solid Mechanics (CSM) models or Computational Fluid Dynamics (CFD) models, where the former typically are Lagrangian and the latter Eulerian, see figure 1.

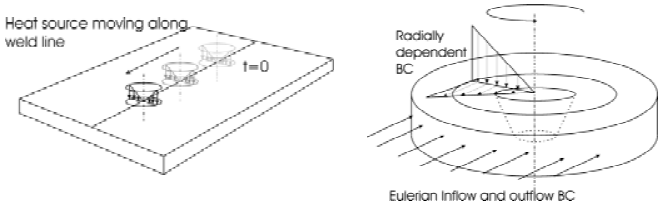


Figure 1: Left: Lagrangian model of FSW. Right: Eulerian model of FSW, [1].

The equations for conservation of momentum, energy and mass, respectively are in the Lagrangian and Eulerian frames given by

Table 1: Governing equations for conservation of momentum, energy and mass, respectively in Lagrangian and Eulerian frames.

Lagrangian frame Typically CSM	**Eulerian frame** Typically CFD
Energy	
$\rho c_p \dot{T} = \left(kT_{,i}\right)_{,i} + \eta s_{ij}\dot{\varepsilon}_{ij}^{pl}$	$\rho c_p \dot{T} = \left(kT_{,i}\right)_{,i} + \eta s_{ij}\dot{\varepsilon}_{ij}^{pl} - u_i(\rho c_p T)_{,i}$
Momentum	
$\rho \ddot{u}_i = \sigma_{ji,j} + p_i$	$\partial\left(\rho \dot{u}_i\right)/\partial t = \sigma_{ji,j} + p_i - \rho(\dot{u}_j \dot{u}_i)_{,j}$
Mass	
No explicit equation	$\dot{\rho} = -(\rho u_i)_{,i}$

These are the fundamental, governing partial differential equations that depending on the continuum mechanical framework that is used will need to be modified and combined with proper constitutive laws for the particular case at hand. In the following, the basis of thermal models, computational solid mechanics (CSM) models, computational fluid dynamics (CFD) models and the new thermal-pseudo-

mechanical model for FSW are shortly presented and the above shown equations are modified accordingly. It should be mentioned that although microstructural models and their coupling to mechanical properites are very important for predicting the behaviour of FSW welds, they are not discussed in the present work.

Thermal models

Thermal modelling has since the late nineties been a central part of modelling of FSW in general (some major contributions are given in [2-10]). One of the reasons for this is that many of the properties of the final weld are directly a function of the thermal history which the workpiece has been exposed to. Secondly, the FSW process itself is highly affected by the heat generation and heat flow and thirdly, from a modelling viewpoint, thermal modelling of FSW can be considered the basis of all other models of the process, be it microstructural, CFD or CSM models. In the FSW process the welding parameters are all chosen such that the softening of the workpiece material enables the mechanical deformation and material flow. However, unlike many other thermomechanical processes, the mechanisms of FSW are fully coupled meaning that the heat generation is related to material flow and frictional/contact conditions and vice versa. Thus, in theory a thermal model alone cannot predict the temperature distribution/history without a pre-knowledge about the heat generation, since the fundamental mechanisms of FSW are not part of a pure thermal model. For this reason, several analytical expressions have been given in literature for the heat generation for a given weld as a function of tool geometry and welding parameters, e.g. tool radius and rotational speed. This means that in the energy equation in table 1, the plastic dissipation term is replaced by a heat source term, see table 2, line 1.

Table 2: Governing equations for conservation of energy in thermal models, respectively in Lagrangian and Eulerian frames

Lagrangian frame	**Eulerian frame**
Energy with source term	
$\rho c_p \dot{T} = \left(kT_{,i}\right)_{,i} + \dot{Q}'''$	$0 = \left(kT_{,i}\right)_{,i} + \dot{Q}''' - u_i\left(\rho c_p T\right)_{,i}$
Energy without source term but all heat prescribed at surface	
$\rho c_p \dot{T} = \left(kT_{,i}\right)_{,i}$	$0 = \left(kT_{,i}\right)_{,i} - u_i\left(\rho c_p T\right)_{,i}$

In all thermal models of FSW the main task now is to solve this energy equation with an appropriate set of initial and boundary conditions. However, in most pure thermal models of FSW, the heat generation from both frictional and plastic dissipation is modelled via a surface flux boundary condition at the tool/matrix interface, see table2, line 2.
The main unknown parameters in these surface flux expressions are either the friction coefficient under the assumption of sliding and the material yield shear stress under the assumption of sticking.

Along this line, the authors have proposed the following generally adopted equation for the total heat generation, [7], (see figure 1)

$$\begin{aligned} Q_{total} &= \delta Q_{sticking} + (1-\delta) Q_{sliding} \\ &= \frac{2}{3}\pi\omega\left[\delta\tau_{yield} + (1-\delta)\mu p\right]\times \\ &\left[\left(R_{shoulder}^3 - R_{probe}^3\right)(1-\tan\alpha) + R_{probe}^3 + 3R_{probe}^2 H\right] \end{aligned} \quad (1)$$

Where δ is the dimensionless slip rate between the tool and the work piece (1 for full sticking and 0 for full sliding).

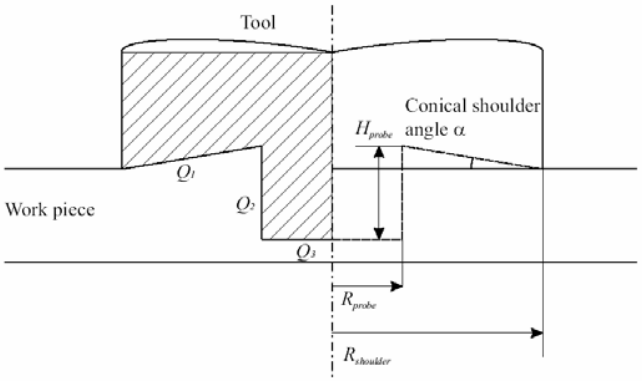


Figure 2: Different contributions from the shoulder and the pin to the surface heat generation, [6].

However, when implementing this into a numerical model using a position dependent surface flux [Wm^{-2}], it is typically used on the following form, [6],

$$\begin{aligned} q_{total} &= \dot{\gamma}\tau_{friction} + (\omega r - \dot{\gamma})\tau_{yield} \\ &= \omega r\left[\delta\tau_{yield} + (1-\delta)\mu p\right] \end{aligned} \quad (2)$$

which in fact is the basis for deriving equation (1). Furthermore, when combining equation (1) and (2) and assuming the simple tool geometry of a flat shoulder only, one obtains the following well-known expression for the heat generation [Wm^{-2}], see e.g. [7],

$$q_{total} = \frac{3Q_{total} r}{2\pi R_{shoulder}^3} \quad (3)$$

which can be applied as a radius dependent surface flux in the model, under the assumption of a constant contact condition close to sliding or in cases of sticking where the shear layer is very thin.

From equation (3) it can be seen that Q_{total} can be considered an input parameter in the same manner as the friction coefficient, the pressure and the material's yield shear stress in equation (1). However, it should be mentioned that having Q_{total} as part of the input parameters for the model in some

situations could conflict with the very objective for thermal modelling of FSW. This is especially the case when the model should predict temperatures for conditions not supported by measurements of the heat input Q_{total}. In these cases it is not straightforward to predict or simulate the effect of e.g. a change in welding or rotational speed, because the total heat generation Q_{total} is a function of these changes in parameters and hence can be considered an internal function of the FSW process; unlike other welding process, e.g. TIG where the heat input is controlled externally.
This dilemma can of course be overcome by supporting the thermal model with a thermo-mechanical CSM or CFD based model which "includes" the underlying physics of the process, namely the material flow producing heat generation by plastic dissipation in the shear layer and frictional contact at the tool/workpiece. As being pretty obvious from the later sections on CSM and CFD models, the efforts for such simulations are highly demanding from both a computational and a human resource point of view, thus leaving pure thermal modelling of FSW as a valuable "simple" alternative for simulation of heat flow having in mind the limitation of such models, which as earlier mentioned primarily lie in the evaluation of Q_{total}.
Having said that, it should be noted that the most utilized procedure for evaluating Q_{total} typically is to rely on experimental findings by simply performing the actual welds and measuring Q_{total} with a dynamometer, and thereby accepting the inherent limitation of the resulting thermal model to predicting only temperatures for a *known* total heat generation.

Since the prescribing of the heat generation is the most single important parameter in thermal models of FSW the authors proposed a classification of different heat sources [7].
One characteristic in this classification is how "detailed" the tool heat generation is resolved. Three levels were evaluated, i.e. i) including shoulder heat generation, without probe heat generation, ii) including shoulder and probe heat generation, with probe material and iii) shoulder and probe heat generation but without probe material.
A second characteristic was whether the convective contribution due to the material flow in the shear layer was taken into account. Two extreme contact conditions were evaluated, i.e. full sliding and full sticking. In the case of sliding the heat was applied as a surface flux and in the case of sticking the heat was applied as a volume flux in a shear layer. This shear layer was prescribed analytically assuming a uniform thickness with a linear velocity profile ramped between the tool velocity at the contact interface and the welding velocity outside the shear layer.
It was concluded that for analysing the temperature field in the volume under the tool, special attention should be paid to how the heat generation and the material "flows" around the tool probe. The main effect of including the probe in the thermal model is to change the material properties to those of the tool; secondly, rotating the probe and modelling the shear layer around it such that it resembles a flow model.

One new procedure to obtain the heat generation is to couple the traditional analytical expression for the heat generation with a constraint based on experimental or phenomenological considerations. It is well-known from the constitutive behaviour of a solid (representative for those alloys used in FSW), that the yield stress dramatically decreases once the temperature approaches the solidus temperature. Above the solidus temperature, the material acts as a fluid. As a consequence, the material close to the tool/matrix interface will reduce its heat generation to negligible values if exceeding the solidus temperature – reducing the temperature level – allowing the material to recover its strength. A self-stabilizing effect will thus establish at a temperature level around the solidus temperature, hence this could be used as an average temperature constraint in a pure thermal model. This is what is done in the thermal model by Tutum *et al*. [10] where the heat generation in FSW of AA2024 is controlled by using an optimization scheme such that an *average* temperature of 500 ^{o}C at the tool/workpiece interface is obtained.
This procedure has given promising results for obtaining temperature fields without prior knowledge regarding Q_{total} – in fact this is an output of the optimization analysis.

Computational Solid Mechanics Models

As mentioned before, the two reasons for doing solid mechanics models of FSW are to predict the flow and to predict the residual stresses. The main governing equations for these types of models are given in table 3. The energy equation will essentially be as shown in table 1, line 1 (giving a fully coupled thermomechanical model) or table 2 (giving a sequentially coupled thermomechanical model) depending on the choice of thermal model. For the solid mechanically based *flow modelling* there are basically two approaches: The simplest is to assume a rigid-visco-plastic material, i.e. the total strain is equal to the visco-plastic strain and use an implicit solver based on quasi-static equilibrium. This is done by several authors in literature, see e.g. [12-14].
The more comprehensive approach is to use the dynamic equilibrium equation and to take all the contributions to the strain into account although the viscoplastic part obviously is the dominating contribution. This is typically done in an ALE formulation where the dynamic equilibrium equation is solved in an explicit manner. This approach has been used by several authors, see e.g. [15-18]. A special feature about the model by Schmidt and Hattel [17] (implemented in ABAQUS Explicit) is that the heat generation between the tool and the matrix is not prescribed but part of the solution itself. This adds to the generality of the model but also to the complexity and hence the need for computational power.
The model [17] has been used to predict among others the plastic strain in the weld which resembles the well-known "flow arm" quite well, see figures 3 and 4 as well as the void

formation behind the tool, see figure 6, being the first model in literature able to predict this.

Table 3: Basic equations for solid mechanics models of FSW for flow and residual stresses, respectively

Flow during welding	**Residual stresses**
Momentum	
Dynamic equilibrium, e.g. explicit ALE $\rho\ddot{u}_i = \sigma_{ji,j} + p_i$	Dynamic equilibrium not relevant
Static equilibrium, e.g. implicit "flow formulation" $\sigma_{ji,j} + p_i = 0$	Static equilibrium. Quasi-static thermomechanical analysis $\sigma_{ji,j} + p_i = 0$
Strain vs. displacements	
Large strain theory $\varepsilon_{ij}^{tot} = \frac{1}{2}(u_{i,j} + u_{j,i} + u_{k,i}u_{k,j})$	Small strain theory $\varepsilon_{ij}^{tot} = \frac{1}{2}(u_{i,j} + u_{j,i})$
Total strain	
Rigid viscoplastic: $\varepsilon_{ij}^{tot} = \varepsilon_{ij}^{vp}$ Explicit ALE $\varepsilon_{ij}^{tot} = \varepsilon_{ij}^{el} + \varepsilon_{ij}^{pl} + \varepsilon_{ij}^{vp} + \varepsilon_{ij}^{th}$	Rate independent plasticity $\varepsilon_{ij}^{tot} = \varepsilon_{ij}^{el} + \varepsilon_{ij}^{pl} + \varepsilon_{ij}^{th}$
Thermal strain	
$\varepsilon_{ij}^{th} = \delta_{ij}\int_{T_1}^{T_2}\alpha(T)dT$	
Constitutive law	
Equivalent stress $\bar{\sigma} = \left(\frac{3}{2}\sigma_{ij}\sigma_{ij}\right)^{1/2}$ Equivalent total strain rate (rigid viscoplastic) $\dot{\bar{\varepsilon}}^{tot} = \left(\frac{2}{3}\dot{\varepsilon}_{ij}^{tot}\dot{\varepsilon}_{ij}^{tot}\right)^{1/2}$ Yield stress, in general: $\sigma_y = \sigma_y(T, \dot{\bar{\varepsilon}}^{pl})$ $\sigma_y \to 0 \quad T \to T_{cut-off} \sim T_{sol}$ Norton power law $\sigma_y = K\dot{\bar{\varepsilon}}^m$ Inverse hyperbolic sine $\sigma_y = \frac{1}{\alpha}\sinh^{-1}\left(\frac{Z}{A}\right)^{1/n}$ $Z = \dot{\bar{\varepsilon}}e^{Q/RT} \qquad m = \frac{\partial \ln \bar{\sigma}}{\partial \ln \dot{\bar{\varepsilon}}}$ Johnson-Cook $\sigma_y = \left(A + B\bar{\varepsilon}^n\right)\left(1 + C\ln\frac{\dot{\bar{\varepsilon}}}{\dot{\varepsilon}_0}\right)$ $\times\left(1 - \left(\frac{T - T_{ref}}{T_{sol} - T_{ref}}\right)^m\right)$	Equivalent stress $\bar{\sigma} = \left(\frac{3}{2}\sigma_{ij}\sigma_{ij}\right)^{1/2}$ Equivalent plastic strain $\bar{\varepsilon}^{pl} = \left(\frac{2}{3}\varepsilon_{ij}^{pl}\varepsilon_{ij}^{pl}\right)^{1/2}$ Yield stress, in general: $\sigma_y = \sigma_y(T, \bar{\varepsilon}^{pl})$ $\sigma_y \to 0 \quad T \to T_{cut-off} \sim T_{sol}$ Hardening law, e.g. Ramberg-Osgood: $\bar{\varepsilon}^{tot} = \frac{\sigma}{E} + \alpha\frac{\sigma_y}{E}\left(\frac{\sigma}{\sigma_y}\right)^n$ Elasticity Hooke's generalized law $\sigma_{ij} = \frac{E}{1+\nu}\left(\varepsilon_{ij}^{el} + \frac{\nu}{1-\nu}\delta_{ij}\varepsilon_{ij}^{el}\right)$

Apart from modelling the flow, the purpose of the computational solid mechanics models of FSW obviously is to find residual stresses, see e.g. [19-22]. As seen from table 3 it is clear that these ressidual stress models typically are based on a thermo-elasto-plastic description, whereas the flow models are based on rigid-visco-plastic models in most cases.

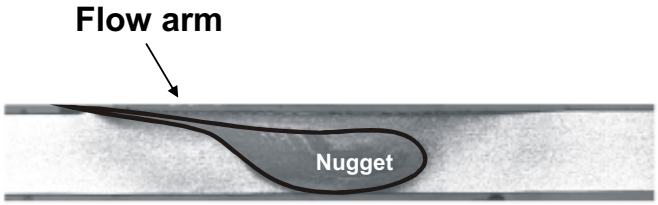


Figure 3: Micrsotructure showing the well-known "flow arm" from the nugget, [17].

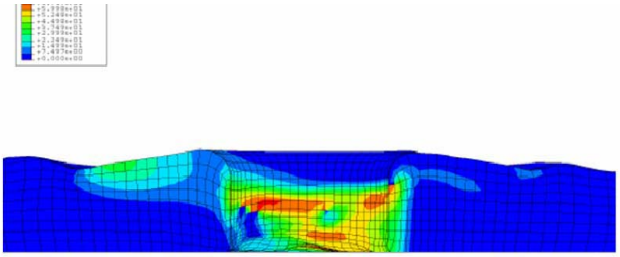


Figure 4: Numerically predicted equivalent plastic strain (in the range of 0-75) showing a similar result in figure 4, [17].

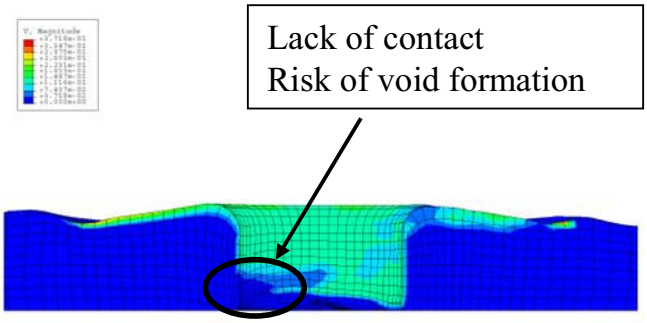


Figure 5: Numerically predicted size of velocity. Note the lack of contact resulting in risk of void formation at the advancing side, [17].

An analysis of the residual stresses and the coupling to the subsequent load situation was carried out by the authors in ABAQUS Standard on a structure where a stringer is friction stir welded onto a plate [22], see figure 6.
The integrated residual stress and load analysis consisted of four steps which are shown in figure 7.
The residual stresses along the dotted line shown in figure 7 after welding and releasing as well as after applying the load are presented in figure 8.
As seen on the figure, applying the load actually slightly reduces the longitudinal stresses in the middlle of the weld. This is of course not a general statement but only valid for this

geometry and load case. Moreover, it should be emphasized, that this type of analysis obviously would benefit a lot from a prediction of the transient evolution of the mechanical properties in the structure close to the weld.

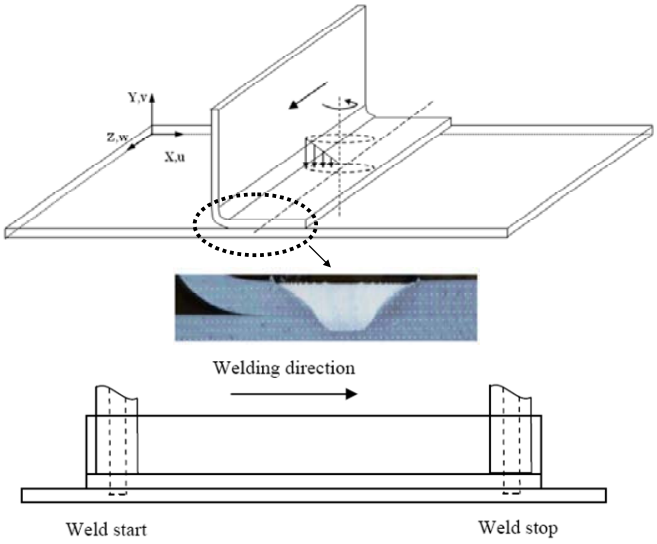


Figure 6: Geometry of a stringer friction stir welded to a plate. Used for coupling of residual stress and in-service load analyses, [22].

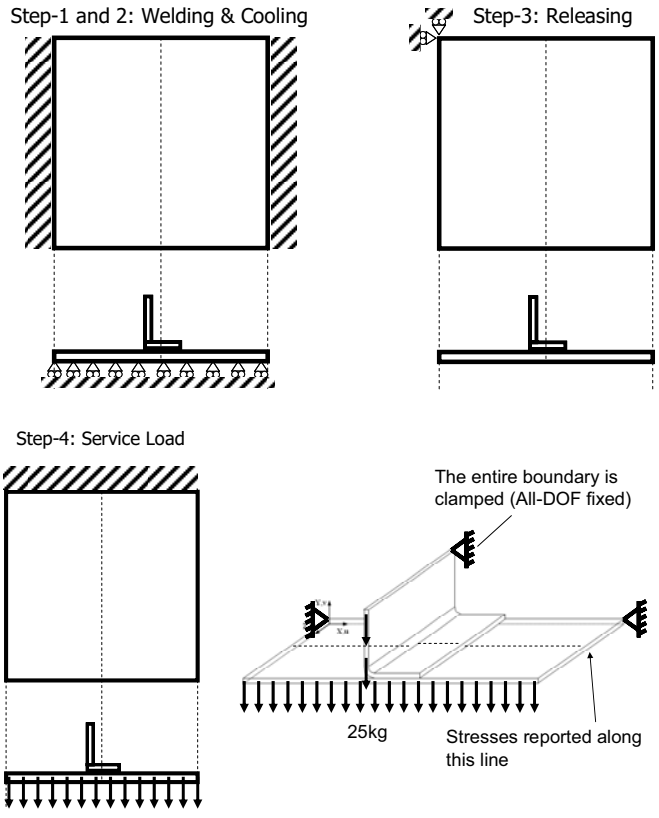


Figure 7: The 4 steps in the coupled residual stress and in-service load analyses, [22]

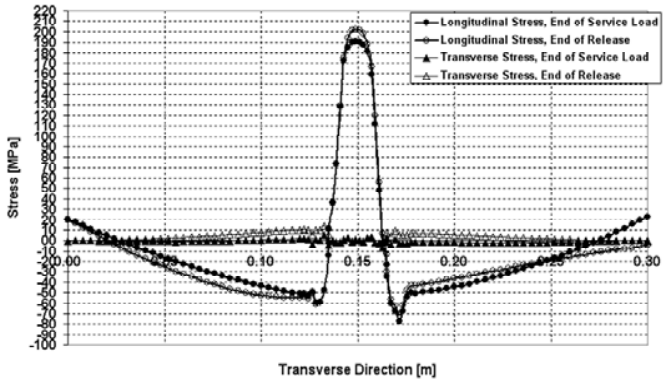


Figure 8: Longitudinal and transverse residual stresses after welding and releasing as well as after applying the load, [22].

Computational Fluid Dynamics Models

Traditionally speaking, CFD models will not be considered as being of "thermomechanical type" in computational welding mechanics. In conventional welding processes, CFD models are normally used for weld pool dynamics and CSM models for residual stresses. However, in FSW both CSM models as the ABAQUS Explicit model mentioned above [17] and CFD models as the one that will be shown in the following can be used to find the material flow during welding. For this reason it makes good sense to present the CFD approach to material flow in FSW, anyway.

The basic assumption in the CFD models obviously is to treat the matrix material as a fluid, see table 4 for an overview of the governing equations. In particular, this calls for a suitable constitutive model from which the viscosity can be expressed. In literature, it is the most common for FSW to use the inverse hyperbolic sine law, see tables 3 and 4, for the yield stress as a function of the shear rate. This expression captures both the power law regime for low strain rates and the power law breakdown at higher strain rates. CFD models are typically formulated in a Eulerian frame, so that the tool is stationary, hence only rotating and the welding speed is accounted for by having incoming and outgoing material flow at the boundaries. The energy equation will be that of table 1, line 1 (right) for a fully-coupled model or table 2, lines 1 and 2 (right) for a sequentially coupled model.

Some important contributions to flow modelling in FSW based on CFD can be found in [23-27].

In Figures 9 and 10, a comparison between experiment, analytically based streamlines and a CFD model (implemented in COMSOL) is shown, [28]. The experiment contained of welding through a line of Cu marker material (MM) in two plates of 2024 and then unscrewing the tool as its center was aligned with the line of MM.

Based on experimental findings in terms of CT pictures and comparisons with theoretical predictions by both analytical and numerical flow models, the authors, [28], have proposed

three characteristic zones for the flow around the tool probe: i) the rotation zone. ii) the transition zone and iii) the deflection zone, see Figure 11. It should be mentioned that this classification is based on a simplified representation of the real, more complex flow – thus not taking into account 3D flow effects, non-symmetrical tool features (e.g. threads) and cyclic contact condition (collapse of the shear layer).

Table 4: Basic equations for fluid mechanics models of FSW for flow

Momentum	
Steady state $0 = \sigma_{ji,j} + p_i - \rho(\dot{u}_j \dot{u}_i)_{,j}$	Steady state, Stokes flow $0 = \sigma_{ji,j} + p_i$
Strain rates vs. displacement rates	
$\dot{\varepsilon}_{ij}^{Tot} = \frac{1}{2}(\dot{u}_{i,j} + \dot{u}_{j,i})$	
Constitutive law	
Incompressible Newtonian $\sigma_{ij} = -\delta_{ij} P + 2\mu\dot{\varepsilon}_{ij}$ Eq. stress vs. eq. strain rate $\bar{\sigma} = K\dot{\bar{\varepsilon}} \quad (m = 1)$ Cut-off temperature $\mu \to 0 \quad T \to T_{cut-off} \sim T_{sol}$	Incompressible non-Newtonian $\sigma_{ij} = -\delta_{ij} P + \mu_{eff} 2\dot{\varepsilon}_{ij}$ Eq. stress vs. eq. strain rate $\bar{\sigma} = K\dot{\bar{\varepsilon}}^m$ Effective viscosity, in general $\mu_{eff} = \frac{\bar{\sigma}}{3\dot{\bar{\varepsilon}}} \Rightarrow \mu_{eff} = \mu\sqrt{3}^{m-1}\dot{\bar{\varepsilon}}^{m-1}$ Effective viscosity, power law $\mu_{eff} = \frac{1}{3}K\dot{\bar{\varepsilon}}^{m-1}$ Effective viscosity, inv.hyp.sine $\mu_{eff} = \frac{\frac{1}{\alpha}\sinh^{-1}\left\{\frac{1}{A}\dot{\bar{\varepsilon}}e^{-Q/RT}\right\}^{1/n}}{3\dot{\bar{\varepsilon}}}$ Cut-off temperature $\mu_{eff} \to 0 \quad T \to T_{cut-off} \sim T_{sol}$

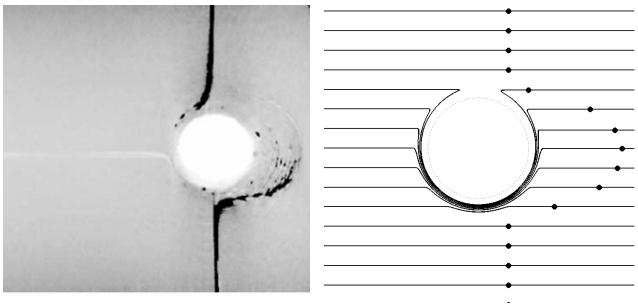


Figure 9: Left: CT picture of weld with transverse MM. Exit hole aligned with the MM plane. Right: Analytically obtained stream lines/tracer particles, [28].

The CFD models are very suitable for analyzing the material flow during FSW, however they have some important limitations that should be mentioned: i) The residual stresses after welding which are of great importance cannot be captured, since a CFD model in nature is rate dependent only. ii) The elastic stress state in the area "far away" from the tool is not properly described. iii) The contact condition at the tool/interface is typically not very well described in the present CFD models of FSW, i.e. an a priori prescribed velocity emulating either sticking or sticking/sliding is normally used. These three limitations mentioned for the CFD models are all more or less eliminated if the material flow is described by computational solid mechanics models. These, on the other hand, often have the disadvantage of high demands for remeshing, which the CFD models, being of Eulerian nature, do not.

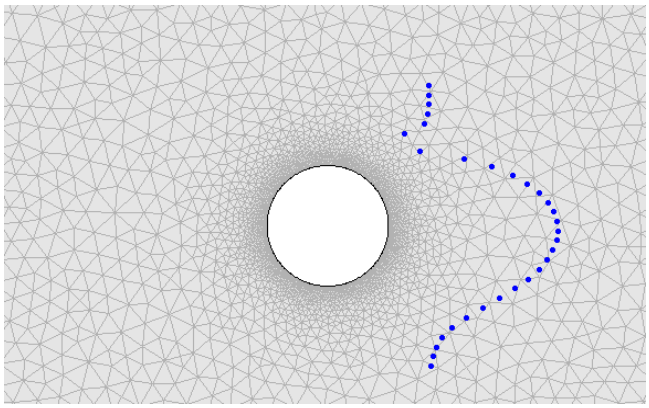


Figure 10: CFD model in COMSOL with simulated tracer particles. Non-Newtonian model with m=0.17, [28].

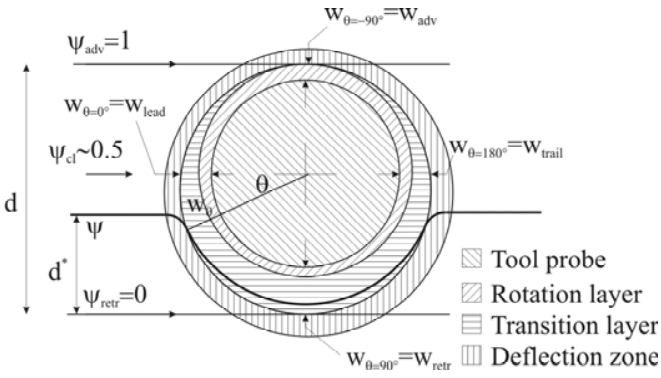


Figure 11: Characteristic zones for the flow around the tool probe: i) the rotation zone. ii) the transition zone and iii) the deflection zone, [28].

Thermal-pseudo-mechanical model

In this paragraph, the author's recently proposed thermal-pseudo-mechanical model [1] is presented. As the name indicates, the model, although being purely thermal from the viewpoint of the classification presented in the present work, still takes some mechanical effects into account without actually solving for any mechanical fields. The main advantage of this is that the obvious simplicity of thermal models is still maintained in the thermal-pseudo-mechanical

model. The basis of the model is to consider the heat generation at a material segment which in general depends on the material flow stress for the given temperature, given strain and given strain rate. The contact stress must be in equilibrium with the material yield shear stress of the underlying material during steady state conditions. This also holds for contact conditions close to sliding, because just for the slightest degree of sticking, i.e. $\delta \sim 0 \ll 1$, see [7] for definition, there will be plastic deformation at the very interface, hence equilibrium between the contact/frictional stress and the material response must prevail. For steady state conditions, this leads to the following relationship

$$\tau_{friction} = \tau_{contact} = \tau_{yield}(T) \tag{4}$$

where $T = T(x, y, z)_{\text{interface}}$ is the non-uniform temperature at the contact interface.

Realizing this, leads to using the material's yield stress to describe the heat generation coming from friction rather than relying on the more classical Coulomb approach which uses the non-uniform pressure distribution and the friction coefficient which in general is a function of the temperature, slip rate, pressure and interface properties (roughness, etc.). Thus, the alternative approach proposed here quantifies the heat generation via the material's yield stress which in general is a function of temperature, strain and strain rates.

However, since it is not the aim to involve a full thermomechanically based solution, only the temperature dependence of the material flow stress is taken into account. This leaves a very important point to be clarified: For which values of ε and $\dot{\varepsilon}$ should the yield stress be chosen? In other words: How should the mechanical history that a material segment experiences during its path through the shear layer be represented? At the present stage, the maximum values of the yield stress resembling a torsion test for a constant strain rate have been chosen.

Summing up, the benefits as compared to classical Coulomb friction are several: a) more data is available for the material's yield response as compared to the friction coefficient μ b) prior knowledge to the pressure distribution is not needed c) the model becomes much more robust towards variations in the input data.

In the thermal pseudo mechanical model, the heat generation contributions from friction and plastic dissipation are now both described as a surface flux and given by

$$\begin{aligned} q_{friction} &= (1-\delta)\omega r \tau_{friction} \\ q_{plastic} &= \delta \omega r \tau_{yield} \end{aligned} \tag{5}$$

Combining these expressions with equation (4) and assuming pure shear yields

$$q_{total} = \omega r \tau_{yield}(T) = \omega r \frac{\sigma_{yield}(T)}{\sqrt{3}} \tag{5}$$

which underlines the fact that the entire heat generation is modelled as a surface flux which is depending on the local temperature. This temperature dependence hence leads to a heat source which is part of the numerical solution itself, making it a non-linear problem to solve. This obviously calls for an iterative procedure. However, it turns out that the price in calculation time is relatively low, i.e. a typical thermal pseudo mechanical model needs approximately twice the calculation time as a normal steady state thermal model of FSW.

It should be underlined that the above described thermal-pseudo-mechanical model will capture the first order effect of the material's mechanical response without making a full thermomechanical model. A very important characteristic about this model is the self-stabilizing effect once temperatures reach the solidus temperature, in which case the heat source "turns itself off". This makes the model effectively more dependent on changes in boundary conditions and welding parameters, rather than e.g. choice of material law. Moreover it should be further underlined that the heat transfer coefficients describing boundary conditions are the only fitting parameters in the model. It is important to emphasize that once these have been found, the effect of welding parameters such as welding velocity and rotational speed can be varied and the model will implicitly capture this in the heat generation and temperature profiles.

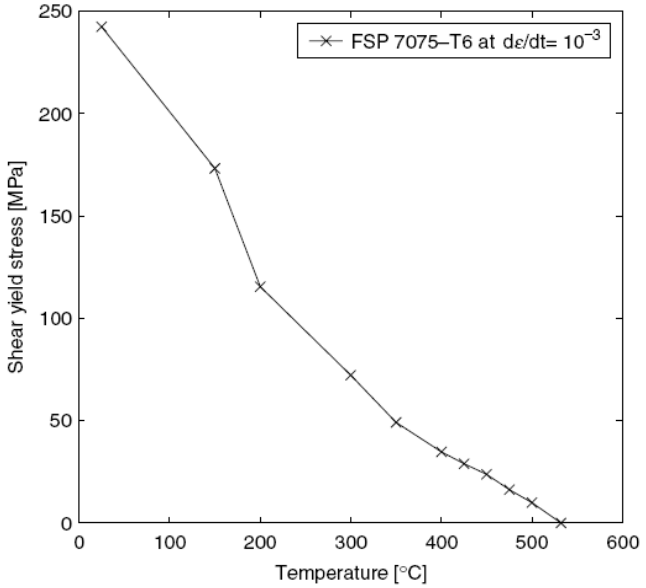


Figure 12: Shear yield stress derived from [29].

The temperature dependent heat source given by equation (5) is implemented in a Eulerian model that, in addition to the workpiece, includes the tool and the backing plate. The thermal model is implemented in Comsol 3.3 and is based on

an experiemental weld of 7075 T6 instrumented with thermocouples. The welding conditions are given in [1] and the thermal properties are taken from [29]. The experimentally found values for the maximal yield stress of friction stir processed 7075 at strain rates of 0.001 1/s are used for the shear yield stress in figure 12. A total of 100.000 degrees of freedom in the model are solved for. A more thorough description of the model can be found in [1].

Figure 13 shows the model results for the local heat generation given by equation (5) evaluated along the intersection between the tool and the joint line. The temperature dependent heat generation denoted by (-) is the result of the thermal-pseudo-mehcanical model and this is compared with the analytically prescribed, linearly dependent heat generation denoted by (--); the latter was evaluated using the same total heat generation as being a result of the former. In the interval under the probe tip, the heat genearation is close to linear since the temperature is nearly constant in this area.

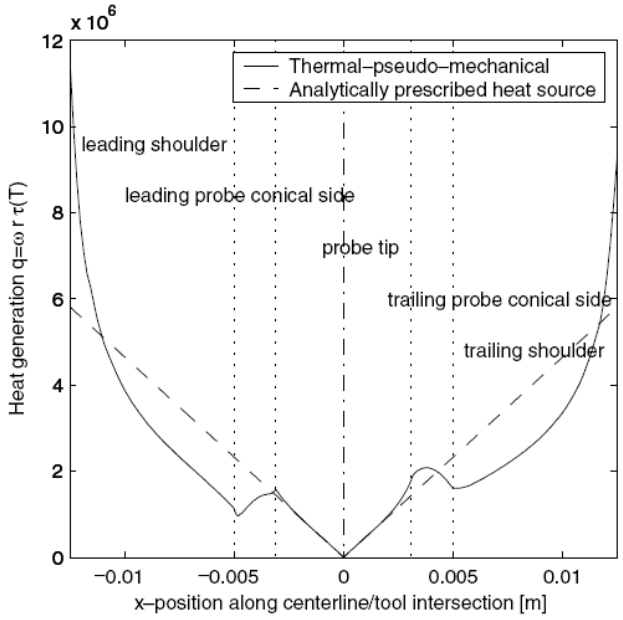


Figure 13: Local heat generation as calculated from eq. (5) evaluated along the intersection between tool and centreline. The total heat generation in the analytical model is prescribed with the value obtained in the thermal-pseudo-mechanical model, [1].

The exponetially increasing heat generation at both the leading and trailing shoulder regions is a combination of the ωr-term and the increase in yield shear stress due to the decrease in temperature for larger radii (here shown as x-values).

The total heat generated at the tool/matrix interface is found to be 1.9 kW (83% from the shoulder, 16% from the conical probe sides and 1% from the probe tip). Moreover, the global thermal efficiency is found to be 88%. Figure 14 shows the corresponding temperatures along the same intersection. It is very obvious that the overestimation of the temperature at the most outer region of the shoulder in the analytical model is avoided by the self-establishing turning off effect of the thermal-pseudo-mehcanical model.

Figure 15 shows a comparison between the experimentally found temperature profiles and the modelling results. The good correleation between the profiels is obtained by adjusting only the heat transfer coefficients/contact resistances. Most important in this context is the coupling between the workpiece and the backing plate.

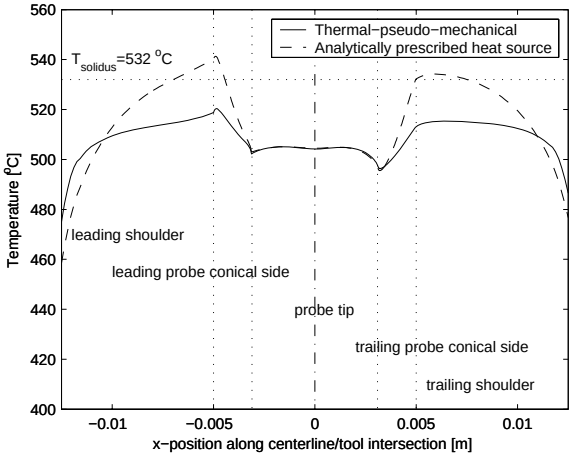


Figure 14: Predicted temperatures along the intersection between tool and centreline, [1].

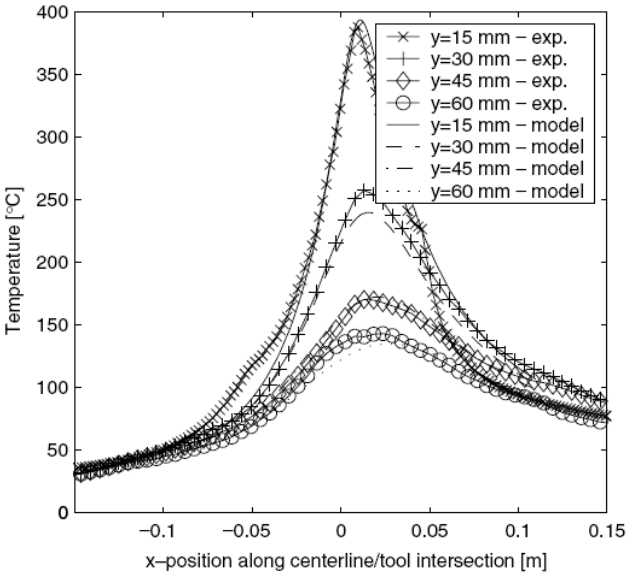


Figure 15: Comparison between the thermal-pseudo-machanical model and experimental results for the far-field temperature profiles at y-values of 15, 30, 45 and 60 mm, [1].

In figure 16 iso-curves for the resulting heat generation are depicted as a function of the welding speed and the rotational speed. Note that for a fixed rotational speed [rpm], increasing the welding speed [mm/s] gives an increase in the heat generation [W]. Similarly for a fixed welding speed, an increase in rotational speed gives an increase in the heat generation. This is also what one would expect.

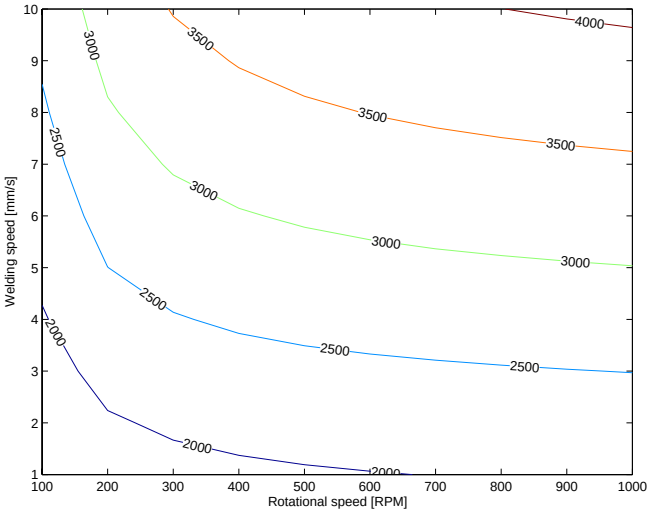


Figure 16: Iso-curves for the resulting heat generation [W] as a function of the welding speed [mm/s] and the rotational speed [rpm], [27].

Figure 17 depicts iso-curves for the maxiamal temperature [ºC] as a function of the rotational speed and welding speed. Here it is noticed that for a fixed rotational speed, increasing the welding speed reduces the maximal temperature and for a fixed welding speed, increasing the rotational speed gives increased maximal temperatures. Note, that the distance between the iso-curves for the maximal temperature becomes increasingly larger when increasing the rotational speed, since the temperature approaches the cut off temperature.

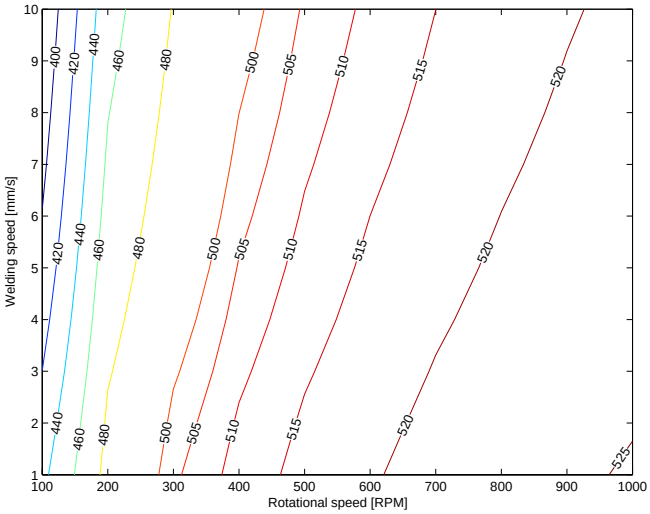


Figure 17: Iso-curves for the resulting maximal temperature [ºC] as a function of the welding speed [mm/s] and the rotational speed [rpm], [27].

The thermal-pseudo-mechanical model presented here is the first attempt to develop a thermal model in which the total heat generation is not an input parameter, but actually a result of the model itself. It should be strongly underlined that the good agreement between the analytically prescribed heat source model and the thermal-pseudo-mechanical model in figure 13 is only achievable when the total heat generation has been found from the latter before being inserted in the former.

Concluding remarks

In the present work, an overview of thermomechanical modelling of FSW has been given. Emphasis has been put on presenting some important basics of the different types of models including thermal models, solid mechanics models and fluid dynamics models. Microstructural models have been excluded from the presentation due to the limited space, however, it should be emphasized that these models are very important for realistic modelling of FSW because of their ability to predict the evolution of the mechanical properties to be used in the thermomechanical models and in particular in models for the in-service behaviour.
Finally, it is fair to say that modelling of FSW now has been established as a field of its own and that we can expect a similar development in this field as in more matured fields of computational welding mechanics (CWM), i.e. a) More combined microstructural-thermomechanical models. b) Integrated process and in-service models. c) More applications on real 3-D structures.

References

1. H. Schmidt and JH. Hattel, *Thermal modelling of friction stir welding, Scripta Materialia.,* 58, 332-337 (2008)
2. M. J. Russel and H. R. Shercliff, *Analytical Modelling of Microstructure Development in Friction Stir Welding*, 1st Int. Symp. On Friction Stir Welding, CA, USA (1999)
3. P. Colegrove, M. Painter, D. Graham and T. Miller, *3-D flow and thermal modelling of the friction stir welding process*, 2nd Int. Symp. On Friction Stir Welding, Sweden (2000)
4. M.Z.H. Khandkar and J.A. Khan, *Thermal model of overlap friction stir welding for Al-alloys, J Materials Processing Manufacture* 10, 91-105 (2001)
5. M. Song and R. Kovacevic, *Thermal modelling of friction stir welding in moving coordinate system and its validation, International Journal of Machine Tool and Manufacturing* 43(6) 605-615 (2003)
6. H. Schmidt, JH. Hattel and J. Wert, *An analytical model for the heat generation in frictions stir welding, Modelling and Simulation in Materials Science and Engineering.,* 12, 143-157 (2004)
7. H. Schmidt and JH. Hattel, *Modelling the heat flow around tool probe in friction stir welding, Science and Technology of Welding and Joining,* 10(2), 176-186 (2005)
8. S. Mandal and K. Williamson, *A thermomechanical hot channel approach for friction stir welding, J Materials Processing Technology*, 190-194 (2006)
9. A. Simar, J. Lecomte-Beckers, T. Pardoen and B. de Meester, *Effect of boundary conditions and heat source*

distribution on temperature distribution in friction stir welding, Science and Technology of Welding and Joining, 170-177 (2006)
10. CC Tutum, H.Schmidt, JH Hattel and MP Bendsøe, *Estimation of the Welding Speed and Heat Input in Friction Stir Welding using Thermal Models and Optimization, Proc. 7th World Congress on Structural and Multidisciplinary Optimization*, Seoul, 2639-2646 (2007)
11. J.K. Raghulapadu, J. Peddieson, G.R. Buchanan and A.C. Nunes, *A rotating plug model of friction stir welding heat transfer, Heat Transfer Engineering*, 29 (3), 321-327, (2008)
12. R.L. Goetz and K.V. Jata, *Modelling friction stir welding of titanium and aluminium alloys*, Proc. Symposium on Friction Stir Welding and Processing, TMS (2001)
13. G. Buffa, J. Hua, R. Shivpuri and L. Fratini, *A continuum based fem model for friction stir welding – model development, Materials Science and Engineering A*, 419, 389-396, (2006)
14. R. K. Uyyury and S.V. Kailas, *Numerical analysis of friction stir welding process, Journal of Materials Engineering and Performance* (2006)
15. A. P. Reynolds, X. Deng, T. Seidel and S. Xu, *Finite element simulation of flow in friction stir welding*, Proc. Joining of Advanced and Specialty Materials, 172-177, MO, USA, (2000)
16. S. Xu and X. Deng, *Two and three-dimensional finite element models for the friction stir welding process*, 4th Int. Symp. On Friction Stir Welding, UT, USA (2003)
17. H. Schmidt and J. Hattel (2005), *A local model for the thermomechanical conditions in friction stir welding, Modelling and Simulation in Materials Science and Engineering*, 13, 77-93, (2005)
18. H. Zhang and Z. Zhang, *Numerical modelling of friction stir welding process by using rate-dependent constitutive model, Journal of Materials Science and Technology*. 23(1), 73-80, (2007)
19. Y.J. Chao and X. Qi, *Heat transfer and thermomechanical analysis of friction stir joining of AA6061-T6 plates*, 1st Int. Symp. On Friction Stir Welding, CA, USA (1999)
20. C.M. Chen and R. Kovacevic, *Finite element modelling of friction stir welding – thermal and thermomechanical analysis, Machine Tools and manufacture*, 43, 1319-1326 (2003)
21. A. Bastier, M.H. Maitournam, K. Dang Van and F. Roger, *Steady state thermomechanical modelling of friction stir welding, Science and Technology of Welding and Joining*, 11(3) 278-288 (2006)
22. C.C. Tutum, H.N.B. Schmidt, J.H. Hattel, *Assessment of benchmark cases for modelling of residual stresses and distortions in friction stir welding*, Proc 7nd Int. Symp. On Friction Stir Welding, Japan (2008)
23. P. A. Colegrove and H.R. Shercliff, *CFD modelling of friction stir welding of thick plate 7449 aluminium alloy, Science and Technology of Welding and Joining*, 11(4), 429-441, (2006)
24. T. Long and A.P. Reynolds, *Parametric studies of friction stir welding by commercial fluid dynamics simulation, Science and Technology of Welding and Joining*, 11(2), 200-208, (2006)
25. R. Nandan, G.G. Roy, T.J. Lienert and T. Debroy, *Three-diemnsional heat and material flow during friction stir welding of mild steel, Acta Materialia*, 55, 883-895, (2007)
26. A.P. Reynolds, *Flow visualization and simulation in FSW, Scripta Materialia*, 58, 338-342, (2008)
27. H. Schmidt and JH Hattel, *Thermal modelling of friction stir welding using a temperature dependent heat source*, Proc 7nd Int. Symp. On Friction Stir Welding, Japan (2008)
28. H.N.B. Schmidt, T.L. Dickerson and J.H. Hattel, *Material flow in butt friction stir welds in AA2024-T3, Acta Materialia*, 54, 1199-1209, (2006)
29. *Metals Handbook*, vol. 2: Properties and Selection: Nonferrous Alloys and Special-Purpose Materials, 10th ed., ASM International, (1990)

E TECHNICAL REPORT-I

C. C. Tutum and J. H. Hattel, *"A benchmark study for unconstrained multi-objective problems"*, September 2009.

Technical Report-1

A Benchmark Study for Unconstrained Multi-objective Optimization Problems

Cem Celal Tutum
Jesper Henri Hattel

Process Modelling Group
Department of Mechanical Engineering
Technical University of Denmark
Produktionstorvet, Kgs. Lyngby
2800, Denmark

September 1st, 2009

Table of Contents

1 Introduction

One of the goals of an ideal multi-objective optimization (MOO) procedure, in which the objectives are conflicting with each other and there is no higher level information (user preferences) beforehand that is leading to a biased search, is to find as many Pareto-optimal solutions as possible. Due to the population based search strategy of an evolutionary algorithm (EA), it is possible to obtain a set of Pareto-optimal solutions in a single simulation run. Besides aiming for a converged solution set as in a single objective optimization (SOO) case, it is also an important task for a multi-objective evolutionary algorithm (MOEA) to have well-distribution or diversity among these trade-off solutions, which is specific to MOO.

The first attempt to accomplish these two tasks mentioned above, although the second task was not fully solved, was the Vector Evaluated Genetic Algorithm (VEGA) which was suggested by Schaffer (1985). Then Goldberg (1989) suggested a better implementation of the domination principle (the non-domination sorting procedure) to emphasize the non-dominated members in the selection scheme as well as the use of the niching strategy to accomplish the second task, i.e. the diversity among the Pareto-optimal solutions. Following his discussions, a number of non-elitist MOEAs were proposed, e.g. the Multi-Objective Genetic Algorithm (MOGA) by Fonseca and Fleming (1993), the Niched Pareto Genetic Algorithm (NPGA) by Horn et al. (1994) and the Non-dominated Sorting Genetic Algorithm (NSGA) by Srinivas and Deb (1994), mainly having a different fitness assignment strategy. A few years later the importance of the elitism concept in MOO research was recognized and experimentally supported by different studies (Parks and Miller, 1998; Zitzler et al., 2000). As a result, a few elitist MOEAs, i.e. the Strength Pareto Evolutionary Algorithm (SPEA) by Zitzler and Thiele (1998, 1999) and the Pareto Archived Evolution Strategy (PAES) by Knowles and Corne (1998), started to show up and outperform the non-elitist approaches. In subsequent studies, further progress has been made and relatively new algorithms were proposed, e.g. the NSGA-II (the successor of the NSGA) by Deb et al. (2002), the Pareto Envelope-based Selection Algorithm (PESA) by Corne et al. (2000), and finally the SPEA2 by Zitzler et al. (2001) eliminating the potential weaknesses of its predecessor (SPEA).

Some of the features that may cause a multi-objective GA difficulty have been systematically investigated and procedures of constructing multi-objective test problems addressing different aspects of difficulties are presented by Deb (1999). The convergence towards the global Pareto-front (among local ones) as well as satisfying the well-distribution of these non-dominated solutions in a complex (convex or non-convex as well as continuous or discontinuous) objective space can be highlighted as major criteria.

This technical report presents several well-known unconstrained MOO benchmark problems (Zitzler et al., 2000; Deb, 2001) in order to investigate some of the difficulties as mentioned above that a MOEA may have to tackle. Three elitist algorithms, i.e. the MOGA-II by Poloni and Pediroda (1997), the NSGA-II by Deb et al. (2002) (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation

by the author in MATLAB with two versions, see the details in Section 2.3), are employed for this purpose. Since these benchmark problems are based on analytical funtions, the exact locations of the true Pareto-fronts are known beforehand and also shown on each figure in order to evaluate the performance of the algorithms clearly, i.e. the convergence to the true Pareto front and the diversity of the solutions along the non-dominated front. These test cases are composed of two minimization objectives and a number of design variables that is varying from 2 to 30.

2 A Brief Description of the Algorithms

In this section, a brief description of the three elitist algorithms, i.e. the MOGA-II by Poloni and Pediroda (1997), the NSGA-II by Deb et al. (2002) (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB), are presented.

2.1 Multi-objective Genetic Algorithm (MOGA-II)

MOGA-II is one of the new elitist evolutionary algorithms designed for MOO problems that is implemented in the commercial MOO software modeFRONTIER (ESTECO s.r.l., 2007). It is an improved version of the MOGA by Poloni and Pediroda (1997), but it should not be confused with the non-elitist MOGA developed by Fonseca and Fleming (1993).

As in classical genetic algorithms (Goldberg, 1989), the internal encoding in MOGA-II is based on binary coding that each design variable is represented as a binary string, i.e. 0100, 1101101, etc., where the length of the string depends on the number of allowed values for the variable. MOGA-II starts with the initial binary-coded population P of size N and an empty elite set $E = \emptyset$. $P' = P \cup E$ is computed every generation. If the population size of P' is greater than the population size of P, the exceeding members in P' are randomly removed in order to reduce the population size to the initial size. Following this step, the evolution takes place in which P' is updated to P'' by applying all MOGA operators, i.e. classical cross-over, directional cross-over, mutation and selection.

The classical (single-point) cross-over is a well-known genetic algorithm operator in which two parents (individuals from the mating pool) are randomly chosen and some portions of these binary strings are exchanged between each other. In single-point cross-over, the crossing site is randomly chosen and all bits on the right side of the crossing site are swapped as shown schematically in Fig. 2.1. In MOGA-II, single point cross-over starts by taking the current individual-i as the first parent. The second parent (individual-j) is chosen by means of a multi-objective tournament selection on a randomly selected population subset (Poles et al., 2007, 2004; Rigoni and Poles, 2005). After the cross-over operation is performed, one of the resulting strings (children) is randomly selected to be the new individual.

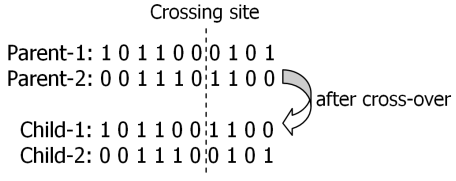


Figure 2.1: Schematic procedure of single-point crossover.

The *directional cross-over* assumes that a *direction of improvement*, which reminds the *search direction* in classical algorithms that is relatively more expensive to compute, can be detected by comparing the fitness values of two reference individuals selected from the same generation similar to the idea presented by Yamamoto and Inoue (1995). The selection of these two reference individuals (parents) can be done using any available selection scheme. In MOGA-II, potential candidates are met in a random walk of assigned number of steps for each individual and then the two of them with the best fitnesses are chosen. The new individual is then created by moving in a randomly weighted direction that lies within the ones originated by the starting point (current individual) and two other parents selected by the local tournament, as shown in Fig. 2.2.

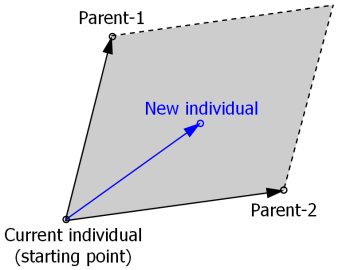


Figure 2.2: Schematic procedure of directional crossover in MOGA-II (Poles et al., 2004, 2007; Rigoni and Poles, 2005).

The *mutation* operator is responsible for perturbing the binary string, i.e. changing 1 to 0 or vice versa, with a given *DNA String Mutation Ratio* (probability). The mutation operator can be said to be responsible of the local search as well as maintaining the diversity in the population (Deb, 2006).

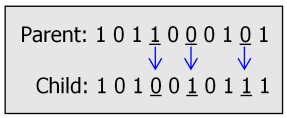


Figure 2.3: Schematic procedure of mutation for a given DNA string mutation ratio, p_m=30%.

The fitness of the population P'' is calculated. Non-domination sorting is performed and all non-dominated designs of P'' are copied to the elite set, E which is then updated by removing duplicated or dominated designs. Next, E is resized if it is larger than the population size N removing randomly the exceeding points. Finally, the first generation is created and P'' is assigned as the new initial population P for the next generation. This iterative loop continues until the stopping criterion is met, i.e. the user-defined total number of generations (Poles, 2003).

2.2 Non-dominated Sorting Genetic Algorithm (NSGA-II)

NSGA-II is a multi-objective evolutionary algorithm (MOEA) proposed by Deb et al. (2002), which enables finding well-spread multiple Pareto-optimal solutions for an MOO

problem by incorporating three substantial features, i.e., elitism, non-dominated sorting, and a diversity preserving mechanism. Elitism, which is a genetic operator used to preserve the good solutions through the generations, is also shown to increase the convergence speed of an MOEA Zitzler et al. (2000).

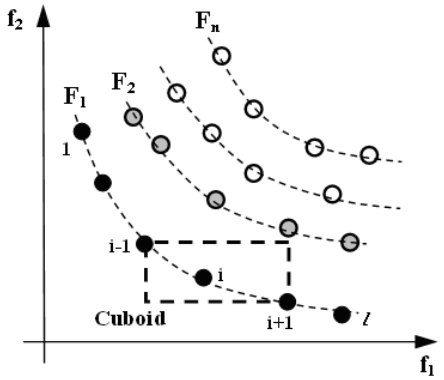


Figure 2.4: The crowding distance calculation (adapted from Deb et al. (2002)).

The preservation of well-distributed (diverse) nondominated solutions is maintained by the crowding distance operator. The crowding distance, which is calculated frontwise, is a measure of how close an individual is to its neighbors. In order to get an estimate of the density of solutions surrounding a particular solution in the population, the average distance between two points on either side of the solution along each of the objectives is calculated. This quantity d_i serves as an estimate of the perimeter of the cuboid formed by using the nearest neighbors as the vertices. A schematic view of the procedure for the computation of the crowding distance for an i^{th} individual is shown in Fig. 2.4, and formulated as (Deb et al., 2002),

$$d_{I_i^m} = d_{I_i^m} + \frac{f_m^{(I_{i+1}^m)} - f_m^{(I_{i-1}^m)}}{f_m^{max} - f_m^{min}}, \qquad for\ i = 2, ..., l-1 \tag{2.1}$$

where I(i) is the solution index of the i^{th} member in the (worse order) sorted list, m is the number of objectives and l is the number of the individuals in that particular front. The first and the last individuals are assigned an infinite distance in order to emphasize their fitness, hence preserving them for the next generation. A large average crowding distance will result in a better diversity in the population.

The schematic view of NSGA-II on a flowchart is shown in 2.5. The procedure starts with an initialization of the parent population, P_0, followed by sorting (ranking) the members of the population based on nondomination into each front, F_i. Each individual is assigned two fitness values, i.e. the one related to its nondomination level (front or rank, i.e. F_i refers to front with rank i in Fig. 2.4) and the crowding distance. Parents are selected by using binary (crowded) tournament selection based on these two measures. An individual is selected if its rank is less than the other, since the minimization is assumed, or if the crowding distance is greater than the other. The selected population generates off-springs, Q_t, by the application of usual genetic operators, i.e. crossover and mutation (simulated binary crossover (SBX), (Deb and Agrawal, 1995; Deb and Kumar,

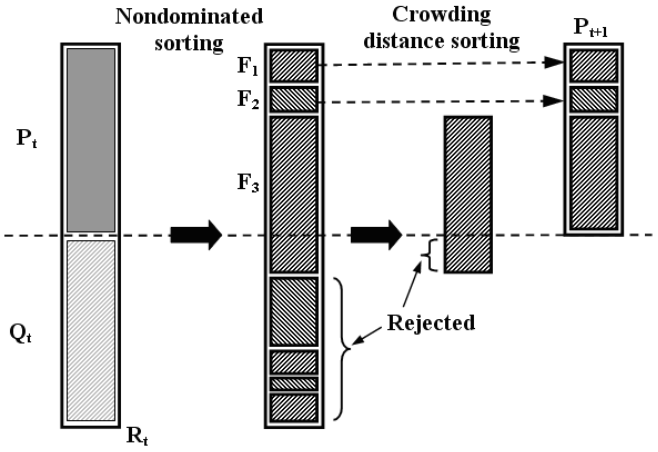


Figure 2.5: Schematic view of the NSGA-II procedure (Deb et al., 2002).

1995) and polynomial mutation for real parameter GA in this case) (Deb, 2001; Goldberg, 1989). Since all the parent and the offspring individuals are included in the combined population, R_t, elitism is ensured. In order to reduce the size of the population of the next generation from $2N$ to N, nondomination sorting is applied and the individuals with relatively lowest possible ranks are accomodated until the size of the population exceeds N. Then the individuals in front F_i are selected based on their crowding distance. This cycle is repeated until the stopping criterion, i.e. the total number of generations, is met. For further details, please consult the original study (Deb et al., 2002).

2.3 Custom implementation of the NSGA-II (cNSGA-II)

This custom (personal) implementation of the NSGA-II (cNSGA-II) by the author should not be confused with the clustered NSGA-II (C-NSGA-II) (Deb et al., 2003) which uses an identical clustering technique to that used in the SPEA (Zitzler and Thiele, 1998, 1999) in order to obtain a well-spread distribution.

The cNSGA-II has a minor change compared to the original NSGA-II in the application of the SBX and polynomial mutation operators. Instead of applying these operators with predefined probabilities for each, a fixed percentage of the population (90%) has been randomly selected as the set of parents in order to produce the children with a predefined distribution index (η_c). Following this, polynomial mutation is applied on the rest of the population (10%), which didn't participate in the SBX operation. By doing this, it is ensured that all members will participate in these operations. The percentage of the population on which the SBX operator is applied is chosen higher compared to the one for the polynomial mutation (i.e. 90% vs. 10%) in order to enforce the search on a larger portion of the search region, in addition to the application of a relatively smaller distribution index (resulting in a wider spread with respect to the parents, aiming at a uniform distribution in the search space). The polynomial mutation operator is applied with a relatively higher distribution index in order to make a finer search (producing the children closer to the parents).

The computation of the fitness values are vectorized instead of looping over each

individual. This kind of minor high performance computing capabilities have been utilized. The performance of the code has been optimized using the MATLAB Profiler. Binary-coding of the parameters has also been implemented, but in this particular study, real-parameter GAs have mainly been considered due to their advantages in continuous search space.

Since the other two algorithms implemented in modeFRONTIER plot the global Pareto-optimal set (not only those obtained at the last generation) after an extra book-keeping, the cNSGA-II is also modified in order to compare the three algorithms on an equal basis. The non-dominated solutions obtained in each generation has been kept in an extra dynamic array. After all non-dominated solutions are enrolled, a non-dominated sorting is applied in order to determine the global Pareto-optimal set. Results of both of these implementations have been presented in this report.

2.4 Algorithm Settings

As mentioned above, the three MOEAs, i.e. the MOGA-II, the NSGA-II (implemented in modeFRONTIER) and the cNSGA-II (custom implementation of the NSGA-II by the author, i.e. with an without extra Pareto archieve for only post-processing issues), were used to investigate some of the difficulties that a MOO algorithm may confront. The settings for each of these three algorithms are given below. Both of the MOGA-II and the NSGA-II algorithms were started with the initial populations provided by the Sobol sequence (Sobol, 1979) which is a *quasi-random* distribution filling the design space in a uniform manner. The cNSGA-II algorithm started with a randomly (provided by MATLAB) distributed initial population. Some common features for the three algorithms, i.e. population size, total number of generations and elitism are given below.

Common settings for the three algorithms:

`Population Size : 100`

`Number of Generations : 250`

`Elitism : Enabled`

The directional cross-over operator, described as in Section 2.1, provides an efficient search, on the other hand, a higher probability of this operator decreases its robustness, i.e. in case of highly non-linear problems the optimizer might be stuck at local optimal solutions (local Pareto-fronts). The probability of the classical (single-point) cross-over is related to the probability of the directional cross-over in the way that the cumulative summation of the probabilities of directional and classical cross-overs, selection and mutation operators is equal to 1. This means that an increase in the probability of directional cross-over results in a decrease in the probability of classical cross-over. A higher probability of the classical cross-over operator provides a more robust search since it allows the algorithm to jump more easily into different regions of the search space, on the other hand, it will decrease the convergence speed as expected.

The MOGA-II algorithm settings (Poles, 2003):

`Initial population : Sobol Sequence`

Prob. of Directional Cross-over : 65%

Prob. of Classical Cross-over : 20%

Prob. of Selection : 5%

Prob. of Mutation : 10%

DNA String Mutation Ration : 10%

The NSGA-II algorithm (suggested as default) settings:

Initial population : Sobol Sequence

Prob. of Cross-over : 90%

Prob. of Mutation : 10%

Distribution Index for Real-coded Cross-over : 20

Distribution Index for Real-coded Mutation : 20

The cNSGA-II algorithm settings:

Initial population : Random Distribution

Probability Distribution for the SBX, η_c : 2

Probability Distribution for the polynomial mutation, η_m : 5

3 Test Functions and Results

In this section, seven unconstrained MOO benchmark problems have been presented. In each problem, minimization of two objective functions has been considered. In the last five well-known test cases, which were proposed by Zitzler et al. (2000), a systematic approach has been followed to investigate a different aspect of difficulty that a MOEA may face. These test problems have a general structure as given in Eq. 3.1,

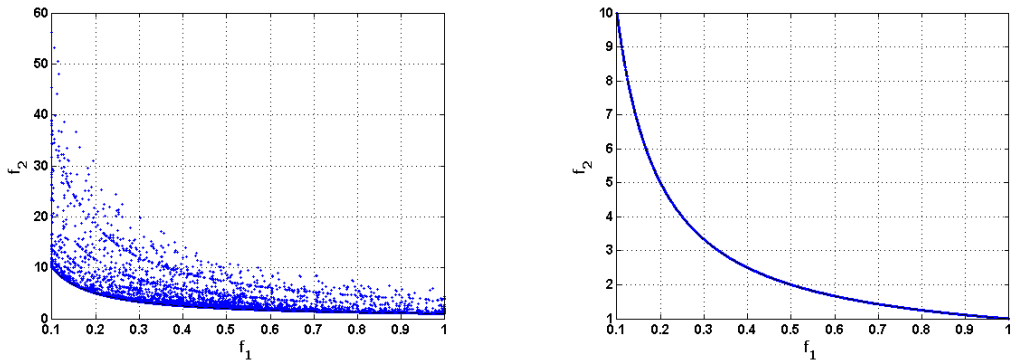
$$\begin{aligned} &Minimize : f_1(\mathbf{x}), \\ &Minimize : f_2(\mathbf{x}) = g(\mathbf{x})\, h(f_1(\mathbf{x}), g(\mathbf{x})). \end{aligned} \tag{3.1}$$

where f_1 is a function of only x_1 in each case, while the formulations of $g(\mathbf{x})$ and $h(f_1(\mathbf{x}), g(\mathbf{x}))$ vary together with the size of the design vector $\mathbf{x}$, i.e. from 2 to 30. In all these five cases, the Pareto-optimal front is formed with $g(\mathbf{x})$=1.

3.1 DEB

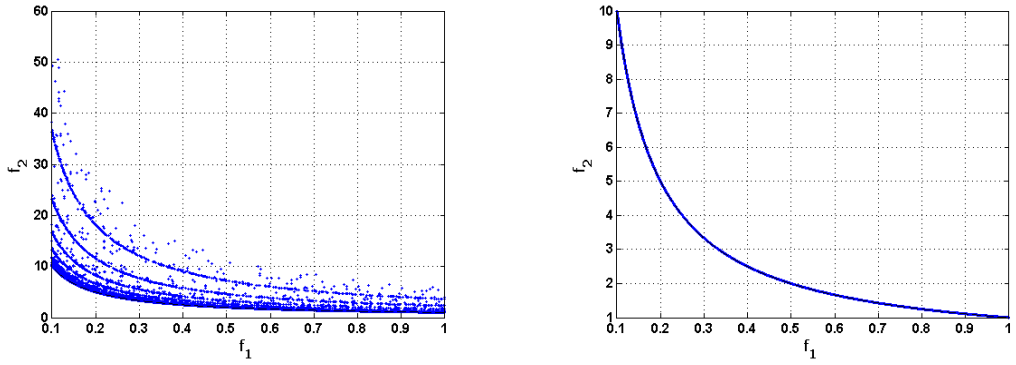
This two-objective, two-variable minimization problem, as shown in Eq. 3.2, has been proposed by Deb (2001) in order to illustrate the conflicting scenarios between both objectives which result in a set of Pareto-optimal solutions. The function f_1 is a function of x_1 only, thus f_1 is limited by the lower and upper limits of the x_1, while f_2 is in contrast with the first one, i.e. increasing f_1 will result in decreasing f_2 and vice versa. The Pareto-front, which corresponds to the continuous curve drawn at $0.1 \leq x_1 \leq 1$ and x_2=0, is convex. The three performances are presented below in Figures 3.1 through 3.4.

$$\begin{aligned} &Minimize : f_1(x_1) = x_1, \\ &Minimize : f_2(f_1, x_2) = \frac{1 + x_2}{x_1}, \\ &subject\ to : x_1 \in [0.1, 1], \\ &\qquad\qquad x_2 \in [0, 5]. \end{aligned} \tag{3.2}$$



(a) (b)

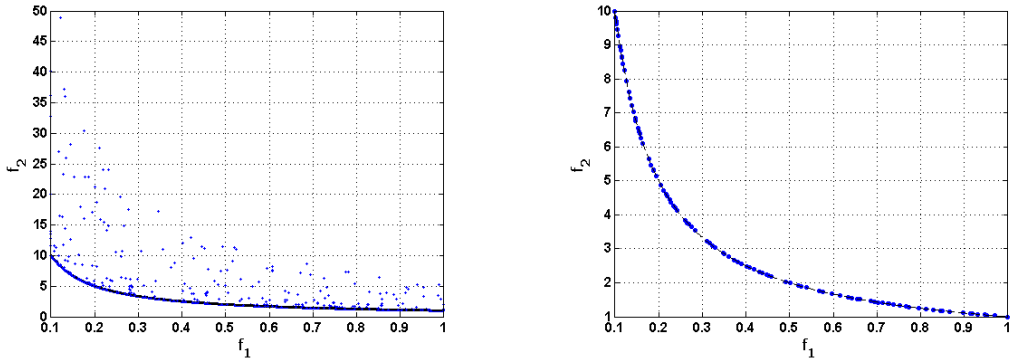
Figure 3.1: a) All individuals obtained during the MOGA-II run for test function DEB. b) Pareto-set obtained at the end of the MOGA-II run for test function DEB.



(a) (b)

Figure 3.2: a) All individuals obtained during the NSGA-II run for test function DEB. b) Pareto-set obtained at the end of the NSGA-II run for test function DEB.

As shown in Figs. 3.1b and 3.2b, the cardinality of the Pareto-sets in application of both algorithms, i.e. MOGA-II and NSGA-II (in modeFRONTIER), is higher than the initial population size (N=100) which can be interpreted as a post-processing issue. A further book-keeping of all the non-dominated solutions in each generation which are kept in an extra array and then sorted with respect to the non-domination criterion as a final graph. The original implementation of the NSGA-II algorithm (Deb et al., 2002) plots the Pareto-set, which corresponds to the set of non-dominated solutions at the last generation, having the initial population size. In order to compare these two algorithms with the cNSGA-II on the same basis, the results of the second version of the cNSGA-II have also been presented.



(a) (b)

Figure 3.3: a) All individuals obtained during the cNSGA-II run for test function DEB. b) Pareto-set obtained at the end of the cNSGA-II run for test function DEB.

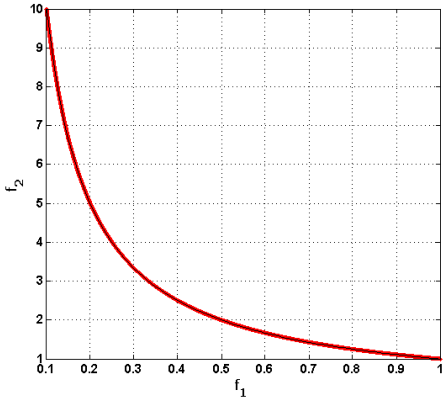
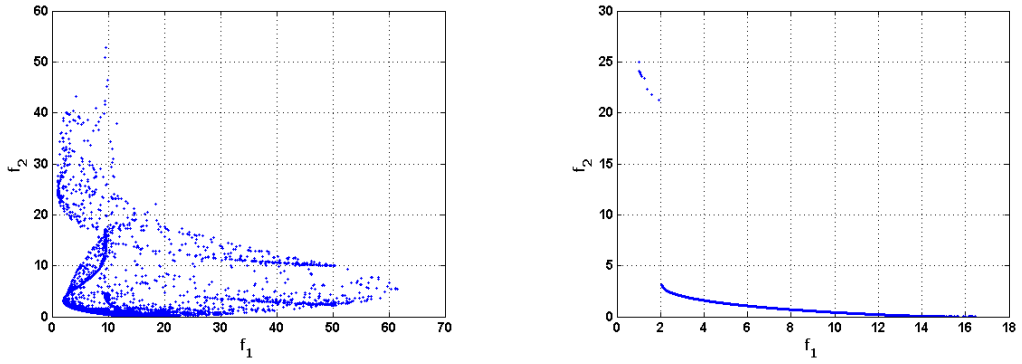


Figure 3.4: Global Pareto set obtained using the cNSGA-II with the extra Pareto archieve for test function DEB.

3.2 POL

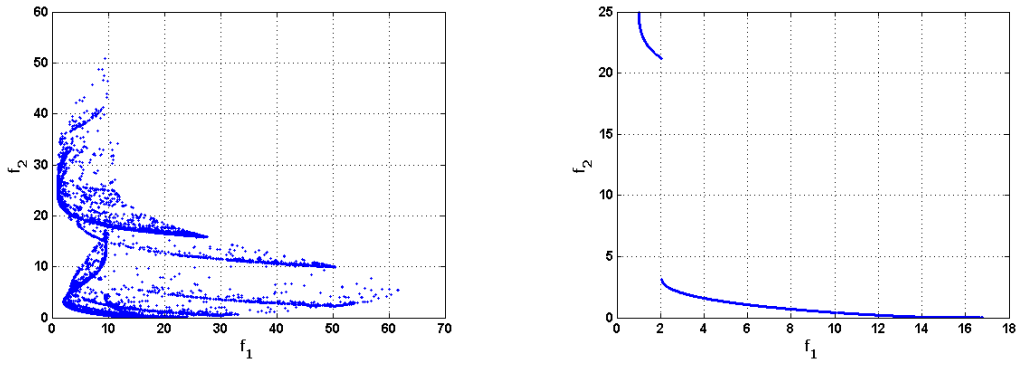
This second benchmark problem, which is composed of two objectives and two variables, is used by Poloni et al. (2000). It has a discontinuous, discretely spaced continuous, Pareto-optimal set which makes it difficult to know the solution beforehand.

$$
\begin{aligned}
&Minimize: f_1(x_1, x_2) = \left[1 + (A_1 - B_1)^2 + (A_2 - B_2)^2\right], \\
&Minimize: f_2(x_1, x_2) = \left[(x_1 + 3)^2 + (x_2 + 1)^2\right], \\
&\quad where \quad A_1 = 0.5\sin(1) - 2\cos(1) + \sin(2) - 1.5\cos(2), \\
&\qquad\qquad A_2 = 1.5\sin(1) - \cos(1) + 2\sin(2) - 0.5\cos(2), \\
&\qquad\qquad B_1 = 0.5\sin(x_1) - 2\cos(x_1) + \sin(x_2) - 1.5\cos(x_2), \\
&\qquad\qquad B_2 = 1.5\sin(x_1) - \cos(x_1) + 2\sin(x_2) - 0.5\cos(x_2), \\
&subject\ to: -\pi \leq x_1, x_2 \leq \pi.
\end{aligned}
\tag{3.3}
$$



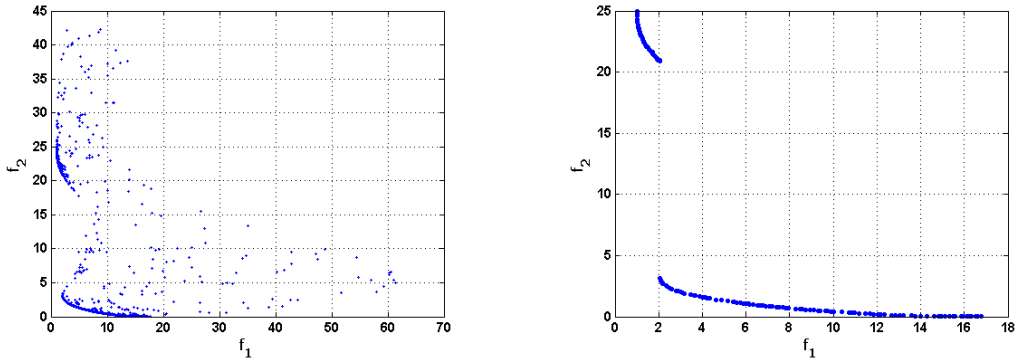
(a) (b)

Figure 3.5: a) All individuals obtained during the MOGA-II run for test function POL. b) Pareto-set obtained at the end of the MOGA-II run for test function POL.



(a) (b)

Figure 3.6: a) All individuals obtained during the NSGA-II run for test function POL. b) Pareto-set obtained at the end of NSGA-II run for test function POL.



(a) (b)

Figure 3.7: a) All individuals obtained during the cNSGA-II run for test function POL. b) Pareto-set obtained at the end of the cNSGA-II run for test function POL.

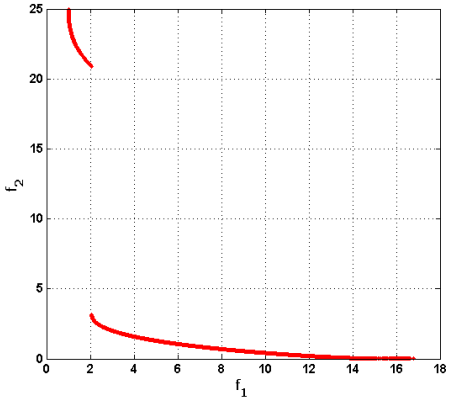
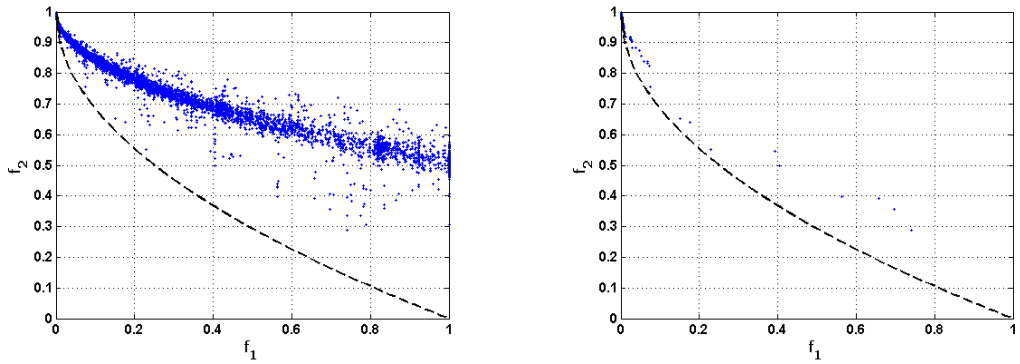


Figure 3.8: Global Pareto set obtained using cNSGA-II with the extra Pareto archieve for test function POL.

3.3 ZDT1

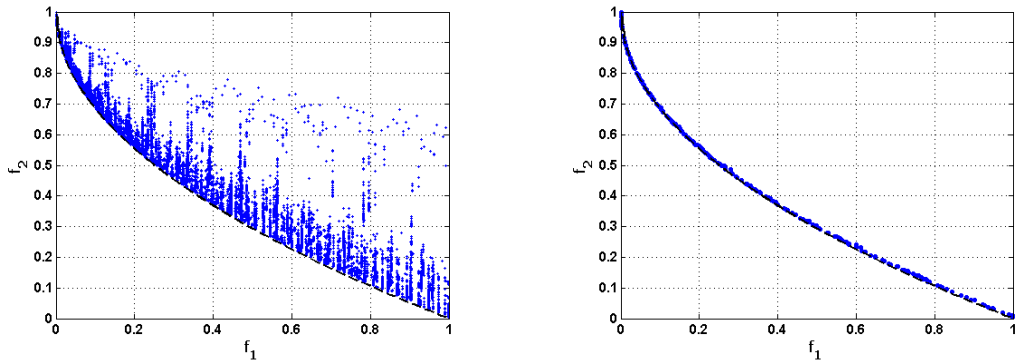
This benchmark problem is part of `Zitzler-Deb-Thiele`'s (ZDT) test problems on MOO composed of 6 cases (Zitzler et al., 2000). It has 2 objective functions and 30 design variables, $m = 30$, resulting in a convex true Pareto-front having a uniform distribution of solutions accross the front. All variables lie in the range $x_i \in [0, 1]$. The Pareto-optimal region corresponds to $0 \leq x_1 \leq 1$ and $x_i = 0$ ($i = 2, 3, ..., m$). The Pareto-optimal front is shown on the following figures 3.9 through 3.12 as the dashed line together with the solutions obtained by the three algorithms.

$$
\begin{aligned}
&Minimize : f_1(x_1) = x_1, \\
&Minimize : f_2(f_1, g) = 1 - \sqrt{\frac{f_1}{g}}, \\
&\quad where \quad g(x_i) = 1 + \frac{9}{m-1}\sum_{i=2}^{m} x_i, \\
&subject\ to : x_i \in [0,1], \quad i = 1, 2..., m. \quad (m = 30)
\end{aligned}
\tag{3.4}
$$



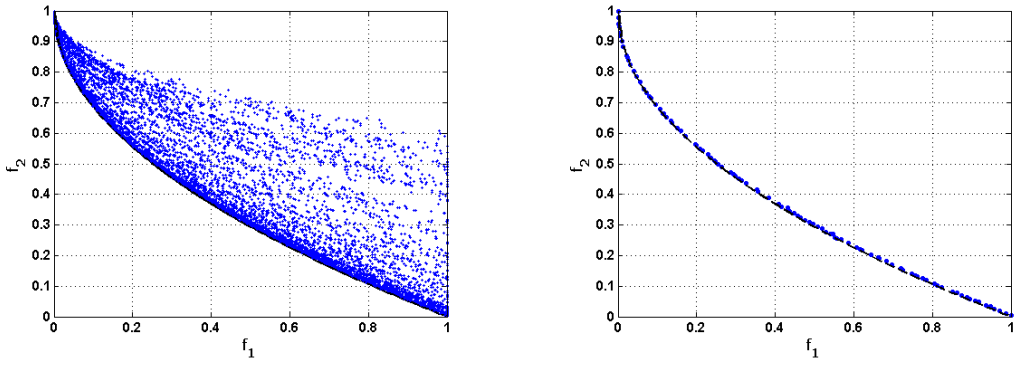
(a) (b)

Figure 3.9: a) All individuals obtained during the MOGA-II run for test function ZDT1. b) Pareto-set obtained at the end of the MOGA-II run for test function ZDT1.



(a) (b)

Figure 3.10: a) All individuals obtained during the NSGA-II run for test function ZDT1. b) Pareto-set obtained at the end of the NSGA-II run for test function ZDT1.



(a) (b)

Figure 3.11: a) All individuals obtained during the cNSGA-II run for test function ZDT1. b) Pareto-set obtained at the end of the cNSGA-II run for test function ZDT1.

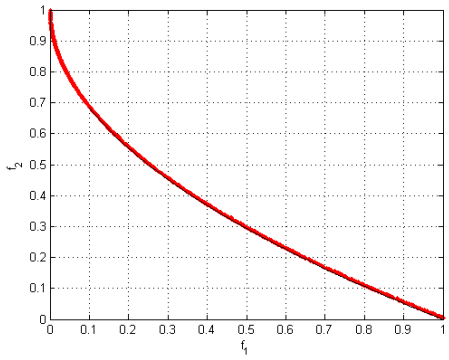
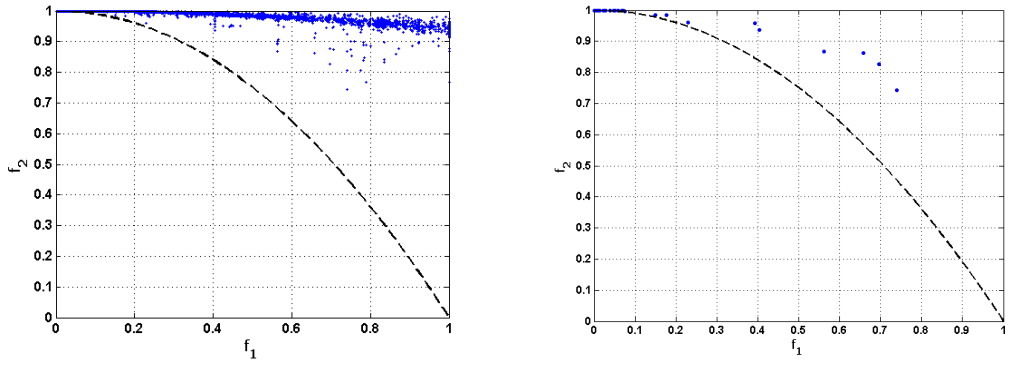


Figure 3.12: Global Pareto set obtained using the cNSGA-II with the extra Pareto archieve for test function ZDT1.

3.4 ZDT2

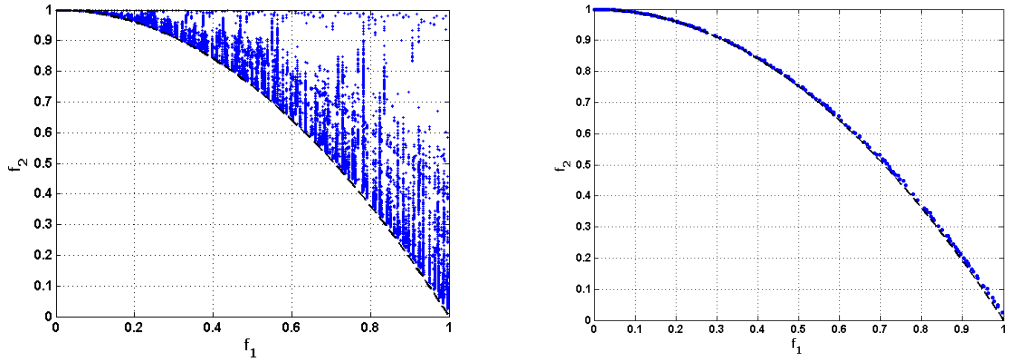
The ZDT2 problem has two objective functions and m=30 variables, which lie in the range $x_i \in [0,1]$. The Pareto-optimal region corresponds to $0 \leq x_1 \leq 1$ and $x_i = 0$ (i=2, 3, ..., m), resulting in a non-convex true Pareto-front having a uniform distribution of solutions accross the front. The Pareto-optimal front for this problem is shown on the following figures 3.13 through 3.16 as the dashed line together with the solutions obtained by the three algorithms.

$$\begin{aligned} &Minimize: f_1(x_1) = x_1, \\ &Minimize: f_2(f_1, g) = 1 - \left(\frac{f_1}{g}\right)^2, \\ &\quad where \quad g(x_i) = 1 + \frac{9}{m-1}\sum_{i=2}^{m} x_i, \\ &subject\ to: x_i \in [0,1], \quad i = 1, 2..., m. \quad (m = 30) \end{aligned} \tag{3.5}$$



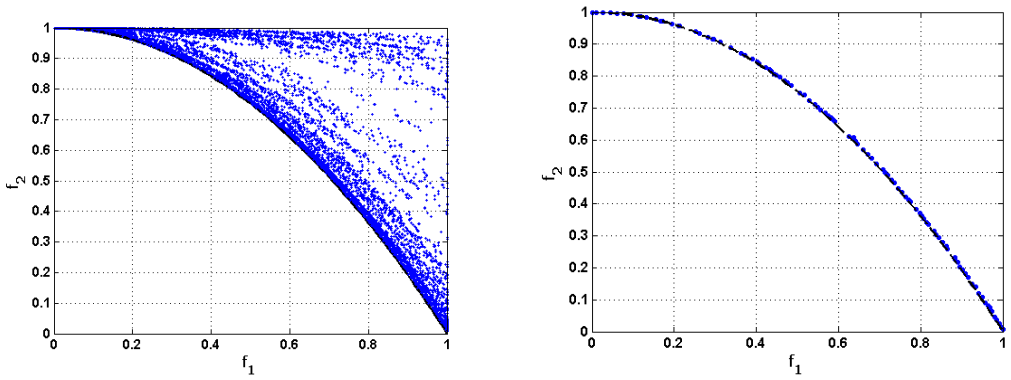
(a) (b)

Figure 3.13: a) All individuals obtained during the MOGA-II run for test function ZDT2. b) Pareto-set obtained at the end of the MOGA-II run for test function ZDT2.



(a) (b)

Figure 3.14: a) All individuals obtained during the NSGA-II run for test function ZDT2. b) Pareto-set obtained at the end of the NSGA-II run for test function ZDT2.



(a) (b)

Figure 3.15: a) All individuals obtained during the cNSGA-II run for test function ZDT2. b) Pareto-set obtained at the end of the cNSGA-II run for test function ZDT2.

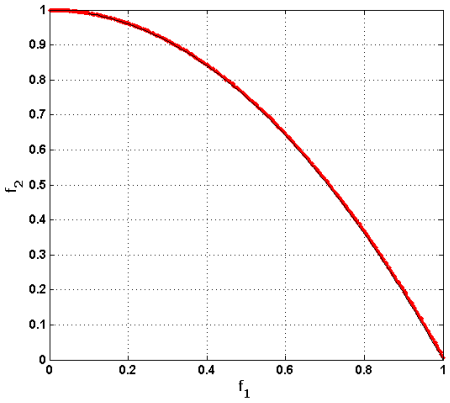
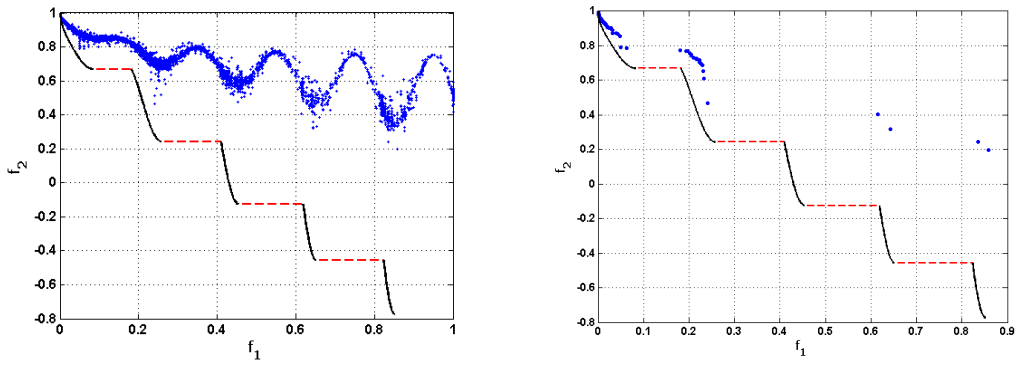


Figure 3.16: Global Pareto set obtained using the cNSGA-II with the extra Pareto archieve for test function ZDT2.

3.5 ZDT3

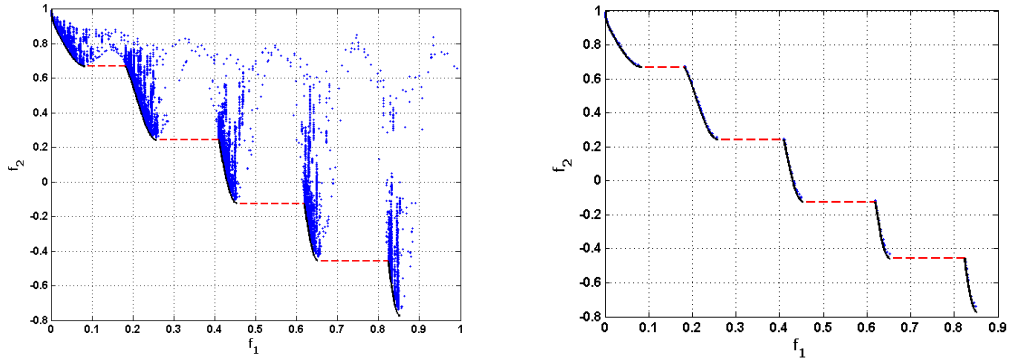
This benchmark problem has also two objective functions and m=30 variables, which lie in the range $x_i \in [0, 1]$. In this case, the Pareto-optimal region corresponds to x_i=0 (i=2, 3, ..., m), but not all 0$\leq x_1 \leq$1, resulting in a number of disconnected Pareto-fronts having uniform distribution of solutions accross the fronts. The Pareto-optimal front for this problem is shown on the following figures 3.13 through 3.16 as the solid black lines (red dashed lines are drawn to distinguish the discontinuous sections) together with the solutions obtained by the three algorithms.

$$
\begin{aligned}
&Minimize: f_1(x_1) = x_1, \\
&Minimize: f_2(f_1, g) = 1 - \sqrt{\frac{f_1}{g}} - \left(\frac{f_1}{g}\right) \sin(10\pi f_1), \\
&\quad where \quad g(x_i) = 1 + \frac{9}{m-1} \sum_{i=2}^{m} x_i, \\
&subject\ to: x_i \in [0,1], \quad i = 1, 2..., m. \quad (m = 30)
\end{aligned}
\tag{3.6}
$$



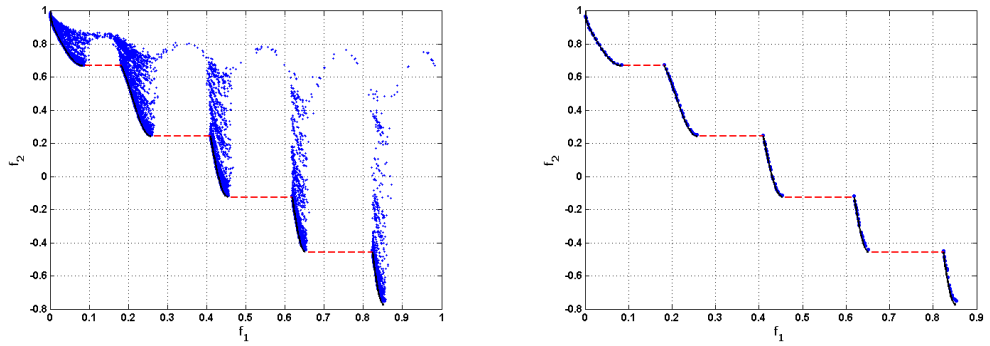
(a) (b)

Figure 3.17: a) All individuals obtained during the MOGA-II run for test function ZDT3. b) Pareto-set obtained at the end of the MOGA-II run for test function ZDT3.



(a) (b)

Figure 3.18: a) All individuals obtained during the NSGA-II run for test function ZDT3. b) Pareto-set obtained at the end of the NSGA-II run for test function ZDT3.



(a) (b)

Figure 3.19: a) All individuals obtained during the cNSGA-II run for test function ZDT3. b) Pareto-set obtained at the end of the cNSGA-II run for test function ZDT3.

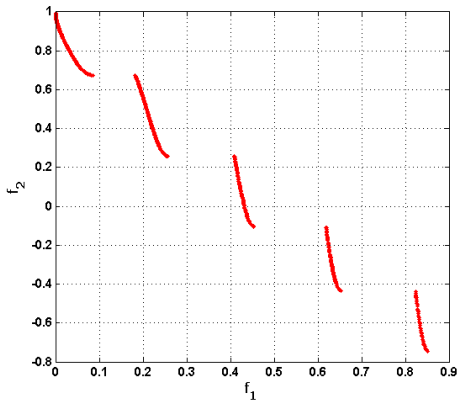
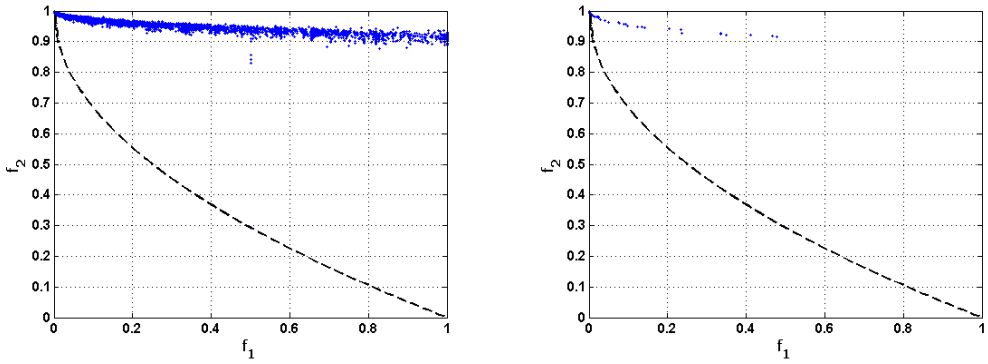


Figure 3.20: Global Pareto set obtained using the cNSGA-II with the extra Pareto archieve for test function ZDT3.

3.6 ZDT4

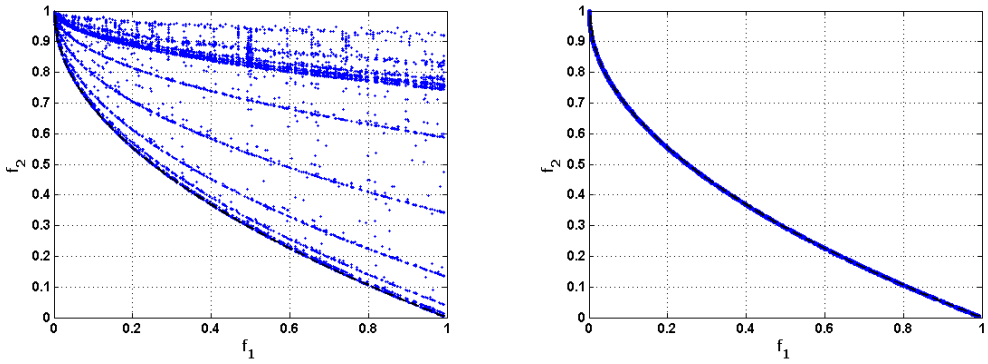
This benchmark problem has m=10 variables together with two objective functions. The variable x_1 lies in the range [0,1], while others lie in [-5,5]. The Pareto-optimal front curve is convex. This problem has been prepared to test a MOEA in the presence of many local Pareto-optimal sets, further discussion can be found in Zitzler et al. (2000) and Deb (2001).

$$
\begin{aligned}
&Minimize: f_1(x_1) = x_1, \\
&Minimize: f_2(f_1, g) = 1 - \sqrt{\frac{f_1}{g}}, \\
&\quad where \quad g(x_i) = 1 + 10(m-1) + \sum_{i=2}^{m} \left(x_i^2 - 10\cos(4\pi x_i)\right), \\
&subject\ to: x_1 \in [0,1], \\
&\qquad\qquad x_i \in [-5,5], \quad i = 2,3...,m. \quad (m = 10)
\end{aligned}
\tag{3.7}
$$



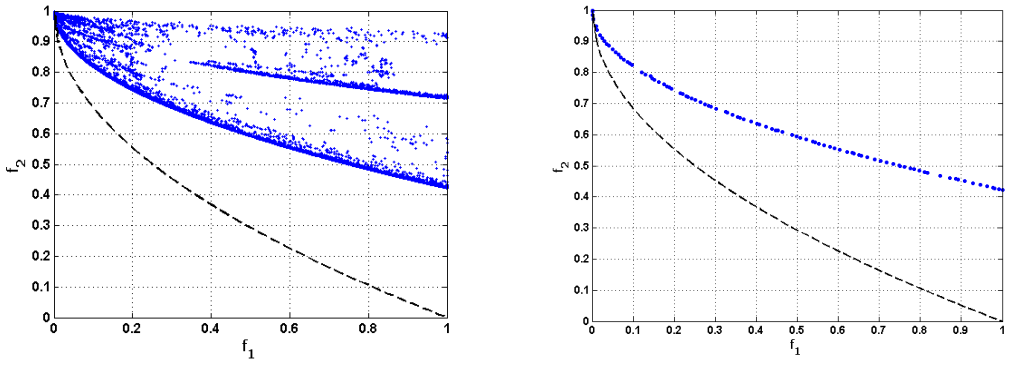
(a) (b)

Figure 3.21: a) All individuals obtained during the MOGA-II run for test function ZDT4. b) Pareto-set obtained at the end of the MOGA-II run for test function ZDT4.



(a) (b)

Figure 3.22: a) All individuals obtained during the NSGA-II run for test function ZDT4. b) Pareto-set obtained at the end of the NSGA-II run for test function ZDT4.



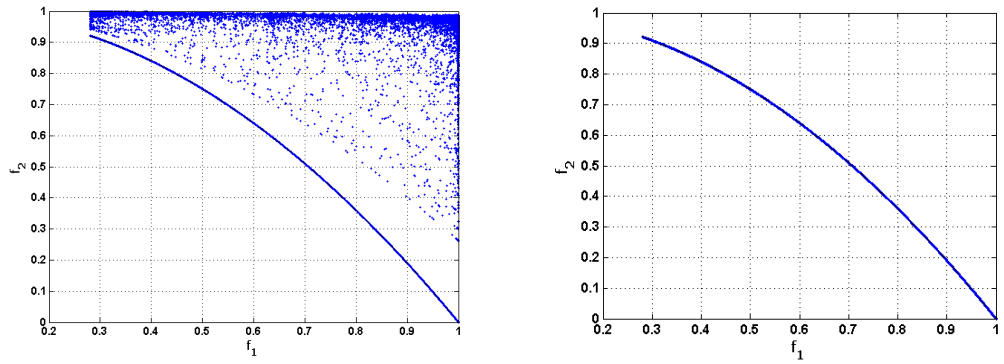
(a) (b)

Figure 3.23: a) All individuals obtained during cNSGA-II run for test function ZDT4. b) Pareto-set obtained at the end of the cNSGA-II run for test function ZDT4.

3.7 ZDT6

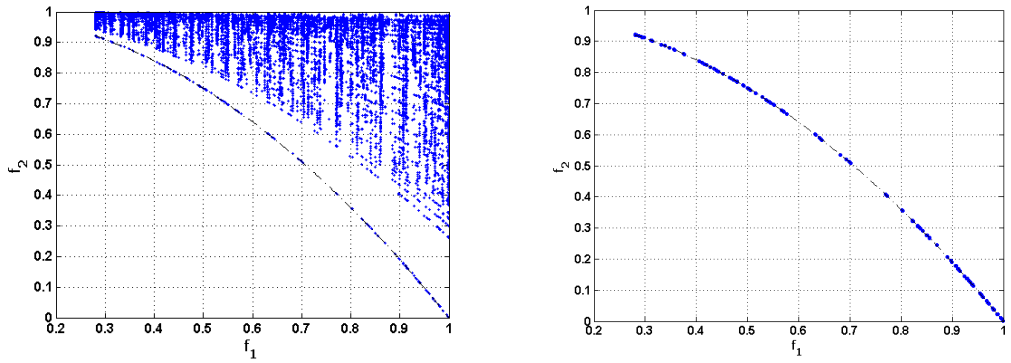
This benchmark problem has also 2 objective functions and m=10 variables having a non-convex Pareto-optimal set. Actually the difficulty in this problem is that the Pareto-optimal region has a non-uniform distribution of solutions and a decreasing density towards the Pareto-optimal front is also observed from the Figs. 3.24 through 3.27.

$$\begin{aligned} &Minimize: f_1(x_1) = 1 - \exp(-4x_1)\sin^6(6\pi x_1), \\ &Minimize: f_2(f_1, g) = 1 - \left(\frac{f_1}{g}\right)^2, \\ &\quad where \quad g(x_i) = 1 + 9\left[\sum_{i=2}^{m}\frac{x_i}{m-1}\right]^{0.25}, \\ &subject\ to: x_i \in [0,1], \quad i = 1, 2..., m. \quad (m = 10) \end{aligned} \tag{3.8}$$



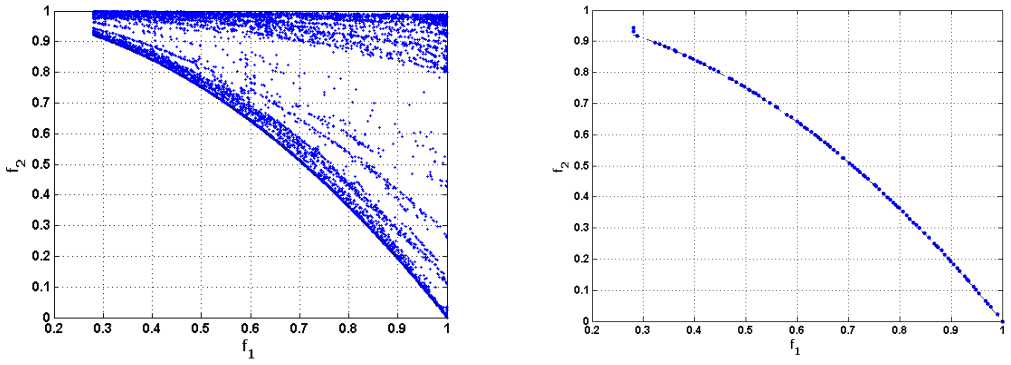
(a) (b)

Figure 3.24: a) All individuals obtained during the MOGA-II run for test function ZDT6. b) Pareto-set obtained at the end of the MOGA-II run for test function ZDT6.



(a) (b)

Figure 3.25: a) All individuals obtained during the NSGA-II run for test function ZDT6. b) Pareto-set obtained at the end of the NSGA-II run for test function ZDT6.



(a) (b)

Figure 3.26: a) All individuals obtained during the cNSGA-II run for test function ZDT6. b) Pareto-set obtained at the end of the cNSGA-II run for test function ZDT6.

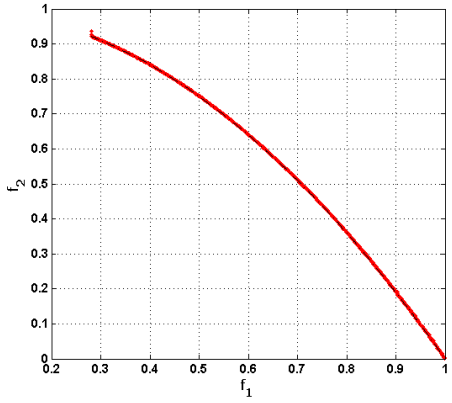


Figure 3.27: Global Pareto set obtained using the cNSGA-II with the extra Pareto archieve for test function ZDT6.

4 Discussion of the Results and Conclusions

In this technical report, several well-known unconstrained MOO benchmark problems (Zitzler et al., 2000; Deb, 2001) have been presented in order to investigate some of the difficulties that a MOEA may have to tackle. Three elitist algorithms, i.e. the MOGA-II by Poloni and Pediroda (1997), the NSGA-II by Deb et al. (2002) (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB), are employed for this purpose. Due to the availability of the exact location and dimensions of the true Pareto-optimal fronts beforehand, the performance of the algorithms (i.e. the convergence to the true Pareto front and diversity of the solutions along the non-dominated front) could be investigated clearly. All of these test cases, which were sorted in an alphabetical order, are composed of two minimization objectives and a number of design variables that is varying between 2 and 30.

The first test function DEB was one of the easiest among all benchmark problems presented. The solutions of all the three algorithms were able to converge to the true Pareto-optimal front without any problem. Another aspect of the MOO problems, which is the diversity among these non-dominated solutions, was also succesfully accomplished by each algorithm. It is noteworthy to distinguish the convergence rate of the algorithms. The MOGA-II and NSGA-II (those implemented in modeFRONTIER), besides having a good overall convergence behavior, had also been iterating between the local Pareto-optimal fronts while converging to the global one. On the other hand, from this perspective, the cNSGA-II had shown a better performance by having fewer computations away from the global Pareto-optimal front. Moreover, the two algorithms (those implemented in modeFRONTIER), post-process the final global Pareto-set that is composed of the non-dominated members which were archived during the generations and finally sorted with respect to domination criterion. In order to compare the performance (i.e. the number or density of the non-dominated solutions accross the Pareto-front) of these two algorithms with the cNSGA-II on an equal basis, this feature was also implemented, indicated as the second version. Thus, two types of post-processing have been performed for the cNSGA-II, i.e. first, plotting the Pareto-set obtained at the last generation, and second, those saved through all generations. For the first test problem, all the three algorithms (the MOGA-II, the NSGA-II and the second version of the cNSGA-II) performed similar in the density of the solutions viewpoint.

The second test problem (POL) presented in this report is the first one which has a discontinuous Pareto-front, besides ZDT3. The three performances are almost similar, but the cNSGA-II shows the same fast convergence behavior as in the first test case. The next two test problems, ZDT1 and ZDT2, both have a continuous Pareto-front, but it is convex in the former and the non-convex in the latter. In both of these problems, the MOGA-II had difficulties to converge to the global Pareto-optimal front, while the other two algorithms performed well. The fifth problem is the ZDT3, which has a discontinuous Pareto-front as mentioned before. The MOGA-II again had difficulties in obtaining a converged set of solutions while the other two didn't have any problems. The next problem, the ZDT4, is one of the most difficult benchmark problems. It has a convex

Pareto-front and less design variables compared to ZDT1, ZDT2 and ZDT3, but the difficulty is that it has many local Pareto-optimal fronts. In this case, both the MOGA-II and the cNSGA-II couldn't succeed to converge to the global one. The second version of the cNSGA-II hasn't been tested due to the failure observed in the first one. The NSGA-II, which has been implemented in the modeFRONTIER outperformed the others for the ZDT5. The last benchmark problem ZDT6, puts forward some difficulties other than discontinuity or non-convexity of the Pareto front. In this case, the Pareto-optimal region has a non-uniform distribution of solutions and a decreasing density towards the true Pareto-front. The MOGA-II has a very good performance regarding both the convergence and the well-spread distribution of the non-dominated solutions accross the Pareto-front. The first version of the cNSGA-II shows a similar performance, moreover the second version has a continuous Pareto-curve identical to the analytical solution. The solutions obtained by the NSGA-II (in modeFRONTIER) are not as dense and well-spread as the those obtained by the other two algorithms.

One could argue whether there is a need for an extra post-processing step, since from a practical point of view, having a well-defined (i.e. full coverage of upper and lower limits of the Pareto-front) and well-distributed trade-off set which allows the user (e.g. an engineer) to have all kinds of combination of each success criterion beforehand and without any initial preference, is crucial enough.

As the last words, it is fair to say that the performance of the cNSGA-II is quite satifying, especially in terms of convergence rate in problems having 2 to 10 design variables. This could be explained by, first, having a slight difference in the application of the cross-over (SBX) and mutation operators, as explained in Section 2.3, and second, due to differences in implementation of the tournament selection operator, which promotes the convergence as well as the diversity, in the three algorithms. At least, concerning only the NSGA-II algorithm and its huge popularity, it should be mentioned that there are a few other versions of implemetations of this algorithm (in MATLAB) which were also tested by the author and it is found out that the crowded tournament selection operator is not implemented as originally proposed by Deb et al. (2002). Two identical pairs of solutions were able to compete in the subsequent tournaments, which stresses out the members with lower (better) rank thus leading to loosing the diversity in the mating pool, hence in the Pareto-optimal front. The last conclusion is, that it could be mentioned that there is a need for an improvement of the cNSGA-II in problems which have many local Pareto-optimal fronts, such as the benchmark problem ZDT4.

Bibliography

Corne, D., Knowles, J., and Oates, M. (2000). The pareto envelope-based selection algorithm for multiobjective optimization. pages 839–848. 6th International Conference on Parallel Problem Solving from Nature.

Deb, K. (1999). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7:205–230.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons, Ltd.

Deb, K. (2006). *Optimization for Engineering Design*. Prentice-Hall of India Pvt. Ltd.

Deb, K. and Agrawal, R. (1995). Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148.

Deb, K., Agrawal, S., Pratap, A., and Meyerivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197.

Deb, K. and Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multi-modal and multi-objective problems. *Complex Systems*, 9:431–454.

Deb, K., Mohan, M., and Mishra, S. (2003). Towards a quick computation of well-spread pareto-optimal solutions. volume 2632 of *Lecture Notes in Computer Science*, pages 222–236. Springer Berlin - Heidelberg.

ESTECO s.r.l. (2007). `http://www.esteco.com`. (version v4.0).

Fonseca, C. and Fleming, P. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. 5th International Conference on Genetic Algorithms, USA.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison Wesley Longman, Inc.

Horn, J., Nafpliotis, N., and Goldberg, D. (1994). A niched pareto genetic algorithm for multiobjective optimization. pages 82–87. First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence.

Knowles, J. and Corne, D. (1998). The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimization. pages 250–259. Parallel Problem Solving from Nature - PPSN V. 5th International Conference.

Parks, G. and Miller, I. (1998). Selective breeding in a multiobjective genetic algorithm. pages 250–259. Parallel Problem Solving from Nature - PPSN V. 5th International Conference.

Poles, S. (2003). MOGA-II, an improved multi-objective genetic algorithm. Technical Report 006, ESTECO s.r.l.

Poles, S., Geremia, P., Campos, F., Weston, S., and Islam, M. (2007). MOGA-II for an automotive cooling duct optimization on distributed resources. volume 4403 of *Lecture Notes in Computer Science*, pages 633–644. Springer Berlin - Heidelberg.

Poles, S., Rigoni, E., and Robič, T. (2004). MOGA-II performance on noisy optimization problems. Bioinspired Optimization Methods and their Applications (BIOMA).

Poloni, C., Giurgevich, A., Onesti, L., and Pediroda, V. (2000). Hybridization of a multiobjective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics. *Computer Methods in Applied Mechanics and Engineering*, 186:403–420.

Poloni, C. and Pediroda, V. (1997). *GA coupled with computationally expensive simulations: tools to improve efficiency*, chapter Genetic Algorithms and Evolution Strategies in Engineering and Computer Science, pages 267–288. John Wiley and Sons.

Rigoni, E. and Poles, S. (2005). NBI and MOGA-II, two complementary algorithms for multi-objective optimizations. In Branke, J., Deb, K., Miettinen, K., and Steuer, R. E., editors, *Practical Approaches to Multi-Objective Optimization*, number 04461 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Schaffer, J. D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 93–100, Hillsdale, NJ, USA. L. Erlbaum Associates Inc.

Sobol, I. (1979). On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis*, 16:790–793.

Srinivas, N. and Deb, K. (1994). Multi-objective optimization using non-dominated sorting in genetic algorithms. *Evolutionary Computation*, 2:221–248.

Yamamoto, K. and Inoue, O. (1995). New evolutionary direction operator for genetic algorithms. *AIAA Journal*, 33:1990–1993.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Emprical results. *Evolutionary Computation*, 8:173–195.

Zitzler, E., Laumanns, M., and Thiele, L. (2001). Spea2: Improving the strength pareto evolutionary algorithm. Technical Report 103, Swiss Federal Institute of Technology (ETH) Zurich.

Zitzler, E. and Thiele, L. (1998). An evolutionary algorithm for multiobjective optimization: the strength pareto approach. Technical Report 43, Swiss Federal Institute of Technology (ETH) Zurich.

Zitzler, E. and Thiele, L. (1999). Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, 3:257–271.

F TECHNICAL REPORT-II

C. C. Tutum and J. H. Hattel, *"A benchmark study for constrained multi-objective problems"*, October 2009.

Technical Report-2

A Benchmark Study for Constrained Multi-objective Optimization Problems

Cem Celal Tutum
Jesper Henri Hattel

Process Modelling Group
Department of Mechanical Engineering
Technical University of Denmark
Produktionstorvet, Kgs. Lyngby
2800, Denmark

October 1st, 2009

Table of Contents

1 Introduction

One of the goals of an ideal multi-objective optimization (MOO) procedure, in which the objectives are conflicting with each other and there is no higher level information (user preferences) beforehand that is leading to a biased search, is to find as many Pareto-optimal solutions as possible. Due to the population based search strategy of an evolutionary algorithm (EA), it is possible to obtain a set of Pareto-optimal solutions in a single simulation run. Besides aiming for a converged solution set as in a single objective optimization (SOO) case, it is also an important task for a multi-objective evolutionary algorithm (MOEA) to have well-distribution or diversity among these trade-off solutions, which is specific to MOO. Some of the problem features that may cause a multi-objective GA difficulty have been systematically investigated and procedures of constructing multi-objective test problems addressing different aspects of difficulties are presented before by Deb (1999b). The convergence to the global Pareto-front (among local ones) as well as satisfying the well-distribution of these non-dominated solutions in a complex (convex or non-convex as well as continuous or discontinuous) objective space can be highlighted as major criteria. Moreover, the presence of constraints, whether they are equality (hard) or inequality, in an MOO problem would make the problems even more severe. Thus the challange of obtaining the two major success criteria from an MOEA (i.e. the convergence and the diversity), can be accomplished only by using a proper constraint handling strategy.

This technical report presents several well-known constrained MOO benchmark problems in order to investigate some of the difficulties as mentioned above that a MOEA may have to tackle. Three elitist algorithms, i.e. the MOGA-II by Poloni and Pediroda (1997), the NSGA-II by Deb et al. (2002) (the versions implemented in modeFRONTIER) and the cNSGA-II (custom NSGA-II implementation by the author in MATLAB with two versions, see the details in Section 2.3), are employed for this purpose. Two types of constraint handling strategies, i.e. the Penalty Function Approach and the Constrained Tournament Method, had also been tested alongside. The MOGA-II uses the former approach, while the NSGA-II and the cNSGA-II use the latter approach. Besides a clear distinction of the different convergence and the diversity behavior of these three algorithms reported, it is also interesting to distinguish the convergence speeds.

2 A Brief Description of the Algorithms

In this section, a brief description of the three elitist algorithms, i.e. the MOGA-II by Poloni and Pediroda (1997), the NSGA-II by Deb et al. (2002) (the versions implemented in modeFRONTIER) and the cNSGA-II (the custom NSGA-II implementation by the author in MATLAB), are presented toether with two different constraint handling strategies, i.e. the penalty function approach in the MOGA-II and the constrained tournament selection in the NSGA-II and the cNSGA-II.

2.1 Multi-objective Genetic Algorithm (MOGA-II)

MOGA-II is one of the new elitist evolutionary algorithms designed for MOO problems that is implemented in the commercial MOO software modeFRONTIER (ESTECO s.r.l., 2007). It is an improved version of the MOGA by Poloni and Pediroda (1997), but it should not be confused with the non-elitist MOGA developed by Fonseca and Fleming (1993).

As in classical genetic algorithms (Goldberg, 1989), the internal encoding in MOGA-II is based on binary coding that each design variable is represented as a binary string, i.e. 0100, 1101101, etc., where the length of the string depends on the number of allowed values for the variable. MOGA-II starts with the initial binary-coded population P of size N and an empty elite set $E = \emptyset$. $P' = P \cup E$ is computed every generation. If the population size of P' is greater than the population size of P, the exceeding members in P' are randomly removed in order to reduce the population size to the initial size. Following this step, the evolution takes place in which P' is updated to P'' by applying all MOGA operators, i.e. classical cross-over, directional cross-over, mutation and selection.

The classical (single-point) cross-over is a well-known genetic algorithm operator in which two parents (individuals from the mating pool) are randomly chosen and some portions of these binary strings are exchanged between each other. In single-point cross-over, the crossing site is randomly chosen and all bits on the right side of the crossing site are swapped as shown schematically in Fig. 2.1. In MOGA-II, single point cross-over starts by taking the current individual-i as the first parent. The second parent (individual-j) is chosen by means of a multi-objective tournament selection on a randomly selected population subset (Poles et al., 2007, 2004; Rigoni and Poles, 2005). After the cross-over operation is performed, one of the resulting strings (children) is randomly selected to be the new individual.

The *directional cross-over* assumes that a *direction of improvement*, which reminds the *search direction* in classical algorithms that is relatively more expensive to compute, can be detected by comparing the fitness values of two reference individuals selected from the same generation similar to the idea presented by Yamamoto and Inoue (1995). The selection of these two reference individuals (parents) can be done using any available selection scheme. In MOGA-II, potential candidates are met in a random walk of assigned

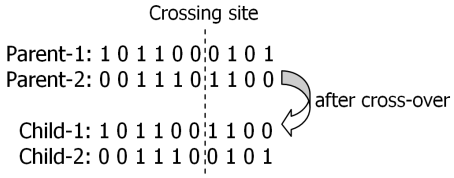


Figure 2.1: Schematic procedure of single-point crossover.

number of steps for each individual and then the two of them with the best fitnesses are chosen. The new individual is then created by moving in a randomly weighted direction that lies within the ones originated by the starting point (current individual) and two other parents selected by the local tournament, as shown in Fig. 2.2.

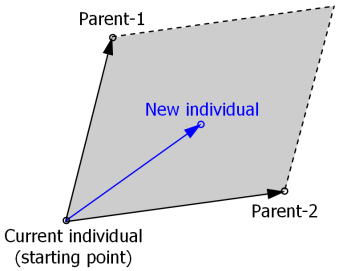


Figure 2.2: Schematic procedure of directional crossover in MOGA-II (Poles et al., 2004, 2007; Rigoni and Poles, 2005).

The *mutation* operator is responsible for perturbing the binary string, i.e. changing 1 to 0 or vice versa, with a given *DNA String Mutation Ratio* (probability). The mutation operator can be said to be responsible of the local search as well as maintaining the diversity in the population (Deb, 2006).

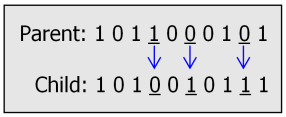


Figure 2.3: Schematic procedure of mutation for a given DNA string mutation ratio, p_m=30%.

The fitness of the population P'' is calculated. Non-domination sorting is performed and all non-dominated designs of P'' are copied to the elite set, E which is then updated by removing duplicated or dominated designs. Next, E is resized if it is larger than the population size N removing randomly the exceeding points. Finally, the first generation is created and P'' is assigned as the new initial population P for the next generation. This iterative loop continues until the stopping criterion is met, i.e. the user-defined total number of generations (Poles, 2003).

2.2 Non-dominated Sorting Genetic Algorithm (NSGA-II)

NSGA-II is a multi-objective evolutionary algorithm (MOEA) proposed by Deb et al. (2002), which enables finding well-spread multiple Pareto-optimal solutions for an MOO problem by incorporating three substantial features, i.e., elitism, non-dominated sorting, and a diversity preserving mechanism. Elitism, which is a genetic operator used to preserve the good solutions through the generations, is also shown to increase the convergence speed of an MOEA Zitzler et al. (2000).

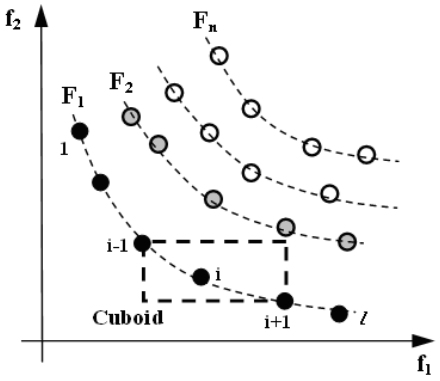


Figure 2.4: The crowding distance calculation (adapted from Deb et al. (2002)).

The preservation of well-distributed (diverse) nondominated solutions is maintained by the crowding distance operator. The crowding distance, which is calculated frontwise, is a measure of how close an individual is to its neighbors. In order to get an estimate of the density of solutions surrounding a particular solution in the population, the average distance between two points on either side of the solution along each of the objectives is calculated. This quantity d_i serves as an estimate of the perimeter of the cuboid formed by using the nearest neighbors as the vertices. A schematic view of the procedure for the computation of the crowding distance for an i^{th} individual is shown in Fig. 2.4, and formulated as (Deb et al., 2002),

$$d_{I_i^m} = d_{I_i^m} + \frac{f_m^{(I_{i+1}^m)} - f_m^{(I_{i-1}^m)}}{f_m^{max} - f_m^{min}}, \qquad for\ i = 2, ..., l-1 \tag{2.1}$$

where I(i) is the solution index of the i^{th} member in the (worse order) sorted list, m is the number of objectives and l is the number of the individuals in that particular front. The first and the last individuals are assigned an infinite distance in order to emphasize their fitness, hence preserving them for the next generation. A large average crowding distance will result in a better diversity in the population.

The schematic view of NSGA-II on a flowchart is shown in 2.5. The procedure starts with an initialization of the parent population, P_0, followed by sorting (ranking) the members of the population based on nondomination into each front, F_i. Each individual is assigned two fitness values, i.e. the one related to its nondomination level (front or rank, i.e. F_i refers to front with rank i in Fig. 2.4) and the crowding distance. Parents are selected by using binary (crowded) tournament selection based on these two measures.

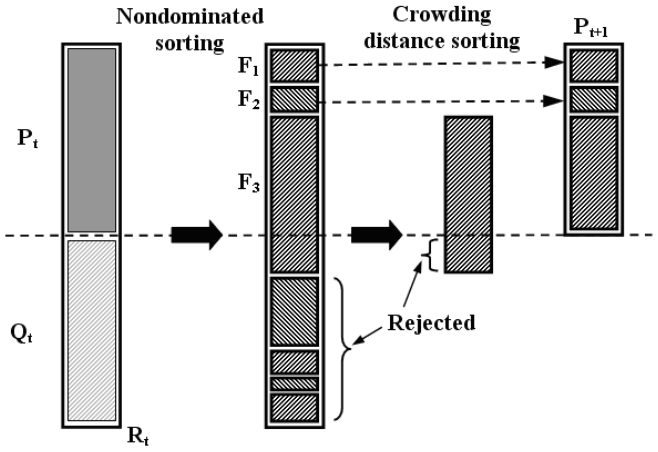


Figure 2.5: Schematic view of the NSGA-II procedure (Deb et al., 2002).

An individual is selected if its rank is less than the other, since the minimization is assumed, or if the crowding distance is greater than the other. The selected population generates off-springs, Q_t, by the application of usual genetic operators, i.e. crossover and mutation (simulated binary crossover (SBX), (Deb and Agrawal, 1995; Deb and Kumar, 1995) and polynomial mutation for real parameter GA in this case) (Deb, 2001; Goldberg, 1989). Since all the parent and the offspring individuals are included in the combined population, R_t, elitism is ensured. In order to reduce the size of the population of the next generation from $2N$ to N, nondomination sorting is applied and the individuals with relatively lowest possible ranks are accomodated until the size of the population exceeds N. Then the individuals in front F_i are selected based on their crowding distance. This cycle is repeated until the stopping criterion, i.e. the total number of generations, is met. For further details, please consult the original study (Deb et al., 2002).

2.3 Custom implementation of the NSGA-II (cNSGA-II)

This custom (personal) implementation of the NSGA-II (cNSGA-II) by the author should not be confused with the clustered NSGA-II (C-NSGA-II) (Deb et al., 2003) which uses an identical clustering technique to that used in the SPEA (Zitzler and Thiele, 1998, 1999) in order to obtain a well-spread distribution.

The cNSGA-II has a minor change compared to the original NSGA-II in the application of the SBX and polynomial mutation operators. Instead of applying these operators with predefined probabilities for each, a fixed percentage of the population (90%) has been randomly selected as the set of parents in order to produce the children with a predefined distribution index (η_c). Following this, polynomial mutation is applied on the rest of the population (10%), which didn't participate in the SBX operation. By doing this, it is ensured that all members will participate in these operations. The percentage of the population on which the SBX operator is applied is chosen higher compared to the one for the polynomial mutation (i.e. 90% vs. 10%) in order to enforce the search on a larger portion of the search region, in addition to the application of a relatively smaller distribution index (resulting in a wider spread with respect to the parents, aiming at a uniform distribution in the search space). The polynomial mutation operator is applied

with a relatively higher distribution index in order to make a finer search (producing the children closer to the parents).

The computation of the fitness values are vectorized instead of looping over each individual. This kind of minor high performance computing capabilities have been utilized. The performance of the code has been optimized using the MATLAB Profiler. Binary-coding of the parameters has also been implemented, but in this particular study, real-parameter GAs have mainly been considered due to their advantages in continuous search space.

Since the other two algorithms implemented in modeFRONTIER plot the global Pareto-optimal set (not only those obtained at the last generation) after an extra book-keeping, the cNSGA-II is also modified in order to compare the three algorithms on an equal basis. The non-dominated solutions obtained in each generation has been kept in an extra dynamic array. After all non-dominated solutions are enrolled, a non-dominated sorting is applied in order to determine the global Pareto-optimal set. Results of both of these implementations have been presented in this report.

2.4 Algorithm Settings

As mentioned above, the three MOEAs, i.e. the MOGA-II, the NSGA-II (implemented in modeFRONTIER) and the cNSGA-II (custom implementation of the NSGA-II by the author, i.e. with an without extra Pareto archieve for only post-processing issues), were used to investigate some of the difficulties that a MOO algorithm may confront. The settings for each of these three algorithms are given below. Both of the MOGA-II and the NSGA-II algorithms were started with the initial populations provided by the Sobol sequence (Sobol, 1979) which is a *quasi-random* distribution filling the design space in a uniform manner. The cNSGA-II algorithm started with a randomly (provided by MATLAB) distributed initial population. Some common features for the three algorithms, i.e. population size, total number of generations and elitism are given below.

Common settings for the three algorithms:

```
Population Size : 100

Number of Generations : 250

Elitism : Enabled
```

The directional cross-over operator, described as in Section 2.1, provides an efficient search, on the other hand, a higher probability of this operator decreases its robustness, i.e. in case of highly non-linear problems the optimizer might be stuck at local optimal solutions (local Pareto-fronts). The probability of the classical (single-point) cross-over is related to the probability of the directional cross-over in the way that the cumulative summation of the probabilities of directional and classical cross-overs, selection and mutation operators is equal to 1. This means that an increase in the probability of directional cross-over results in a decrease in the probability of classical cross-over. A higher probability of the classical cross-over operator provides a more robust search since it allows the algorithm to jump more easily into different regions of the search space, on the other

hand, it will decrease the convergence speed as expected.

The constraint handling strategy in the MOGA-II depends on the penalization of the objectives which is a very popular one (due to its easy and relatively effective implementation) among others, e.g. ignoring infeasible solutions, Jiménez-Verdegay-Goméz-Skarmeta's method, Ray-Tai-Seow's Method, for further details see Deb (2001). For a given multi-objective minimization problem (a maximization problem can also be converted into a minimization problem by the duality principle, i.e. the objective function is multiplied by -1) having inequality constraints (which are normalized to have the same order) of type $g_j(\mathbf{x^i}) \geq 0$ (for j=1,2, ..., n_{ineq}), first, all constraint violations (w_j) are computed for each individual x^i, then summed up to give the overall constraint violation $\Omega(\mathbf{x^i})$ as follows,

$$w_j(\mathbf{x^i}) = \begin{cases} |g_j(\mathbf{x^i})|, & \text{if } g_j(\mathbf{x^i}) < 0, \\ 0 & \text{otherwise.} \end{cases} \Rightarrow \Omega(\mathbf{x^i}) = \sum_{j=1}^{n_{ineq}} w_j(\mathbf{x^i}) \tag{2.2}$$

next, this overall constraint violation ($\Omega(\mathbf{x^i})$) is multiplied with a penalty parameter R_m (the subscript-m is the number of the objectives, hence R_m is a vector of penalty terms), and then the product is added to each of the objective function values.

$$F_m(\mathbf{x^i}) = f_m(\mathbf{x^i}) + R_m\Omega(\mathbf{x^i}) \tag{2.3}$$

Here $F_m(\mathbf{x^i})$ denotes the modified objective functions which take the constraints into account via a penalty parameter. The overall constraint violation will be zero for a feasible solution, which means that F_m will be equal to the original objective function f_m. Obviously, an infeasible solution will have a modified objective function greater than the original one ($F_m > f_m$) due to the addition of the penalty term. For a problem of the minimization of the objectives, an MOEA will first assign rank on each solution based on its modified objective, by this way, it will try to promote the solutions which have lower fitness (objective) values, in other words, since the feasible solutions will dominate the infeasible ones, an MOEA will enhance the probability of survival of the feasible solutions. Further details about the application of the penalty function approach are given in the literature, e.g. Deb (1999a) (for MOO problems) and Michalewicz (1992) (for SOO problems). However, this approach is not accurate enough, since ranking depends on the chosen penalty parameter and as a result, infeasible solutions which are close to the constraint boundary can occupy the same front with feasible solutions.

The MOGA-II algorithm settings (Poles, 2003):

```
Initial population : Sobol Sequence

Prob. of Directional Cross-over : 65%

Prob. of Classical Cross-over : 20%

Prob. of Selection : 5%

Prob. of Mutation : 10%

DNA String Mutation Ration : 10%
```

`Constraint Handling Algorithm : Penalty Function Method`

The NSGA-II uses the Constrained Tournament Method, developed by K. Deb, which incorporates a definition or condition of constrain-domination (modified version of the usual definition of domination in the presence of constraints) for the Binary Tournament Selection (BTS) operator, which does not need any penalty parameter. Each solution can be either feasible or infeasible in a constrained problem. The BTS operator needs to compare two solutions in order to select the better one, hence it needs deal with three conditions: *i*) both solutions can be feasible, *ii*) one of them can be feasible, and the other one can be infeasible, or *iii*) both solutions can be infeasible. It is obvious to decide for the second case (*ii*), a feasible solution is preferred. In the last case, the infeasible solution with the lower constraint violation is chosen. In the first case, which is not obvious as in other cases, the solutions will be compared with respect to their non-domination levels (ranks) and if they belong to the same front, the crowding distance sorting operator will resolve the tie by selecting the solution located in a less crowded region. This approach will classify the infeasible solutions more precisely compared to the penalty approach (almost each infeasible solution will belong to a different front unless they have an identical constraint violation), while having a similar front distribution in the feasible region.

The NSGA-II algorithm (suggested as default) settings:

`Initial population : Sobol Sequence`

`Prob. of Cross-over : 90%`

`Prob. of Mutation : 10%`

`Distribution Index for Real-coded Cross-over : 20`

`Distribution Index for Real-coded Mutation : 20`

`Constraint Handling Algorithm : Constrained Tournament Method`

The cNSGA-II algorithm settings:

`Initial population : Random Distribution`

`Probability Distribution for the SBX,` η_c `: 2`

`Probability Distribution for the polynomial mutation,` η_m `: 5`

`Constraint Handling Algorithm : Constrained Tournament Method`

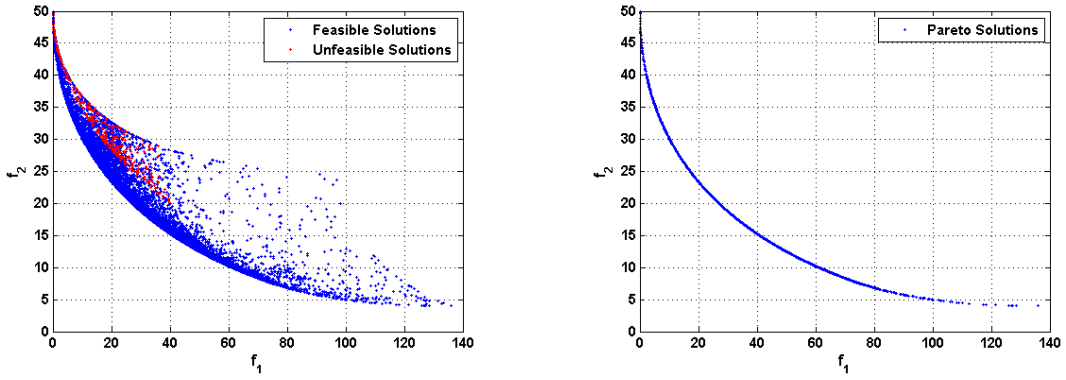
3 Test Functions and Results

In this section, five well-known constrained MOO benchmark problems, i.e. BNH, DEB, OSY, SRN and TNK, have been presented in an alphabetical order. In each problem, the minimization of two objective functions, having 2 to 6 design variables, has been considered. The convergence and the diversity performances of the three algorithms, the MOGA-II and the NSGA-II (implemented in the modeFRONTIER) as well as the cNSGA-II (the custom NSGA-II implementation by the author in MATLAB), have been compared.

3.1 BNH

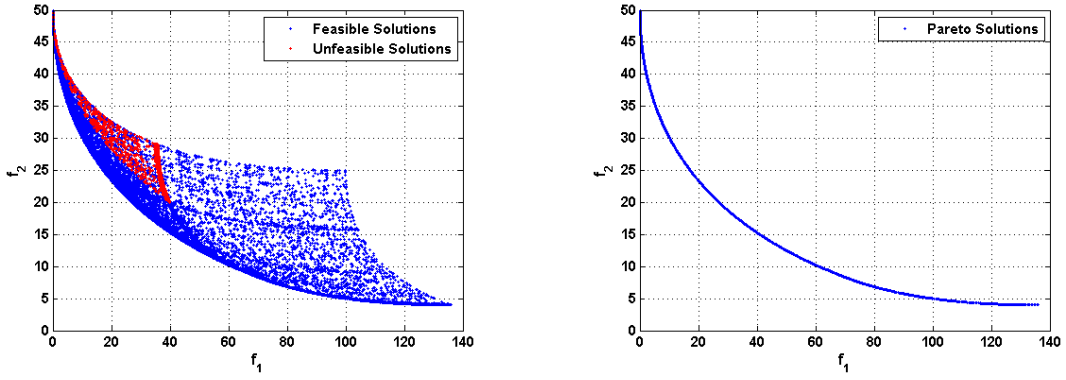
This benchmark case is a two-objective (f_1 and f_2 to be minimized), two-variable (x_1 and x_2) optimization problem with two nonlinear constraints presented by Binh and Korn (1997) for testing the performance of their Multi-Objective Evolution Strategy (MOBES) on nonlinear constraints. The results obtained by the three algorithms (the MOGA-II, the NSGA-II and the cNSGA-II) are presented below. First graph on a row presents all the solutions obtained through all generations. The feasible and unfeasible solutions are denoted by blue and red markers, respectively. In the second graph on the same corresponding raw, Pareto solutions are drawn. As mentioned before, the MOGA-II and the NSGA-II (which are implemented in the modeFRONTIER) post-process all the Pareto-optimal solutions obtained along generations. Three graphs of the cNSGA-II for each problem are shown following the results of the first two aforementioned algorithms. The last two of these three results in each problem present the Pareto-optimal solutions obtained by the cNSGA-II. The second graph of the Pareto-front is obtained with an identical post-processing step as in the MOGA-II and the NSGA-II. This presentation scheme is followed for each benchmark problem.

$$
\begin{aligned}
&Minimize: f_1(x) = 4x_1^2 + 4x_2^2, \\
&Minimize: f_2(x) = (x_1 - 5)^2 + (x_2 - 5)^2, \\
&subject\ to: g_1(x) = (x_1 - 5)^2 + x_2^2 \leq 25, \\
&\qquad\qquad g_2(x) = (x_1 - 8)^2 + (x_2 + 3)^2 \geq 7.7, \\
&\qquad\qquad 0 \leq x_1 \leq 5, \\
&\qquad\qquad 0 \leq x_2 \leq 3.
\end{aligned}
\tag{3.1}
$$



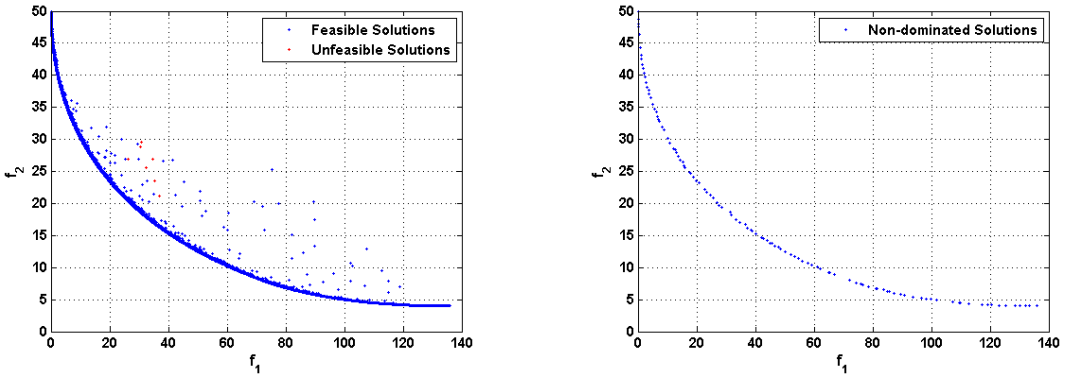
(a) (b)

Figure 3.1: a) All individuals obtained during the MOGA-II run for the test function BNH. b) The Pareto-set obtained at the end of the MOGA-II run for the test function BNH.



(a) (b)

Figure 3.2: a) All individuals obtained during the NSGA-II run for the test function BNH. b) The Pareto-set obtained at the end of NSGA-II run for the test function BNH.



(a) (b)

Figure 3.3: a) All individuals obtained during the cNSGA-II run for the test function BNH. b) The Pareto-set obtained at the end of the cNSGA-II run for the test function BNH.

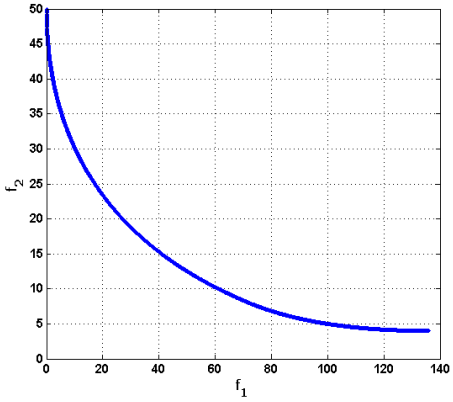
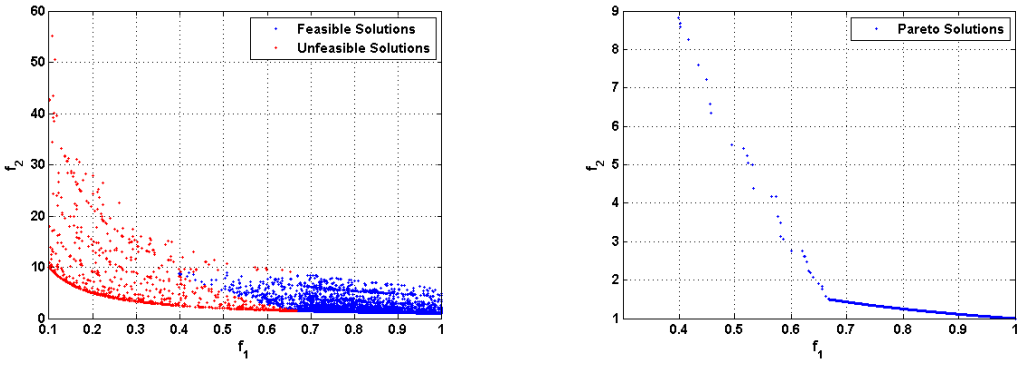


Figure 3.4: The global Pareto set obtained using the cNSGA-II with Pareto archieve strategy for the test function BNH.

3.2 DEB (CONSTR)

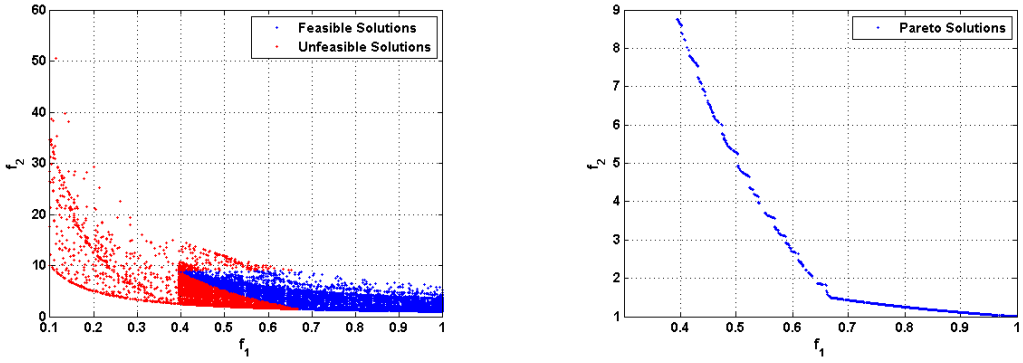
The test function DEB (CONSTR) is used in Deb et al. (2002) and discussed thoroughly in Deb (2001). The first objective function f_1 depends only on x_1, the second objective function has two variables.

$$\begin{aligned} Minimize : f_1(x) &= x_1, \\ Minimize : f_2(x) &= \frac{1+x_2}{x_1}, \\ subject\ to : g_1(x) &= x_2 + 9x_1 \geq 6, \\ g_2(x) &= -x_2 + 9x_1 \geq 1, \\ 0.1 \leq x_1 &\leq 1, \\ 0 \leq x_2 &\leq 5. \end{aligned} \tag{3.2}$$



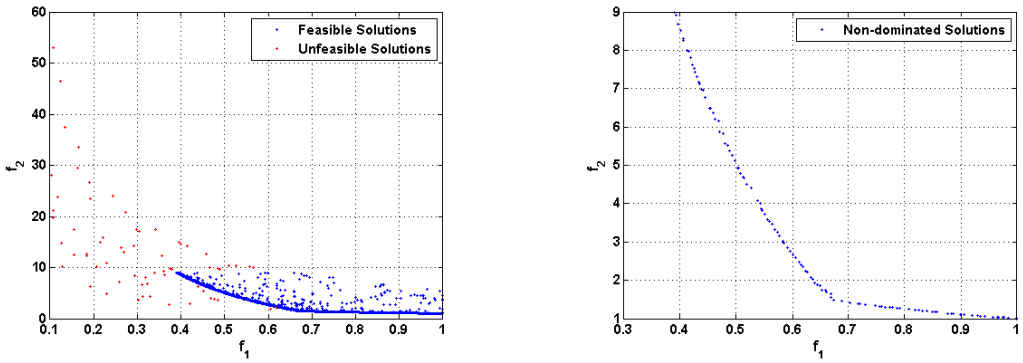
(a) (b)

Figure 3.5: a) All individuals obtained during the MOGA-II run for the test function DEB. b) The Pareto-set obtained at the end of the MOGA-II run for the test function DEB.



(a) (b)

Figure 3.6: a) All individuals obtained during the NSGA-II run for the test function DEB. b) The Pareto-set obtained at the end of the NSGA-II run for the test function DEB.



(a) (b)

Figure 3.7: a) All individuals obtained during the cNSGA-II run for the test function DEB. b) The Pareto-set obtained at the end of the cNSGA-II run for the test function DEB.

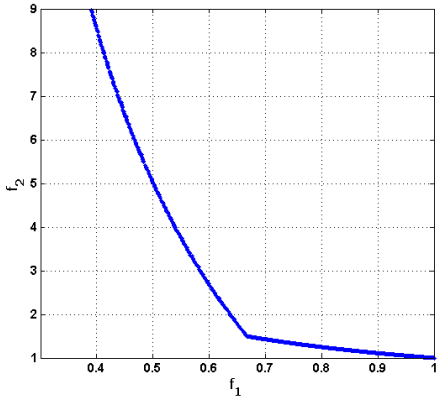
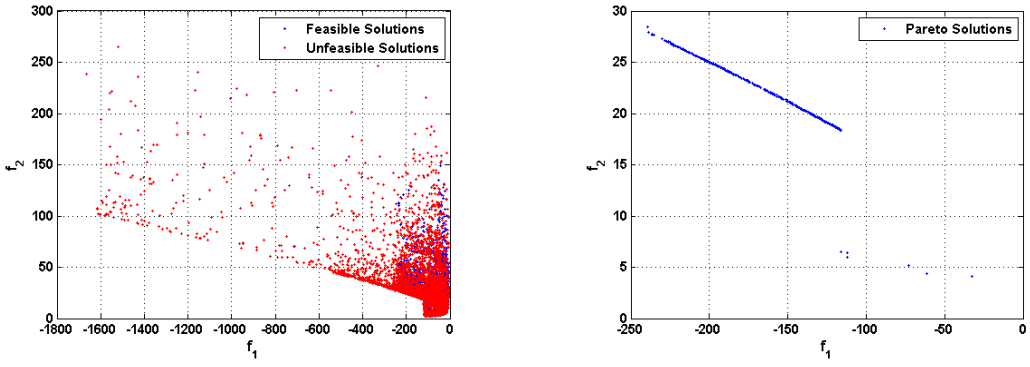


Figure 3.8: The global Pareto set obtained using the cNSGA-II with the Pareto archieve strategy for the test function DEB.

3.3 OSY

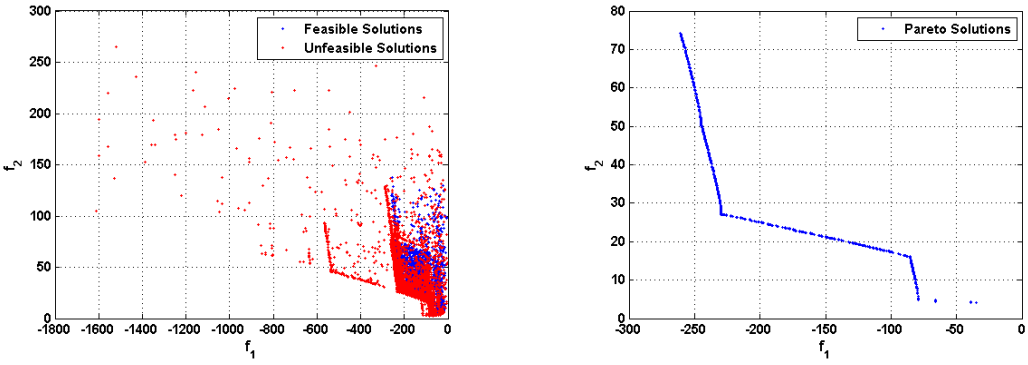
This benchmark problem, which is used by Osyczka and Kundu (1995), contains two nonlinear objective functions and six constraints (four of them are linear). This test case is the most difficult test case among five benchmark problems, since the Pareto-optimal front is composed of five regions and each region demands an MOEA to maintain its subpopulations at different intersections of constraint boundaries (Deb, 2001).

$$
\begin{aligned}
&Minimize: f_1(x) = -[25(x_1-2)^2 + (x_2-2)^2 + (x_3-1)^2 + (x_4-4)^2 + (x_5-1)^2],\\
&Minimize: f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2,\\
&subject\ to: g_1(x) = x_1 + x_2 - 2 \geq 0,\\
&\qquad g_2(x) = 6 - x_1 - x_2 \geq 0,\\
&\qquad g_3(x) = 2 - x_2 + x_1 \geq 0,\\
&\qquad g_4(x) = 2 - x_1 + 3x_2 \geq 1,\\
&\qquad g_5(x) = 4 - (x_3-3)^2 - x_4 \geq 0,\\
&\qquad g_6(x) = (x_5-3)^2 + x_6 - 4 \geq 0,\\
&\qquad 0 \leq x_1, x_2, x_6 \leq 10,\\
&\qquad 1 \leq x_3, x_5 \leq 5,\\
&\qquad 0 \leq x_4 \leq 6.
\end{aligned}
\tag{3.3}
$$



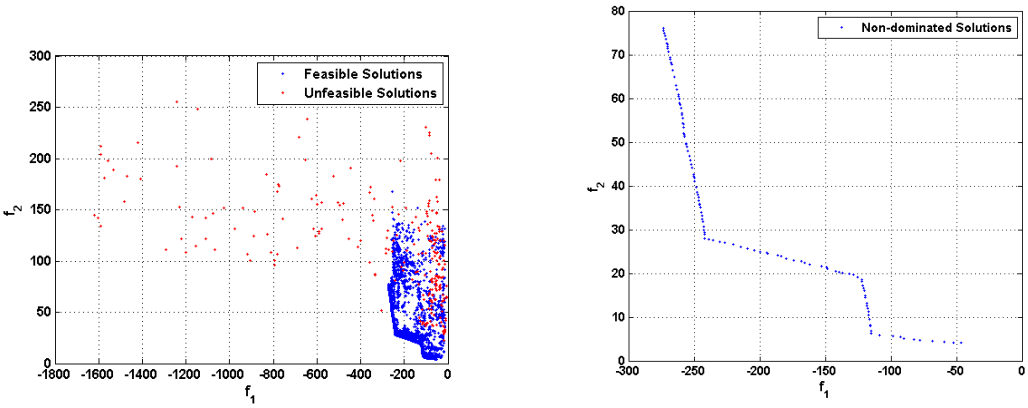
(a) (b)

Figure 3.9: a) All individuals obtained during the MOGA-II run for the test function OSY. b) The Pareto-set obtained at the end of the MOGA-II run for the test function OSY.



(a) (b)

Figure 3.10: a) All individuals obtained during the NSGA-II run for the test function OSY. b) The Pareto-set obtained at the end of the NSGA-II run for the test function OSY.



(a) (b)

Figure 3.11: a) All individuals obtained during the cNSGA-II run for the test function OSY. b) The Pareto-set obtained at the end of the cNSGA-II run for the test function OSY.

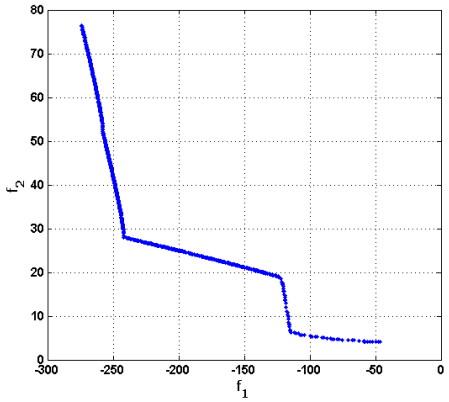
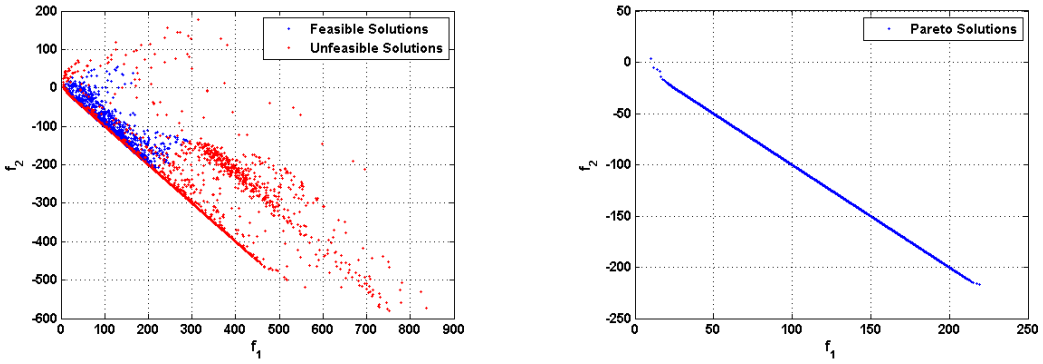


Figure 3.12: The global Pareto set obtained using the cNSGA-II with the Pareto archieve strategy for the test function OSY.

3.4 SRN

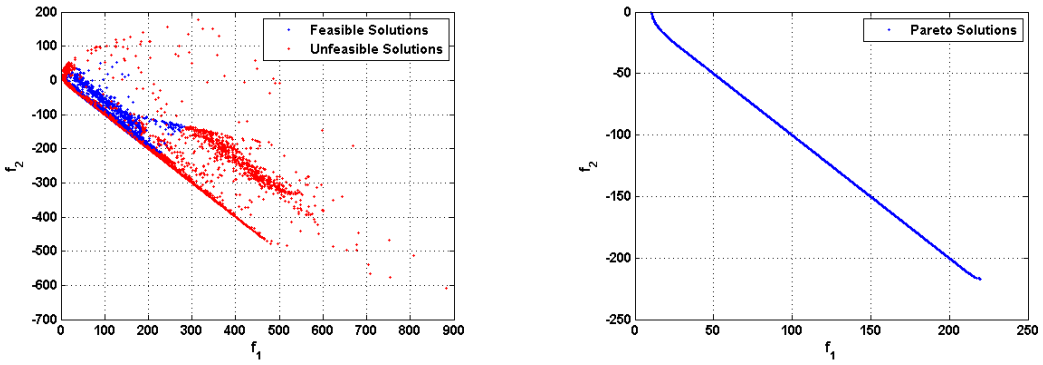
The test problem presented here is used by Srinivas and Deb (1994).

$$\begin{aligned} Minimize: f_1(x) &= 2 + (x_1 - 2)^2 + (x_2 - 1)^2, \\ Minimize: f_2(x) &= 9x_1 - (x_2 - 1)^2, \\ subject\ to: g_1(x) &= x_1^2 + x_2^2 \leq 225, \\ g_2(x) &= x_1 - 3x_2 + 10 \leq 0, \\ -20 &\leq x_1 \leq 20, \\ -20 &\leq x_2 \leq 20. \end{aligned} \tag{3.4}$$



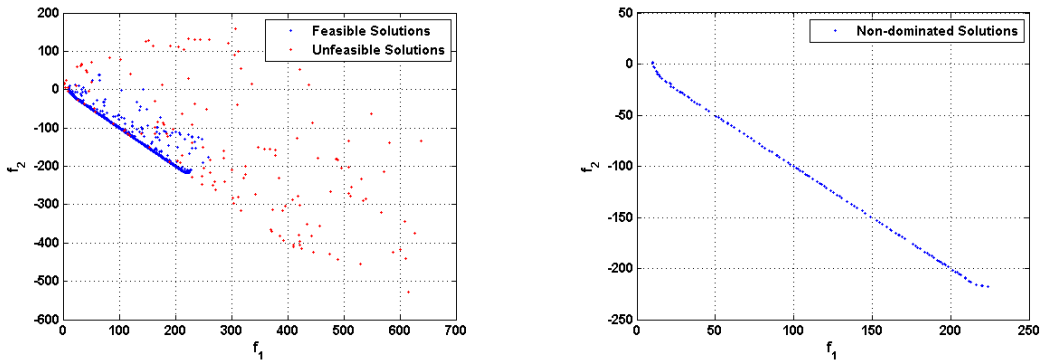
(a) (b)

Figure 3.13: a) All individuals obtained during the MOGA-II run for the test function SRN. (b)The Pareto-set obtained at the end of the MOGA-II run for the test function SRN.



(a) (b)

Figure 3.14: a) All individuals obtained during the NSGA-II run for the test function SRN. b) The Pareto-set obtained at the end of the NSGA-II run for the test function SRN.



(a) (b)

Figure 3.15: a) All individuals obtained during the cNSGA-II run for the test function SRN. b) The Pareto-set obtained at the end of the cNSGA-II run for the test function SRN.

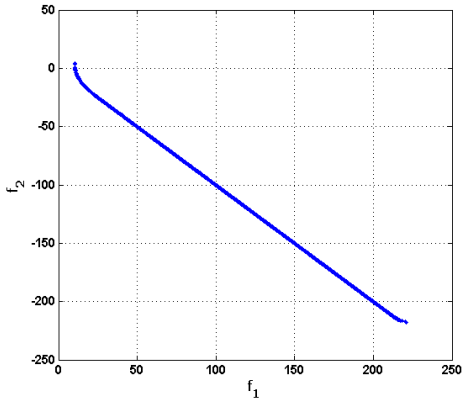
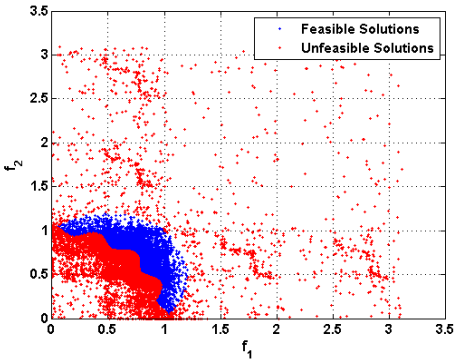


Figure 3.16: The global Pareto set obtained using the cNSGA-II with the Pareto archieve strategy for the test function SRN.

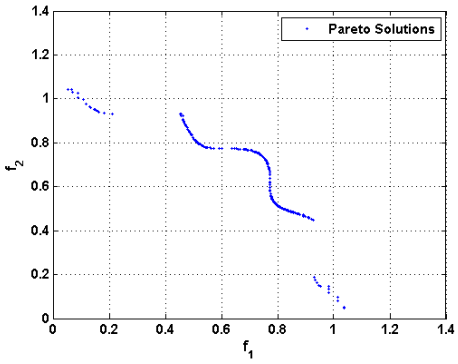
3.5 TNK

Tanaka (1995) used the following two-variable problem of which the solution, i.e. the Pareto-optimal front, is discontinuous.

$$
\begin{aligned}
&Minimize: f_1(x) = x_1, \\
&Minimize: f_2(x) = x_2, \\
&subject\ to: g_1(x) = x_1^2 + x_2^2 - 1 - 0.1\cos\left(16\arctan\left(\frac{x_1}{x_2}\right)\right) \geq 0, \\
&\qquad\quad g_2(x) = (x_1 - 0.5)^2 + (x_2 - 0.5)^2 \leq 0.5, \\
&\qquad\quad 0 \leq x_1 \leq \pi, \\
&\qquad\quad 0 \leq x_2 \leq \pi.
\end{aligned}
\tag{3.5}
$$

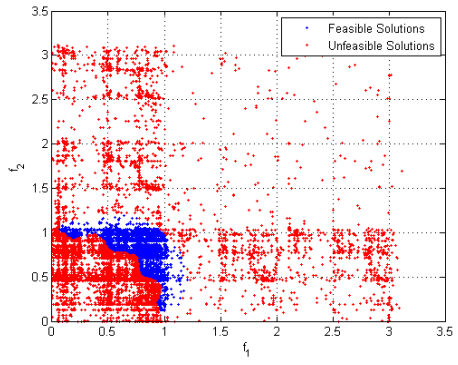


(a)

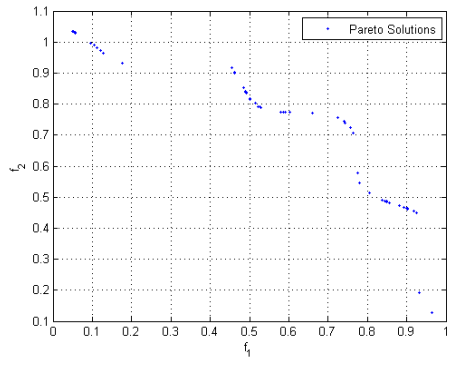


(b)

Figure 3.17: a) All individuals obtained during the MOGA-II run for the test function TNK. b) The Pareto-set obtained at the end of the MOGA-II run for the test function TNK.

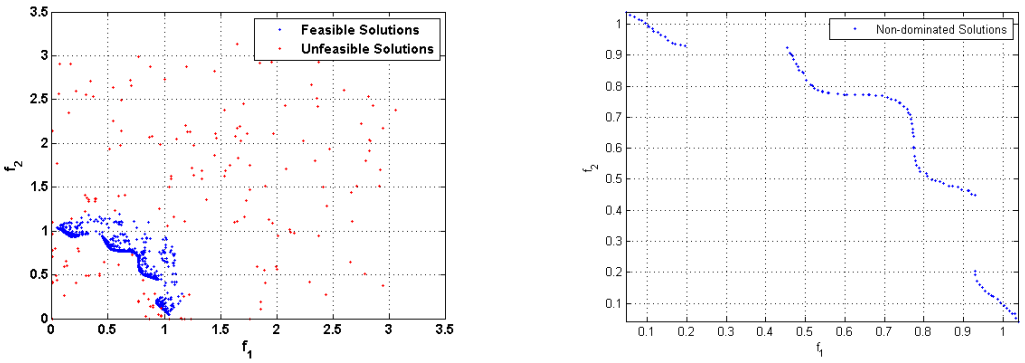


(a)



(b)

Figure 3.18: a) All individuals obtained during the NSGA-II run for the test function TNK. b) The Pareto-set obtained at the end of the NSGA-II run for the test function TNK.



(a) (b)

Figure 3.19: a) All individuals obtained during the cNSGA-II run for the test function TNK. b) The Pareto-set obtained at the end of the cNSGA-II run for the test function TNK.

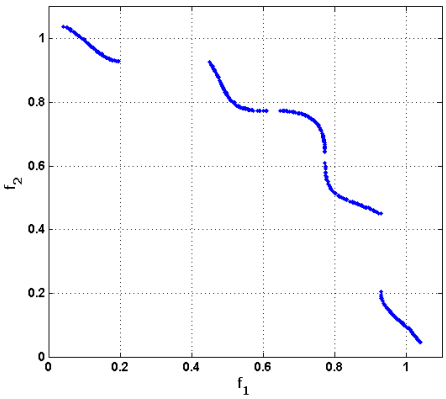


Figure 3.20: The global Pareto set obtained using the cNSGA-II with the Pareto archieve strategy for the test function TNK.

4 Discussion of the Results and Conclusions

In this technical report, five well-known constrained MOO benchmark problems, i.e. BNH, DEB (CONSTR), OSY, SRN and TNK, have been presented. Some of the difficulties that an MOEA may have to tackle have been investigated by using three elitist algorithms, the MOGA-II and the NSGA-II implemented in the modeFRONTIER and the cNSGA-II implemented by the author in MATLAB. Two types of constraint handling strategies, i.e. the Penalty Function Approach and the Constrained Tournament Method, have also been tested alongside. The MOGA-II uses the former approach, while the NSGA-II and the cNSGA-II use the latter approach. All of the test problems, which are sorted in an alphabetical order (i.e. not according to difficulty level), are composed of two minimization of objectives and a number of design variables varying from 2 to 6. Moreover, both linear and non-linear inequality constraints are used varying between 2 and 6, excluding the bounds on the variables.

In the first test function BNH, the solutions of all the three algorithms were able to converge to the true Pareto-optimal front without any problem and the diversity among the non-dominated solutions was also succesfully obtained by each algorithm. It is noteworthy to distinguish the convergence rate of the algorithms. The MOGA-II has spent less time (i.e. less number of computations) near the Pareto front compared to the NSGA-II, but on the other hand the NSGA-II has fully covered the true Pareto-front while the density of the solutions obtained by the MOGA-II has been decreasing towards the right boundary of the front. The cNSGA-II clearly outperformed these two algorithms by converging much faster and spending less computations near the Pareto front. In the second test problem DEB (CONSTR), although the NSGA-II was more succesful in finding a larger amount of non-dominated solutions compared to the MOGA-II, the number of unfeasible solutions was also much higher which is not preferable. On the other hand, the MOGA-II was not able to cover the left arm of the Pareto-front. The performance of the cNSGA-II was again satisfying compared to the others in the sense that it both converged faster and found a larger amount of non-dominated solutions which were distributed accross the Pareto-front in a very uniform way. The third test problem OSY is the most difficult one of all five benchmark cases, since the Pareto-optimal front consists of five regions and each region demands an MOEA to maintain its subpopulations at different intersections of constraint boundaries. As seen from the results, the MOGA-II was clearly unsuccessful in this problem since it could only find the middle (the third) distinct part of the Pareto-front. The NSGA-II was able to capture almost all five regions of the Pareto-front, but the density in the lower region is quite low. The cNSGA-II performed well considering both having converged and well-spread distributed non-dominated solutions, although it spent more computations outside the Pareto-front compared to the previous test cases. In the next test problem SRN, the convergence behavior of the two algorithms, the MOGA-II and the NSGA-II, was similar. A nice (uniform and dence) distribution of the non-dominated solutions accross the Pareto-front is seen clearly. The convergence rate of the cNSGA-II was again similar to the previous cases, i.e. fast convergence with less computations outside the Pareto-front. In the last benchmark case TNK, which has a discontinuous Pareto-optimal front, the performance of the cNSGA-II was dominating

the other two algorithms.

These results show the performance of both different MOEAs and the two different constraint handling strategies, i.e. the penalty function approach and the constrained tournament method. Although the NSGA-II and the cNSGA-II uses the same constraint handling strategies (constrained tournament method), they show very different convergence characteristics. The NSGA-II is claimed to be implemented as in the original study by Deb (2001), but at least for the constrained problems presented here, it is fair to say that it is not as succesfull as the original one. On the other hand, the success of the cNSGA-II could be explained by a slightly different strategy in applying the SBX operator, i.e. ensuring all members to be updated with a chosen low probability distribution number (η_c=2), hence obtaining a large diversity in the begining of the evolution, then towards the end, narrowing the search which helps convergence. Moreover, the constrained tournament operator (proposed by K. Deb), which incorporates the constrain-domination definition, helps the cNSGA-II to converge to the true Pareto-front in a substantially quick way, besides supplying a very uniform distribution. Finally, it can be concluded that the performance of the cNSGA-II is satisfying especially regarding the convergence rate for the constrained problems presented here (with up to 6 design variables). It is planned to apply it on real-world (e.g. engineering) applications. An evolutionary hybrid algorithm, which uses the cNSGA-II as a robust search technique together with a local search algorithm, e.g. the Sequential Quadratic Programming (SQP), which improves the accuracy and the number of function evaluations, is under preparation.

Bibliography

Binh, T. and U. Korn (1997). Mobes: A multi-objective evolution strategy for constrained optimization problems. pp. 176–182. The Third International Conference on Genetic Algorithms (Mendel 97).

Deb, K. (1999a). *Evolutionary algorithms for multi-criterion optimization in engineering design*, pp. 135–161. Evolutionary Algorithms in Engineering and Computer Science. Chichester, UK: John Wiley and Sons.

Deb, K. (1999b). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation 7*, 205–230.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons, Ltd.

Deb, K. (2006). *Optimization for Engineering Design*. Prentice-Hall of India Pvt. Ltd.

Deb, K. and R. Agrawal (1995). Simulated binary crossover for continuous search space. *Complex Systems 9*, 115–148.

Deb, K., S. Agrawal, A. Pratap, and T. Meyerivan (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation 6*, 182–197.

Deb, K. and A. Kumar (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multi-modal and multi-objective problems. *Complex Systems 9*, 431–454.

Deb, K., M. Mohan, and S. Mishra (2003). Towards a quick computation of well-spread pareto-optimal solutions. Volume 2632 of *Lecture Notes in Computer Science*, pp. 222–236. Springer Berlin - Heidelberg.

ESTECO s.r.l. (2007). `http://www.esteco.com`. (version v4.0).

Fonseca, C. and P. Fleming (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. 5th International Conference on Genetic Algorithms, USA.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison Wesley Longman, Inc.

Michalewicz, Z. (1992). *Genetic Algorithms+Data Structures=Evolution Programs*. Springer-Verlag.

Osyczka, A. and S. Kundu (1995). A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm. *Structural Optimization 10*, 94–99.

Poles, S. (2003). MOGA-II, an improved multi-objective genetic algorithm. Technical Report 006, ESTECO s.r.l.

Poles, S., P. Geremia, F. Campos, S. Weston, and M. Islam (2007). MOGA-II for an automotive cooling duct optimization on distributed resources. Volume 4403 of *Lecture Notes in Computer Science*, pp. 633–644. Springer Berlin - Heidelberg.

Poles, S., E. Rigoni, and T. Robič (2004). MOGA-II performance on noisy optimization problems. Bioinspired Optimization Methods and their Applications (BIOMA).

Poloni, C. and V. Pediroda (1997). *GA coupled with computationally expensive simulations: tools to improve efficiency*, Chapter Genetic Algorithms and Evolution Strategies in Engineering and Computer Science, pp. 267–288. John Wiley and Sons.

Rigoni, E. and S. Poles (2005). NBI and MOGA-II, two complementary algorithms for multi-objective optimizations. In J. Branke, K. Deb, K. Miettinen, and R. E. Steuer (Eds.), *Practical Approaches to Multi-Objective Optimization*, Number 04461 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Sobol, I. (1979). On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis 16*, 790–793.

Srinivas, N. and K. Deb (1994). Multi-objective optimization using non-dominated sorting in genetic algorithms. *Evolutionary Computation 2*, 221–248.

Tanaka, M. (1995). Ga-based decision support system for multi-criteria optimization. pp. 1556–1561. The International Conference on on Systems, Man and Cybernetics.

Yamamoto, K. and O. Inoue (1995). New evolutionary direction operator for genetic algorithms. *AIAA Journal 33*, 1990–1993.

Zitzler, E., K. Deb, and L. Thiele (2000). Comparison of multiobjective evolutionary algorithms: Emprical results. *Evolutionary Computation 8*, 173–195.

Zitzler, E. and L. Thiele (1998). An evolutionary algorithm for multiobjective optimization: the strength pareto approach. Technical Report 43, Swiss Federal Institute of Technology (ETH) Zurich.

Zitzler, E. and L. Thiele (1999). Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation 3*, 257–271.