
This version is available at https://strathprints.strath.ac.uk/52499/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk
A new distributed cooling method for mitigating residual stress in Friction Stir Welding

Xingguo Zhou, Donald Mackenzie*, Wenke Pan

Abstract

An analytical and Finite Element investigation of the effect of different cryogenic cooling nozzle configurations on temperature and residual stress in a model friction stir weld is presented. A new configuration adopting a distributed cooling approach is proposed based on an analytical cooling model. Finite element models are implemented to verify the effect of distributed cooling on welding temperature and longitudinal residual stress. The results presented indicate that new active cooling methods can improve mitigation of welding induced residual stress.

Keywords: distributed cooling, active cooling, residual stress, friction stir welding, FEA

1 Introduction

Friction Stir Welding, FSW, invented by TWI in 1990 [1, 5], is a solid state welding process in which heat from a rotating tool softens material along a join line in the workpiece, enabling mechanical mixing of the material along the weld. The process is known for its capability to join metals which are difficult to weld using conventional fusion welding techniques, such as high strength aluminium alloys [6, 10]. Figure 1 shows a schematic of the FSW butt welding process. Compared to conventional fusion welding, FSW usually results in lower residual distortion due to the relatively low thermal input to the workpiece. However, the process induces residual stress that could influence the mechanical performance [11, 13] and the mechanical-chemical performance [14, 15] of the welds.
To investigate the factors that influence residual stress, researchers have investigated various materials and process parameters using numerical methods such as the Finite Element Method. These studies have mainly focused on residual stress caused by unbalanced heat progress during welding. Using a simplified 2D model incorporating the pin effect to material flow patterns, Zhang et al. found that longitudinal residual stress of the weld was symmetric to the welding line, with an *M-shaped* transverse profile. In addition, the maximum longitudinal residual stress was found to vary with welding velocity.

Khandkar et al. investigated thermal residual stress in sequentially coupled 3D simulation, considering aluminium alloys and steel. The properties of all three materials were treated as temperature dependent. Predicted results showed good agreement with data reported by other researchers and indicated that longitudinal tensile stress was the main component of residual stress in all three cases. More complex and realistic material models, heat source models and boundary conditions have been studied by other workers. These investigations indicated that thermal-induced stress contributed the majority to the residual stress with the largest component in the longitudinal direction.

Use of active cooling to control residual stress in both conventional fusion welding and friction stir welding has been widely reported. The effect of a cooling source was
investigated through finite element modelling and experimental validation by van der Aa [27] for aluminium alloy and steel, though there is a risk of martensite formation and consequent cracking for carbon steels due to the rapid cooling. Results indicated that cooling sources trailing the welding torque can reduce residual stress, as well as buckling distortion induced by fusion welding. Further, it was observed that increasing cooling power above 6000 Wm\(^{-2}\)K did not have a significant effect on further reducing residual stress. Staron et al. [28] applied liquid CO\(_2\) cooling media to friction stir welded Al sheets by experiment and found tensile residual stresses in the weld zone was reduced significantly and it was even possible to generate compressive residual stress, which is considered to be able to improve fatigue life for components. Richards et al. [29] modelled the progress of active cooling applied to friction stir welding using liquid CO\(_2\), and also identified that there is a cooling power limit, above which little reduction in residual stress occurs. They also investigated a two nozzle strategy aimed at further reduction in longitudinal residual stress but without success. This was attributed to the near-seam areas getting colder and yield strength become higher, making it difficult to generate more tensile plastic strain.

This paper presents an analytical and numerical investigation of the effect of cooling nozzle configuration and a new distributed cooling method proposed. As the welding temperature plays an important role in friction stir welding [30], the effect of high cooling power on welding temperature is investigated when applied close to the heating source. The resulting longitudinal residual stress was studied in particular, as this has been shown to be a significant stress component [31,32].

2 Nozzle configurations and analysis
According to investigations in [29], active cooling can generate tensile plastic strain to compensate compressive plastic strain incurred during welding. It is therefore only effective when the nozzle is placed trailing the welding tool. A common configuration for active cooling is shown in Figure 2a. The left circle indicates that the cooling nozzle used is a cylinder perpendicularly placed to workpiece top surface. The cooling nozzle’s diameter $D_c$ is usually not smaller than that of the tool diameter $D_t$. The cooling rate within the nozzle is usually considered to be uniform. For convenience of investigation, the cooling nozzle in this paper is rectangular, as shown in Figure 2b (referred as non-distributed cooling), and its width $W_{ca}$ and length $L_{ca}$ are 10mm and 20mm respectively.

Another nozzle configuration (referred as distributed cooling) is presented in Figure 3. The cooling nozzle is divided evenly into 4 zones, as shown in Figure 3a, each with adjustable distance centre to centre $D_{cc}^i$ (for $i=1$ and 2; $i=3$ or 4 is not shown due to symmetry). Theoretically, the cooling rate $q_i$ ($i=1$, 2) of the zones can be different, as shown in Figure 3b, but physically precise control of its distribution might be difficult in practice and this approach is not considered here.
2.1 Welding temperature

Provided the cooling nozzle is close enough to the heat source, applying active cooling will decrease welding temperature of material under the welding tool. To simplify the analysis presented in this section, 2D planar heat transfer is assumed and the material properties related to heat transfer calculations are temperature independent.

The cooling area is equally discretized into \( n \) points (or lines through the thickness), as shown in Figure 4. The temperature increment of point \( P \) during the time increment from 0 to \( t \) due to one of the \( n \) cooling points, point \( j \), can be expressed as Equ.1 [33], which is derived from Laplace's equation:

\[
T = \frac{Q}{4\pi \alpha c p h} \exp \left( \frac{-r^2}{4\alpha t} \right)
\]

where: \( T \) represents temperature after time \( t \); \( Q \) is the instantaneous energy input (negative); \( h \) is the thickness of plate; \( \alpha, c \) and \( p \) are constant material property parameters thermal diffusivity; specific heat and density; \( r \) is the distance between point \( P \) and point \( j \); and \( t \) is the time duration since the line heating source applied.
Based on temperature cumulative law, the temperature induced by a non-distributed cooling source and a distributed cooling source can be expressed as Equ.2 and Equ.3 respectively,

\[
T_{\text{cool}}^{nc} = \sum_{j=1}^{n} \int_{0}^{t_{0}} \frac{qdt}{4\pi cp\alpha h(t_{0} - t)} \exp\left(\frac{-(x_{j}^{nc} - vt)^{2} + y_{j}^{nc2}}{4\alpha(t_{0} - t)}\right)
\]

Equ.2

\[
T_{\text{cool}}^{dc} = \sum_{j=1}^{n} \int_{0}^{t_{0}} \frac{qdt}{4\pi cp\alpha h(t_{0} - t)} \exp\left(\frac{-(x_{j}^{dc} - vt)^{2} + y_{j}^{dc2}}{4\alpha(t_{0} - t)}\right)
\]

Equ.3

where: \(n\) is the total amount of discrete points; \(T_{\text{cool}}^{nc}\) and \(T_{\text{cool}}^{dc}\) are temperature increment of point P from time 0 to \(t_{0}\) in the non-distributed and distributed cooling models respectively; \(q\) is the cooling rate of point j in the non-distributed and distributed cooling models and \(x_{j}^{nc}\), \(y_{j}^{nc}\) and \(x_{j}^{dc}\), \(y_{j}^{dc}\) are the X and Y co-ordinate distances between point j and point P in the non-distributed cooling model and distributed cooling model; \(v\) is welding velocity. When \(t=t_{0}\), \(x_{j}^{nc} - vt > 0\).

\[\begin{align*}
x_{j}^{dc} > x_{j}^{nc} & \quad \text{and} \quad y_{j}^{dc} = y_{j}^{nc} \quad \text{when point C belongs to zone 1} \\
x_{j}^{dc} = x_{j}^{nc} & \quad \text{and} \quad y_{j}^{dc} = y_{j}^{nc} \quad \text{when point C belongs to zone 2}
\end{align*}\]

Thus \(T_{\text{cool}}^{dc} > T_{\text{cool}}^{nc}\), indicating that the welding temperature in the distributed cooling model is higher than that of non-distributed cooling.

2.2 Temperature gradient

The temperature gradients between point A and B, as shown in Figure 5 for distributed cooling and non-distributed cooling are expressed by Equ.4 and Equ.5 respectively (material properties are assumed constant).

![Figure 5 Schematics of temperature gradient analysis](image-url)
where: $\Delta T_{cool}^{dc}$ is the temperature difference between point A and point B in the distributed cooling; and $\Delta T_{cool}^{nc}$ is the temperature difference between point A and point B in the non-distributed cooling.

For point j in zone 2,

$$\Delta T_{cool}^{dc} = \Delta T_{cool}^{nc}$$

For point j in zone 1,

$$\Delta T_{cool}^{dc} < \Delta T_{cool}^{nc}$$

Therefore, the temperature gradient for material under the welding tool could be reduced by distributed cooling. The lower temperature gradient indicates lower local temperature-induced tensile stress after the tool has passed.

### 2.3 Temperature Dependant Material Properties

The mechanical properties of metals like aluminium alloys and steels are temperature dependent and they become softer at higher temperature. Figure 6 shows a temperature-yield strength curve for aluminium alloy AA2024. The temperature dependence of the strength influences the effect of active cooling. To generate tensile plastic strain, it is better to place the active cooling power as close as possible to the welding tool, where material is hot and temperature gradient is high. However, as the temperature drops quickly between the tool and cooling source, the material quickly becomes more difficult to yield.
Though the longitudinal temperature gradient in distributed cooling will be smaller than that of non-distributed cooling the local temperature itself is higher, giving a possibility of lower longitudinal residual stress.

3 FEA modelling

The analyses presented in Section 2.1 and Section 2.2 are simplified and provide a qualitative description of the effect of distributed cooling. To achieve a quantitative assessment of the effect of the distributed cooling arrangement, ABAQUS Standard based thermal-mechanical coupled Finite Element models were used. The FE models include the thermal effect of the welding tool but do not include mechanical effects such as material stirring and local forces. Three cooling arrangements are considered: no active cooling, non-distributed cooling and distributed cooling.

In the ABAQUS based models, the main heat source is the friction between the welding tool and the material, which is modelled as a heat flux in the FE model. It moves synchronously with the welding tool and is applied as a user defined load implemented by user subroutine DFLUX \(^34\). DFLUX is programed in FORTRAN and it defines the heat flux.
distribution and variation with time. It is called at each step increment and heat flux inputs for elements are updated.

The passed-in parameters in DFLUX include step time, total time, coordinates, flux type, surface name, etc. The step time and total time can be used to calculate the reference position based on welding velocity. The coordinates are current coordinates of current point and are represented by a vector. In case that multiple types of heat flux are involved, the heat flux type is identified as either body flux or surface flux. The surface name helps to identify different surface heat fluxes in different surfaces if necessary.

The basic implementation of DFLUX to define the moving heat source in the present investigation is shown in Figure 7. When the ABAQUS main program needs the heat flux value at an integration point, it calls DFLUX. First it calculates the current position of heat source centre based on the current step time and constant welding velocity. Then the distance between the integration point and the heat source centre is judged if within the radius of the welding tool’s shoulder. The heat flux value of the integration point is determined by the total power input and the distance and is returned to the ABAQUS main program.

Figure 7 Work flow of user subroutine DFLUX
The active cooling and heat loss between the workpiece and support plate are defined as user defined interaction and implemented in a user subroutine FILM [34]. The passed-in parameters of FILM include step time, total time, coordinates, flux type, surface name etc. The step time and total time are used to calculate the tool position based on the welding velocity. The coordinates are those of the current point and are represented by a vector. The flux type defines the film condition type (when multiple types of non-uniform film conditions are involved). The surface name relates different surface film conditions to different surfaces. The parameters to be defined in the FILM include the film coefficient and sink temperature for the current point and film coefficient rate with respect to time.

Figure 8 illustrates the implementation of FILM for the distributed active cooling. First, the requested film coefficient defines whether there is active cooling or natural heat loss, based on the surface name of the current integration point and its coordinates. If it is for active cooling, the current coordinates of the welding tool centre is determined, based on step time and welding velocity. Then the location of the current integration point is identified as either inside or outside the cooling area. Finally the film coefficient value for active cooling is calculated.

![Figure 8 Work flow of user subroutine FILM](image)
The models can be divided into two separated steps, namely the welding step and natural cooling step. In the welding step, the heating source and active cooling (if applicable) are loaded.

3.1 Geometry

The model workpiece comprises of two plates with dimensions 50mm*150mm*3mm. Only half is modelled due to symmetric geometry, boundary conditions and loads, as shown in Figure 9. The welding tool moves along the X+ axis.

3.2 Material properties

The material investigated is AA2024, which has properties:
- Thermal conductivity = 121 Wm$^{-1}$C$^{-1}$
- Thermal expansion coefficient = 2.268E-5 $^\degree$C$^{-1}$
- Density = 2780 kg/m$^3$
- Young's Modulus = 73.1 GPa
- Poisson’s ratio = 0.33

Temperature dependent yield strength properties \[35\] are listed in Table 1.

<table>
<thead>
<tr>
<th>Temperature/$^\degree$C</th>
<th>25</th>
<th>100</th>
<th>149</th>
<th>204</th>
<th>260</th>
<th>315</th>
<th>371</th>
<th>424</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield strength/MPa</td>
<td>324</td>
<td>310</td>
<td>248</td>
<td>131</td>
<td>62.0</td>
<td>41.0</td>
<td>27.6</td>
<td>15.3</td>
</tr>
</tbody>
</table>

Temperature dependent values for Young’s Modulus from \[36\] and thermal expansion coefficient from \[37\] are also examined to assess the sensitivity of residual stress to temperature dependence of these properties.

3.3 Boundary conditions and loads

Physically, the workpiece is in contact with a rigid backing plate. As representing this contact interaction using a contact algorithm is computationally expensive, this is simplified to a zero Z direction (vertical) constraint applied to the bottom face of the workpiece in the mechanical model. Heat loss from the other faces to air through radiation and convection is
incorporated in the film condition, where the ambient temperature is constant 25°C and film coefficient is 15 Wm⁻²°C⁻¹. A symmetry boundary condition is applied in the Z-X plane.

In the welding step, the heat loss rate for the bottom face area directly under the welding tool should be larger than other areas because of its Z axis force. Here, the assumed film coefficient in this area is 500 Wm⁻²°C⁻¹, while for other areas it is 200 Wm⁻²°C⁻¹. The total heat input power is programmed in DFLUX as 700W surface heat flux in a circular area of radius 10mm. The welding speed is 5mm/s.

In the non-distributed cooling models and distributed cooling models, the cooling source moves at the same velocity as the tool. The Y direction of one side of the plate is constrained, as shown in Figure 9. In the natural cooling step, the heat input and active cooling are removed, as well as the Y direction constraint on the side.

![Finite element model for friction stir welding](image)

Points A to J are reference points spacing at 1 mm. Point S and T are welding start point and end point respectively. The red circle represents the moving welding tool and the blue rectangle indicates cooling nozzle.

**Figure 9 Finite element model for friction stir welding**
The welding starts 10mm away to one side and ends 10mm away to the other side, as shown in Figure 9. In the natural cooling step, all loads and clamps are released.

3.4 Mesh

The element type used in the model is C3D8RT. The mesh is graduated such that elements increase in size as the location moves from the welding path. A mesh dependence study was performed to determine a suitable mesh size near the welding path. As shown in Figure 10, changing mesh size from 1\text{mm}^3 to 2\text{mm}^3 does not significantly change the temperature curve. The model contains 16200 elements and 22348 nodes.

3.5 Output and solver control

To balance calculation accuracy/convergence and computational cost, the time increment size in both welding step and natural cooling step are automatically controlled. The maximum increment is limited to 0.1 second in the welding step and 1.0 second in the natural cooling step. The non-linear geometry feature is turned on in the solver. Results are output every second in the welding step and every 20 seconds in natural cooling step.

4 Results & Discussion
4.1 No Active Cooling Model

In the model with no active cooling, the temperature profile in the moving coordinate system attached to the welding tool is similar to an oval, as shown in Figure 11. The temperature field for weld time greater 14 seconds is seen to be quasi-static, with the maximum welding temperature increment in the region of $2^\circ$C/second. The reference points identified in Figure 9 achieve their maximum temperature at this moment. Their longitudinal residual stress distribution in the transverse direction is representative of most workpiece area.
The maximum longitudinal residual stress (also referred to as MAX S11) calculated for the no active cooling model is 176.5 MPa. The longitudinal residual stress of the weld is tensile except in the welding start and end sections, and a high value is observed near the tool edge, as shown in Figure 12.

![Figure 11](image1.png)

Figure 11 Temperature field for the no active cooling model at different weld time

Figure 12 shows the longitudinal residual stress distribution along the reference line. The peak value is located 8mm from the welding path (where the tool shoulder radius is 10mm).
4.2 Non-Distributed Cooling Model

The temperature distribution in the model with active cooling at time \( t=14 \) s is shown in Figure 14. The distribution close to the cooling nozzle is significantly different to the distribution for no active cooling shown in Figure 11. The welding temperature under the tool is also reduced.

Active cooling also lowers the longitudinal residual stress. In case of \( D_{cc}=15 \) mm and cooling rate=\( 1.6E4 \) Wm\(^{-2}\)C\(^{-1}\), which is the maximum cooling rate achieved in [29], the max longitudinal residual stress is 122.8 MPa, 53.7 MPa smaller than that of the no active cooling model. In addition, the longitudinal residual stress distribution along transversal direction is...
different, as shown in Figure 13. The disparity between the tool centre and near the edge is increased: on reference line it is 101 MPa, compared to 40 MPa in the no active cooling model.

The welding temperature drops dramatically when $D_{cc}$ is decreased in Figure 15 though the longitudinal residual stress also gets lower. When $D_{cc}=15\text{mm}$, the welding temperature at weld time $t=14\text{s}$ drops from original $433.2^{\circ}\text{C}$ to $348.1^{\circ}\text{C}$. These findings agree well with [29].

![Figure 15 Longitudinal residual stress and welding temperature variation with $D_{cc}$ for non-distributed cooling](image)

4.3 Distributed Cooling Model

Two distributed cooling arrangements are considered. In the first arrangement, Zone 1 (defined in Figure 3) is moved a further 2mm away from the tool centre. In the second arrangement, it is moved another 2mm.

The temperature field for distributed cooling, shown in Figure 16 for time $14\text{s}$, differs from that found for non-distributed cooling. First, the temperature contour under the cooling nozzle is almost rectangular for non-distributed cooling, as seen in Figure 14 while in distributed cooling this area becomes smaller and is not rectangular. Secondly, the
welding temperature is increased. These changes in response are related to the adjusted cooling areas with greater centre-to-centre distances. Consequently, max S11 drops from 122.8 MPa to 118.8 MPa.

The temperature of the material under the welding tool is significantly increased in the distributed cooling models compared to the non-distributed cooling, as shown in Figure 18 at weld time t=14s, where Temp-dis1 identifies the first arrangement and Temp-dis2 the second. In addition, the temperature gradient is decreased, as shown in Figure 17. Here the temperature gradient is quantified by temperature difference between two 5mm spaced nodes (near the reference line) in the longitudinal direction.
In Figure 13, the longitudinal residual stress near the welding path is increased, compared to the non-distributed cooling, whilst for material near the tool edge it is slightly decreased. Though the temperature is higher at the tool centre, as shown in Figure 17, the temperature gradient is significantly decreased in distributed cooling, indicating lower tensile plastic strain will be generated after the tool has passed. For material near the tool edge, the temperature gradient is almost the same but the temperature is increased in distributed cooling. These findings support the simplified analysis of the welding temperature and temperature gradient presented in Section 2.

5 Conclusions

This paper compares the temperature, temperature gradient and longitudinal residual stress for FSW processes including a non-distributed cooling model and two distributed cooling models. The main findings of the investigation are:

1. The simplified analytical and FE models considered in the investigation indicate that distributed cooling may result in higher welding temperature than conventional cooling. In practice, this response would be influenced by factors like the actual input heat flux, the material flow introduced by the rotating tool, the real heat conductivity, the actual cooling rate of the cooling nozzle and the convection rates.

2. Distributed cooling changes the temperature gradient distribution in FSW. The temperature gradient near the welding path is significantly affected by distributed cooling, whilst for regions away from the welding path, it is almost unaffected.

3. The results presented indicate that distributed cooling arrangements such as those considered could lead to improved mitigation of FSW residual stress compared to a conventional cooling arrangement of the same cooling power. The possibility of
achieving both higher welding temperature and lower residual stress through distributed cooling is identified in the numerical models, indicating that the welding temperature plays an import role in reducing residual stress.

4. Experimental investigation of the results from the analytical and numerical models is required to validate the findings of the investigation and will be undertaken in future work.

References


