Recent advances in friction-stir welding – Process, weldment structure and properties

R. Nandana, T. DebRoy, H.K.D.H. Bhadeshia

Article info

Article history:
Received 7 May 2008
Accepted 18 May 2008

Abstract

Friction-stir welding is a refreshing approach to the joining of metals. Although originally intended for aluminium alloys, the reach of FSW has now extended to a variety of materials including steels and polymers. This review deals with the fundamental understanding of the process and its metallurgical consequences. The focus is on heat generation, heat transfer and plastic flow during welding, elements of tool design, understanding defect formation and the structure and properties of the welded materials.

© 2008 Elsevier Ltd. All rights reserved.

Contents

1. Introduction ............................................................ 981
2. FSW process ............................................................ 982
   2.1. Heat generation .................................................. 982
   2.2. Principles of heat transfer and material flow ................. 984
   2.3. Material flow and mechanism of joining ..................... 988
   2.4. Temperature fields and cooling rates ....................... 993
   2.5. Welding variables ............................................. 997
   2.6. Tool design ................................................... 998
   2.7. Defects ....................................................... 1001
   2.8. Residual stress ............................................... 1003
3. Microstructure and properties of friction-stir welded alloys .......... 1005
   3.1. Aluminium alloys ............................................. 1005
   3.2. Magnesium alloys ............................................. 1009
   3.3. Copper alloys ................................................ 1009
   3.4. Titanium alloys .............................................. 1009
   3.5. Steels .......................................................... 1010

* Corresponding author.
E-mail addresses: debroy@psu.edu (T. DebRoy)

0079-6425/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pmatsci.2008.05.001
1. Introduction

Friction-stir welding (FSW) is a solid-state, hot-shear joining process [1–3] in which a rotating tool with a shoulder and terminating in a threaded pin, moves along the butting surfaces of two rigidly clamped plates placed on a backing plate as shown in Fig. 1. The shoulder makes firm contact with the top surface of the work-piece. Heat generated by friction at the shoulder and to a lesser extent at the pin surface, softens the material being welded. Severe plastic deformation and flow of this plasticised metal occurs as the tool is translated along the welding direction. Material is transported from the front of the tool to the trailing edge where it is forged into a joint. Although Fig. 1 shows a butt joint for illustration, other types of joints such as lap joints and fillet joints can also be fabricated by FSW.

The half-plate where the direction of rotation is the same as that of welding is called the advancing side, with the other side designated as being the retreating side. This difference can lead to asymmetry

Fig. 1. (a) Schematic illustration of the friction-stir welding process. (b) An FSW weld between aluminium sheets. (c) An actual tool, with a threaded-pin.
in heat transfer [4], material flow and the properties of the two sides of the weld. For example, the hardness of particular age-hardened aluminium alloys tends to be lower in the heat-affected zone on the retreating side, which then becomes the location of tensile fracture in cross-weld tests [5]; this is also the case for pure titanium [6].

Since its discovery in 1991 [1], FSW has evolved as a technique of choice in the routine joining of aluminium components; its applications for joining difficult metals and metals other than aluminium are growing, albeit at a slower pace. There have been widespread benefits resulting from the application of FSW in for example, aerospace, shipbuilding, automotive and railway industries [7].

FSW involves complex interactions between a variety of simultaneous thermomechanical processes. The interactions affect the heating and cooling rates, plastic deformation and flow, dynamic recrystallization [8] phenomena and the mechanical integrity of the joint. A typical cross-section of the FSW joint consists of a number of zones (Fig. 2) [9–12]. The heat-affected zone (HAZ) is similar to that in conventional welds although the maximum peak temperature is significantly less than the solidus temperature, and the heat source is rather diffuse. This can lead to somewhat different microstructures when compared with fusion welding processes. The central nugget region containing the “onion-ring” appearance is the one which experiences the most severe deformation, and is a consequence of the way in which a threaded tool deposits material from the front to the back of the weld. The thermomechanically affected zone (TMAZ) lies between the HAZ and nugget; the grains of the original microstructure are retained in this region, but often in a deformed state.

A unique feature of the friction-stir welding process is that the transport of heat is aided by the plastic flow of the substrate close to the rotating tool. The heat and mass transfer depend on material properties as well as welding variables including the rotational and welding speeds of the tool and its geometry. In FSW, the joining takes place by extrusion and forging of the metal at high strain rates. Jata and Semiatin estimated a typical deformation strain rate of 10 s\(^{-1}\) by measuring grain-size and using a correlation between grain-size and Zener–Holloman parameter which is temperature compensated strain rate [13]. Kokawa et al. estimated an effective strain rates in the range 2–3 s\(^{-1}\) [14]. The plastic flow must clearly feature in any theory for the process, and the behaviour of the metal at high strain rates, its dynamic recrystallization behaviour and the effects of heating and cooling must also be considered.

The role of welding parameters, tool design, the initial process modelling, and the microstructure and properties of the FSW joints have been reviewed previously [15,16]. It is nevertheless a relatively new process and one which is evolving rapidly. As a result, a periodic critical assessment of our understanding of the welding process as well as the structure and properties of the welded materials is likely to be helpful. This is the aim of the paper.

2. FSW process

2.1. Heat generation

During FSW, heat is generated by friction between the tool and the work-piece and via plastic deformation. A fraction of the plastic deformation energy is stored within the thermomechanically processed region in the form of increased defect densities. In the weld, a mixture of recovery and

![Fig. 2. Schematic cross-section of a typical FSW weld showing four distinct zones: (A) base metal, (B) heat-affected, (C) thermomechanically affected and (D) stirred (nugget) zone [9].](image)
recrystallization phenomena occur simultaneously [17]. Deformation not only increases the dislocation density but also the amount of grain surface and grain edge per unit volume [18] and by cutting precipitates may force them to dissolve [19–24].

In FSW, the tool moves along the weld joint at a constant speed $U$, as it rotates about its axis with speed $\omega$. At any point on the tool–work-piece interface, the tangential speed of the tool with respect to the work-piece is given by $v_t = \omega r - U \sin \theta$ where $r$ is the radial distance from the tool–axis and $\theta$ is the angle between radial vector, $r$, and the welding direction. The term $U \sin \theta$ may be neglected when $\omega r$ is much larger. Heat is generated due to friction and plastic deformation at the tool–work-piece interface and due to plastic deformation in the TMAZ. The local interfacial heat generation due to friction is the product of frictional force and the sliding velocity. The interfacial deformation heat is the product of shear stress and the velocity of the work-piece material which sticks to the tool as it moves.

The local heat generation rate due to friction, $d\dot{e}_f$, is approximately given by [15,25–36]:

$$d\dot{e}_f = \delta(\omega r - U \sin \theta)\mu_d p dA$$

where $\delta$ is the extent of slip, $\mu_d$ is the friction coefficient and $p$ is the local pressure applied by the tool on the elemental area $dA$. When $\delta$ is 1, no material sticks to the tool and all the heat is generated by friction. In contrast, when $\delta = 0$, all the heat is generated by plastic deformation. Schmidt et al. [33] provide a comprehensive discussion on the calculation of interfacial heat generation rates during FSW. A problem in the calculations of heat generation is that the friction coefficient cannot be determined from fundamental principles or it seems, by straightforward representative experiments of relevance to the conditions of FSW.

When the work-piece material sticks to the tool, heat is generated at the tool–work-piece contact due to shear deformation. The resulting heat generation, assuming 100% efficiency of conversion of deformational work into heat, may be approximated as

$$d\dot{e}_s = (1 - \delta)(\omega r - U \sin \theta)\tau_V dA$$

where $\tau_V$ is the shear yield stress related to that in tension ($\sigma_V$) by the von Mises criterion $\tau_V = \sigma_V / \sqrt{3}$. It is possible to use an effective $\tau_V$, back calculated from the experimentally determined average power input to the system [37,38], or estimated from the local shear stress using a temperature and strain rate dependent yield stress [39–41].

Plastic deformation also occurs away from the tool–work-piece interface and its effect on the local heat generation rate may be estimated as $d\dot{e}_p = \beta \mu_\phi dV$, where $\beta$ is the fraction of plastic deformation work which is dissipated as heat, $\mu$ is the non-Newtonian viscosity of the plasticized material and $\phi$ is given by [42]:

$$\phi = 2 \sum_{i=1}^{3} \left( \frac{\partial u_i}{\partial x_1} \right)^2 + \left( \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right)^2 + \left( \frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right)^2 + \left( \frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3} \right)^2$$

When materials are well mixed, the fraction of plastic deformation work which is converted to heat ranges between 0.6 for low strain rate ($1 \text{ s}^{-1}$) and 0.8 for high strain rate ($3000 \text{ s}^{-1}$) for the FSW of $\alpha$-titanium [43,44]. However, in FSW heat is generated by deformation and grain boundary sliding. Bastier et al. [45] showed that the overall magnitude of the heat generation rate by viscous dissipation is small, about 4.4% of total heat generation rate.

In practice, the errors in the calculation of heat generation rates can arise from a number of sources. For example, the value of the friction coefficient can vary with temperature, relative velocity and other factors. As the work-piece is heated, localised softening reduces friction and the heat generation rate. Some investigators have considered this effect by adjusting the coefficient of friction. However, there is no straightforward method to estimate the coefficient of friction or how it changes with temperature or relative velocity. In the context of heating resulting from plastic deformation, the decrease in the yield strength with increasing temperature leads to a reduction in the heat generation rate by this mechanism. Since strain rates depend on velocity gradients, which diminish rapidly away from the tool, most of the heat generation due to plasticity occurs close to the tool–work-piece interface. Verification of the heat generation rate calculations has been done indirectly either by measuring thermal cycles or by measuring mechanical energy expended during FSW [38]. Su et al. performed calorimetric
studies on friction-stir spot welds of AA 6111 [46]. After welding, they transferred the work-piece into an insulated flask of oil and measured the temperature increase of the oil thus determining the heat content by calorimetry [46].

As Eqs. (1) and (2) indicate, the heat generation rate is influenced by the rotational speed but not by the welding speed [40]. Surprisingly, for the same shoulder dimension, tool pin design does not seem to have a significant effect on the total heat generation rate during FSW (Fig. 3) [44,47].

2.2. Principles of heat transfer and material flow

Except for the initial and the final periods of welding, i.e., during tool pin insertion and extraction, the heat generation occurs at a constant rate if the tool rotates and moves forward at a constant speed. This steady-state assumption is justified by the fact that the weld profile and properties remain roughly constant during the welding phase. The temperature and velocity fields in pseudo-steady-state are commonly obtained by solving the generally available continuity, momentum and energy equations for incompressible single-phase flow assuming steady state. The steady-state thermal energy conservation equation in index notation, for \( i = 1, 2 \) and 3, representing \( x, y \) and \( z \)-directions, respectively, is given by

\[
\rho C_p \frac{\partial (u_i T)}{\partial x_i} = -\rho C_p U \frac{\partial T}{\partial x_i} + \frac{\partial}{\partial x_i} \left( k \frac{\partial T}{\partial x_i} \right) + S_b
\]

where \( \rho \) is the density, \( u_i \) is the material velocity in \( i \)-direction, \( T \) is the temperature, \( x_i \) is \( i \)-coordinate, \( C_p \) is the specific heat capacity at constant pressure, \( k \) is the thermal conductivity of the work-piece and \( S_b = \frac{\partial \phi}{\partial V} = \beta \mu \phi \) is the heat generation rate per unit volume, due to plastic deformation in the work-piece away from the interface.

Of the heat generated at the shoulder–work-piece interface, some of it is transported into the tool material while the rest enters the work-piece. The total heat generated at the shoulder–work-piece interface can be partitioned between the work-piece \( (J_W) \) and the tool \( (J_T) \), based on their thermo-physical properties [48]:

\[
f = \frac{J_W}{J_T} = \left( \frac{k \rho C_p}{k \rho C_p} \right)^{1/2}
\]

Eq. (5) assumes steady-state, one-dimensional heat flow from the interface of dissimilar metal joints [49]. At 1000 K, the estimated heat flux into the work-piece is calculated to be 43% of total heat

Fig. 3. Heat input at various welding speeds used to weld 6.35 mm thick AA 7075-T7351 plates using MX-Triflute, MX-Trivex and Trivex tools [44].
generated in 1018 Mn steel welded using tungsten tool. This relation has been examined experimentally by Lienert et al. [48] and found to be reliable.

The extent of slip can be estimated by curve fitting the measured values at various relative velocities [50]:

$$\delta = 0.2 + 0.8 \left(1 - \exp \left(-\delta_0 \frac{\omega R}{\omega_0 R_S}\right)\right)$$

(6)

where $\delta_0$ is an adjustable parameter, $R_S$ is the radius of the tool shoulder, $\omega_0$ is the normalising rotational velocity which can be taken as the mid-point of the range of rotational speeds.

Values of friction coefficient can be estimated considering the relative velocity between the tool and the work-piece according to previous work in the context of friction [51]. The relative velocity increases from zero at the axis of rotation to $\omega R_S$ at the periphery of the tool shoulder. The available data can be fitted in the following form [51]

$$\mu_t = \mu_0 \exp \left(-\frac{\delta_0 \frac{\omega R}{\omega_0 R_S}}{\rho_0 R_S}\right)$$

where $\mu_0$ is a fitting constant.

The boundary condition for heat exchange between the top surface of the work-piece and the surroundings beyond the shoulder involves consideration of both convective and radiative heat transfer as

$$-k \frac{\partial T}{\partial z}_{\text{top}} = \sigma \epsilon (T^4 - T_a^4) + h(T - T_a)$$

(7)

where $\sigma$ is the Stefan–Boltzmann constant ($5.67 \times 10^{-12}$ J·K$^{-4}$·cm$^{-2}$·s$^{-1}$), $\epsilon$ is the emissivity, $T_a$ is the ambient temperature and $h$ is the heat transfer coefficient at the top surface.

At the bottom surface, the heat transfer into the backing plate is modelled by an enhanced heat transfer coefficient [38]:

$$k \frac{\partial T}{\partial z}_{\text{bottom}} = h_b(T - T_a)$$

(8)

where $h_b$ is the heat transfer coefficient at the bottom surface. Several investigators [37, 38, 52, 49] have examined the important role played by the boundary conditions at the bottom surface. Khandkar et al. [38] reported that for FSW of AA 6061-T651 plates, an overall convective heat transfer coefficient of 1000 W·m$^{-2}$·K$^{-1}$ might be appropriate for the bottom surface of the work-piece if the backing plate is not considered. However, they observed that when a stainless steel backing plate is taken into account, a variable gap conductance would be appropriate for the work-piece/backing plate interface, and recommended an average gap conductance somewhat less than 5000 W·m$^{-2}$·K$^{-1}$. They suggested that significant variations in the heat transfer rates can occur depending on the specific experimental conditions and recommended determining the rate experimentally. Since the accuracy of the computed temperature field depends on the value of the heat transfer coefficient, the uncertainty in the heat transfer coefficient can significantly affect the reliability of the computed temperature field. Significant improvement in the reliability of the model predictions can be achieved by determining the uncertain parameters, such as the heat transfer coefficient using a limited volume of experimental data. This approach has been tested for FSW of dissimilar aluminum alloys [53] and for fusion welding [54–57].

In recent years there has been a growing recognition of the importance of materials flow in the calculations of heat transfer during friction-stir welding. These calculations have adapted two approaches in the modelling of flow. In many calculations, the plasticised material was treated as a high viscosity fluid and the flow field was obtained using computational fluid dynamics. In some other cases the plastic deformation was modelled from the principles of solid mechanics using the finite volume method to solve for displacements.

The continuity equation for incompressible flow is given by

$$\frac{\partial u_i}{\partial x_i} = 0$$

(9)
The momentum conservation equations with reference to a coordinate system with origin at the tool-axis and moving with the tool at a constant speed $U$ along the $x$-axis are:

$$p \frac{\partial u_i}{\partial x_j} = - \frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_i} \left( \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right) - \rho U \frac{\partial u_j}{\partial x_i}$$

(10)

where $P$ is the pressure. Notice that in contrast to Eq. (1) where $p$ is the pressure applied by the tool on the work-piece, $P$ is a relative pressure which drives flow.

The speeds with which material moves at the tool pin periphery, $u$, $v$ and $w$, along the welding direction, the normal to the welding direction in the plane of the plate being welded, and normal to the plane of the plate, respectively, are given by

$$u = \frac{1}{\sinh h_e} (\omega \sin \theta - U), \quad v = (1 - \delta) \omega \cos \theta, \quad w = \Psi \omega$$

(11)

where $\Psi$ is the pitch of the threads on the cylindrical tool. At the shoulder $w = 0$.

The non-Newtonian viscosity $\mu$ must be estimated as a function of strain rate and temperature using experimentally determined constitutive equations for the material of interest. According to Perzyna, $\mu$ can be expressed in terms of effective flow stress $\sigma_e$ (i.e. deviatoric stress) and effective strain rate $\dot{\varepsilon}$ (i.e. deviatoric strain rate) [58],

$$\mu = \frac{\sigma_e}{3\dot{\varepsilon}}$$

where $\dot{\varepsilon} = \left( \frac{2}{3} \dot{\varepsilon}_i \dot{\varepsilon}_j \right)^{\frac{1}{2}}$ and $\dot{\varepsilon}_ij = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$

(12)

The strain rate is found to correlate with the flow stress and temperature as follows [59]:

$$\dot{\varepsilon} = A (\sinh \alpha \sigma_e)^n \exp \left( - \frac{Q}{RT} \right)$$

(13)

where the material constants $A$, $n$, $\alpha$ and the apparent activation energy $Q$ are derived by fitting the equation to experimental data, and are all independent of temperature. On rearranging this equation [60,61],

$$\sigma_e = \frac{1}{\alpha} \sinh^{-1} \left[ \left( \frac{Z}{A} \right)^{\frac{1}{n}} \right]$$

(14)

where $Z = \dot{\varepsilon} \exp (Q/RT)$ is known as the Zener–Holloman temperature compensated strain rate [62]. Sheppard and Jackson determined the fitting constants for a number of aluminium alloys [63]. Bruschi et al. [64] studied hot workability of Ti–6Al–4V alloys and estimated values of the constants in the constitutive equation for this alloy. Material constants could be determined for other alloys from hot working literature.

The computed variation of viscosity as a function of strain rate and temperature has been reported for several materials: 6061 aluminium alloys [39], 304 stainless steel [40] and 1018 C–Mn steel [41]. Fig. 4 shows that the viscosity decreases significantly with both strain rate and temperature with the former being the dominant factor for the conditions typical of FSW.

It should be emphasised that in Eq. (13), the parameters $A$, $n$ and $\alpha$ are strictly functions of strain in a work-hardening solid. However, they should become independent of strain once a steady-state is reached during deformation in which the work-hardening is balanced by recovery-softening. Generally this state is reached before strain drops to a low value, $\varepsilon = 1$, and since the strains involved in FSW are much larger, it may be assumed that a single set of material constants corresponding to steady-state deformation can be utilised with the assumption that they are independent of strain.

Most of the recent studies have considered the temperature dependent thermal conductivity, specific heat and yield stress for the work-piece [39,41,40,66,67,65]. The tool material commonly used for the FSW of aluminium alloys, steels and Ti–6Al–4V are tool steel and tungsten, respectively. Table 1 presents the thermophysical properties for these alloys.

Plastic deformation can be modeled using solid mechanics, which, solves for displacements instead of velocities. Ulysse [65] used a solid mechanics approach assuming a rigid-visco-plastic material where the flow stress depended on strain rate and temperature. The heat generation rate was
expressed as the product of the effective stress and the effective strain rate. Temperature profiles were determined for the work-piece and the tool, using a three-dimensional finite analysis code.

Schmidt et al. [36] used adaptive boundary conditions to determine conditions for void-free welds using finite element analysis and flow stress determined according to the Johnson–Cook law. Buffa et al. [68] developed a thermomechanically coupled, rigid-visco-plastic, three-dimensional finite element model to study the effect of tool geometry and welding velocity on material flow pattern and the grain-size distribution in the welded joints.

In the finite element method, the mechanical response on an element is obtained using force balance:

$$q \varepsilon u + c _ { \dot{u}} + k u = p$$  \(15\)

where \(q\) is the density, \(c\) the damping coefficient, \(k\) the stiffness coefficient, \(p\) the body force, \(u\) is the displacement vector, \(\dot{u}\) is the velocity vector and \(\ddot{u}\) is the acceleration. In a finite element framework, the equation can be written as

$$M \ddot{u} + C \dot{u} + Ku = P$$  \(16\)

where \(M\) is the discrete mass matrix, \(C\) the viscous damping matrix, \(K\) the stiffness matrix, \(P\) the vector of external discrete forces, which include body forces, surface forces and concentrated loads acting on the system, and \(u, \dot{u}\) and \(\ddot{u}\) are the nodal displacement, velocity and acceleration vectors, respectively.

The thermal response is governed by an energy balance:

$$\rho C_p \dot{T} = \frac{d}{dx_i} \left( k \frac{d T}{dx_i} \right) + \eta S_s \dot{e}_p^{pl}$$  \(17\)

where \(C_p\) is the specific heat capacity at constant pressure, \(\dot{T}\) is the rate of change of temperature with time, \(k\) the thermal conductivity, \(\eta\) the fraction of plastic energy dissipation, \(S_s\) the deviatoric stress tensor and \(\dot{e}_p^{pl}\) is the plastic strain rate tensor. In FE framework, we obtain

$$CT + BT = S$$  \(18\)

where \(C\) is the discrete capacity matrix, \(B\) the conductivity matrix, \(S\) the source vector accounting for all thermal sources and \(T\) and \(\dot{T}\) are the nodal temperature and temperature rate vectors, respectively.

Since FSW involves large plastic deformation analogous to ballistic impact, an Eulerian finite-difference code that simulates shock-wave in 3D by solving time-dependent continuum mechanics equations has also been used to model friction-stir welding in aluminium [69,70]. This code is well suited for modelling very large deformations at high strain rate.

In polycrystalline materials, large plastic deformation alters microstructure and changes properties. Boyce et al. [71] suggested that to adequately model the deformation of a polycrystalline material,
it is necessary to understand the interaction of the material on two length scales. Multi-scale modelling of deformation in FSW has been performed with one length scale of the order of the dimension of the work-piece (in cm), and the other of the order of the grain-diameter (in μm). Temperature and velocity fields were calculated and the texture evolution was predicted using the velocity gradients along streamlines of the flow field together with a model for polycrystal plasticity. Large variation in textures was obtained suggesting strong variation in mechanical properties across the weld zone.

### 2.3. Material flow and mechanism of joining

Recent experimental and computational works have provided significant insight about several interesting features of materials flow in FSW and the joining process. Most of the material flow occurs through the retreating side and the transport of the plasticized material behind the tool forms the welded joint. Three types of flow affects the overall transport of plasticized materials during FSW. First, near the tool, a slab of plasticized material rotates around the tool. This motion is driven by the rotation of the tool and the resulting friction between the tool and the work-piece. Second, rotational motion of the threaded pin tends to push material downward close to the pin which drives an upward motion of an equivalent amount of material somewhat farther away. Finally, there is a relative motion between the tool and the work-piece. The overall motion of the plasticized material and the formation of the joint results from the simultaneous interaction of these three effects.

The plastic flow models for FSW have been used to predict velocities around the tool pin. The velocities have also been estimated from strain rates which, in turn, were obtained from the correlation between grain-size and strain rate [13]. Comparison of the shape of the TMAZ predicted by flow models with macrostructural observation has shown satisfactory match [39]. Good agreement between the torque values obtained using dynamometers and the computed values from the flow models for 304L stainless steel [40], 1018 Mn steel [41] and Ti–6Al–4V alloy [72] indicates the usefulness of the models to understand the FSW process. Since torque is a measure of the shear stress on the tool and since the shear stress on the work-piece is responsible for both heat generation [38] and plastic flow, validation of the model predictions by experiments indicates that it is appropriate to use the models for the estimation of several important parameters. Though numerical modelling of plastic flow can aid tool design and the optimisation of weld quality, there does not appear to have been an application of models towards the prediction of practical processing maps [73].
The computed stream traces on horizontal planes around the tool pin at three different elevations are shown in Fig. 5. The stream lines indicate the presence of rotational zone, which clearly shows the recirculating flow of a plug of material around the tool pin. The thickness of the recirculating material flow region is affected by material properties, welding parameters and rate of heat transfer into the tool. This zone occupies larger areas at higher elevation planes due to greater momentum transport from the rotating shoulder. The streamlines indicate that beyond the region of recirculating plastic flow, i.e. in the transition zone, material transfer occurs mainly on the retreating side. Fig. 5 also shows a flow reversal in the advancing side close to the pin, leading to a relatively stagnant zone, which forms closer to the pin at lower elevations. An important consequence of the lack of adequate material flow on the advancing side has been related to the formation of “wormhole” defects [74]. The streamlines show that beyond the rotational zone the material transport occurs mainly along the retreating side. Flow visualisation using tracers also indicates the presence of a zone where the material rotates and advances with the tool and transitional zone where the materials move on the retreating side [75–77].

The maximum velocity is attained near the edge of the shoulder at the top surface of the work-piece followed by a rapid decrease away from this region [39–41], as shown in Fig. 6. At planes near the bottom of the work-piece, the peaks in velocity were attained near the tool surface. The computed viscosity contours at different horizontal planes (Fig. 6) shows that the viscosity lies in the range of $10^5$ to $5 \times 10^6$ Pa s for FSW of aluminium alloys. It is also seen that no significant flow occurs when the...
viscosity is higher than about $5 \times 10^6$ Pa s and that the region of plastic flow decreases with depth. It was observed that plastic flow ceases beyond a certain critical value of viscosity. This cut-off viscosity surface defines the geometry of the thermomechanically affected zone [39].

The flow during FSW differs in character from material flow in the liquid state during conventional fusion welding. This is evident from studies of the welding of dissimilar metals. During fusion welding, the weld pool becomes compositionally homogenous after solidification. However, during FSW of dissimilar metals, mixing does not occur in atomic scale, and it is possible to find larger concentration differences in the weld metal and the region is far from homogenous. For example, Fig. 7 shows a diffusion couple formation between Fe and Ni at a certain location in stir region, when pure Fe and Ni plates are joined together using FSW. The diffusion couple has a length scale of only 2–3 μm. Such diffusion couples would not exist had melting and homogenisation of weld metal occurred.

During FSW, the material near the top of the work-piece is stirred under the action of the shoulder and vertical component of motion occurs mainly due to threading on the tool pin. The stirred material from the top is carried down by the threads and deposited in the weld nugget. The vertical mixing becomes prominent at low weld pitch, which is the ratio of welding speed to rotational speed. One way to

---

**Fig. 6.** Spatial variation of viscosity and velocity in AA 6061-T6 [39] at planes corresponding to $z = 1.27, 4.67, 8.07$ and $11.47$ mm for a plate thickness of 12.7 mm. Distances in $x$- and $y$-direction were equivalent, but that in the $z$-direction was increased eight-fold to enhance clarity. The welding velocity was 1.59 mm s$^{-1}$ and the rotational speed was 637 rpm.

**Fig. 7.** Concentration profiles of Fe and Ni at a location in stir region for FSW of pure Fe and Ni. Points represent data obtained by atomic emission spectroscopy and the solid lines indicate computed results [78].
understand material flow experimentally is to use inert markers before starting the weld [77], and then determine their final positions using serial sectioning parallel to the top surface. The tracer technique does not reveal the actual flow path of the material but it shows the final position of the tracer. Schmidt et al. [75] estimated the average velocity of material flow through shear layers during FSW of aluminium alloy based on experimental investigation of tracer flow. They estimated the average velocity to be approximately 0.1–0.3 times the shoulder rotational speed. The order of magnitude of the material velocity is comparable with the reported values from computational flow models [39].

An interesting experiment involved the introduction of copper markers parallel and perpendicular to the weld centreline, in both real experiments and computational models during FSW of AA 7075-T6 [79]. The positions of the markers after welding were then compared for the two cases, enabling some qualitative conclusions to be made about the material flow as a function of position, and on the nature of the material bonding that takes place in the advancing side. It was concluded that the weld pitch (welding velocity/rotational speed) and the tool pin shape determined whether the flow was simple or complex in nature. When pitch was small, markers were continuous rather than dispersed, indicating better welds [80]. A conical tool pin led to more effective material flow with fewer defects. This is analogous to chip formation in the machining processes – just as chip formation morphology changes from discontinuous to segmented to wavy to continuous with increase in cutting speed, the flow pattern in FSW also changes with weld pitch [79] as shown in Fig. 8.

Steel balls have been used to trace flow in butt welds of 6061 and 7075 aluminium alloys [82], their positions after welding being determined using radiography. Several initial configurations of balls were investigated and the “stop action technique” was used by suddenly interrupting the forward motion of the pin which is then quickly unscrewed from the work-piece, leaving the threading in the keyhole intact. When the trailing edge of the keyhole was examined using metallography, it showed microstructural banding of the extruded material. The bottom portion indicated vertical striations, whereas horizontal striations were observed close to the shoulder indicating downward and rotational flows, respectively.

![Fig. 8. Analogy between chip morphology in machining and material flow in FSW [79]. Mid-plane horizontal sections of AA 7075-T6 show the re-distribution of copper about the center-line after welding when copper foil is placed between the plates to be welded. Tool was made of C40 steel, with shoulder diameter of 12 mm. The cylindrical pin had a diameter of 3.0 mm while the conical pin had a major diameter equal to 5.00 mm and minor diameter of 1.90 mm and both were 2.70 mm in height (for the conical pin, this corresponds to cone angle 30°).](image-url)
Some idea about the flow of materials can be obtained when microscopic observations are combined with knowledge of the crystallographic texture and microtexture [83]. Far from the tool, the undeformed grains have low-angle boundaries while near the tool, elongated subgrains form because of the rotation induced by the tool. These subgrains eventually develop high-angle (≥ 15°) grain boundaries [84]. Orientation imaging microscopy can provide shear texture maps which, in turn, gives an idea of the strength of flow.

Macrostructural observations of transverse cross-sections of FSW-welded aluminium alloy specimens reveals “onion-ring” shaped structures of the type illustrated in Fig. 9. These have been linked to the nature of material flow during FSW [85,86,81], but the detailed mechanism of pattern formation is not completely understood. The pattern has been attributed to the geometry of the extrusion of cylindrical sheets of material during each rotation of the tool [86]. Measurement of the spacing of the markings has been found to be equal to the weld pitch, i.e., the length the tool moves forward in one rotation. Electron back scattered diffraction studies indicate that the bands have different densities of second-phase particles, rather than any significant difference in the local grain structure or texture [87,88]. This is confirmed by studies of crack propagation in FS-welds [89], where it is observed that cracks deflect around the onion-ring structure, consistent with local variations in hardness. In AA 2024-T351 rolled sheet, segregated banded microstructure consisting of alternating hard particle-rich and hard particle-poor regions form with the crack path tending to follow regions of high particle density [90].

In FSW of dissimilar aluminium alloys of cast A356 and wrought 6061, onion-ring patterns consist of lamellar mixture of the two alloys of equal width [91]. When AA 2024 and AA 6061 are welded, the intercalated flow pattern can be visualized by differential etching of the AA 2024 producing contrast relative to AA 6061 [92–94]. Apart from welding parameters, the base metal microstructure also affects the rings. For example, direct-chill casting of AA 5182 produces a dense and uniformly distributed intermetallic particle structure while during strip casting intermetallics are broken-up during hot rolling and distributed along the primary grain boundaries [95]. Strip-cast base material, with intermetallic particles present in bands, show a definite onion-ring structure in the weld. In contrast, the diffuse rings appear in the weld when the starting material is homogeneous [96].

As a result of recent research, several important features of material flow during FSW are now well understood. The flow path of the plasticized materials across both the advancing and the retreating sides of the tool and the process of joining is now better understood. However, recent research has also uncovered several complexities and special features of the material flow in FSW such as the lack of mixing of the plasticized materials in atomic scale and the formation of certain interesting patterns that still require further work.

![Fig. 9. Optical image showing the macroscopic features [81] in a transverse section of a friction-stir welded 2195-T81 Al–Li–Cu alloy. Note the onion-ring and the adjacent large upward movement of material.](image-url)
2.4. Temperature fields and cooling rates

The welding variables and the material properties affect the temperature profiles, cooling rates, microstructure and the resulting properties of the welded joint. The temperature fields in FSW exhibit certain special features. The peak temperatures are significantly lower than those attained in conventional fusion welding processes. Furthermore, diffuse heat source in the FSW process and relatively low welding speed often results in low cooling rate in FSW.

Thermal cycles in FSW can be measured using thermocouples (Fig. 10) [97] and in situ neutron-diffraction (Fig. 11) [98] where measured lattice-distortion is related to change in temperature. When experimental data are not available, a recourse is to estimate the temperature profiles and the cooling rates using a computational model after the model has been properly validated with experiments.

Fig. 10. 3D plot of thermal cycles at the bottom surface along the weld centerline of AA 6061-T6 [97]. The thermal cycle at the front is for the location where tool pin is inserted into the work-piece while other thermal cycles correspond to locations along the weld centreline in the direction of tool movement.

Fig. 11. Instantaneous temperature distribution, measured using neutron-diffraction, on the top surface along the weld centerline behind the the tool pin, for FSW of AA 6061-T6 [98].
Asymmetry in the temperature field is also a special feature of FSW. For example, in welds performed on 304L stainless steel, the experimental and the computed peak temperatures were about 100 K higher on the advancing side than the retreating side, as shown in Fig. 12 [4]. This large temperature difference in the two sides is higher than the values typically observed in the FSW of other materials under most welding conditions. However, the asymmetry in temperature can be appreciated owing to very low thermal conductivity of stainless steel compared to most other alloys.

Some of the early work on heat transfer during FSW was based on simple analytical or numerical heat conduction models that neglected convective heat transfer due to motion of the plasticised material. Models which neglect convective heat transfer are limited to materials with high thermal conductivity (aluminum alloys) or low Péclet number, $Pe$, the latter defined as

$$Pe = \frac{\rho C_p u_c L}{k}$$

where $u_c$ is the characteristic velocity, and $L$ is the characteristic length. The Péclet number indicates the relative importance of heat transfer by convection to that by conduction. When $Pe$ is much lower than one, heat is transported mainly by conduction. Under these conditions, the errors in the calculation of temperature fields would not be significant if convection were ignored. Let us consider a typical FSW of an aluminum alloy. Considering approximate values of $\rho$, $C_p$ and $k$ as 2700 kg m$^{-3}$, 3.77 J kg$^{-1}$ K$^{-1}$ and 7.02 J m$^{-1}$ s$^{-1}$ K$^{-1}$, respectively, and assuming $L$ to be the average length of stir zone (6 mm) and $u \approx 150$ mm s$^{-1}$, the Péclet number is 12. Therefore, even for a high thermal conductivity material such as aluminium, convective heat transfer is an important mechanism for heat transfer near the tool during FSW. The typical order of magnitude of the Péclet number for steel is 100. For such high values of $Pe$, the temperature profiles cannot be accurately calculated ignoring convective heat transfer. When convective heat transfer is ignored, the peak temperature and the size of the deformation zone are overestimated and the flow stress and the tool force are underestimated [65,43]. Furthermore, since the two-dimensional models [74,99] do not consider vertical mixing during FSW, their application in high Péclet systems can lead to overprediction of local temperature.

Plastic flow in FSW is quite complex due to an interplay of variation in strain rates and flow stress leading to variation in viscosity which affects the flow. However, comprehensive three-dimensional heat transfer and material flow calculations considering spatially variable thermophysical properties and non-Newtonian viscosity have been reported for the FSW of aluminium alloys, steels and Ti–6Al–4V [39–41]. It was found that the typical peak strain rate during FSW reaches about 100 s$^{-1}$ at

![Fig. 12. Peak temperature distribution along the intersection of horizontal midplane and transverse plane through the tool-axis, for the FSW of 304L stainless steel using a tool rotational speed of 600 rpm and a welding speed of 10.2 mm min$^{-1}$. The tool shoulder and tool pin diameters were 25.4 and 5.6 mm, respectively [4].](image)
locations where the velocity gradient is highest, such as near the shoulder edge at the work-piece surface and at the pin surface at lower elevations. The strain rate drops sharply to about 30 s\(^{-1}\) a few mm below the top surface. The strain rate decreases rapidly with depth due to a significant decrease in velocities through viscous dissipation. The computed strain rates were comparable with the 20 s\(^{-1}\) value estimated based on measured grain-size and calculated peak temperature from a thermal model reported by Frigaard et al. [9].

The computed temperature profiles along the longitudinal and transverse sections through tool-axis and at the top surface of Ti–6Al–4V work-piece are shown in Fig. 13a–c, respectively. The temperature profiles on the longitudinal mid-section (Fig. 13a) and on the top surface of the work-piece (Fig. 13c) are compressed in front of the tool and expanded behind it. The computed results are consistent with the fact that heat is supplied rapidly to the cold region of the work-piece ahead of the tool while heat is transported at a slower rate to material already preheated behind the tool.

Peak temperatures in the work-piece are attained close to the edge of the tool shoulder and significant spatial gradients of temperature exist in the vicinity of the tool surfaces. Measurements of temperatures close to a rotating tool are difficult for two reasons. The material transport caused by the motion of the tool makes it difficult to focus on a single location; any embedded thermocouple may be displaced due to plastic flow. The strong temperature gradient near the tool means that a small error in thermocouple location can lead to a large error in temperature.

Thermal cycles affect the structure and properties of welded materials. For example, in precipitation-strengthened aluminum alloys, the hardness profile depends greatly on the precipitate distribution and only slightly on the grain-size [100,101]. If a peak temperature greater than 675 K is attained

![Fig. 13. Computed temperature profiles (K) in (a) xz-plane, (b) yz-plane and (c) z-plane, i.e. top surface of 7.2 mm thick work-piece of Ti–6Al–4V alloy welded with translational speed of 1.6 mm s\(^{-1}\) and rotational speed of 275 rpm [72].](image-url)
at any location, complete dissolution and re-precipitation takes place there. The dissolution and growth of the precipitates during the welding thermal cycle may lead to softening. Even when the peak temperature is lower than 675 K, the density of the strengthening precipitate reduces as the peak temperature gets closer to 675 K [100].

Dimensional analysis, a useful tool for understanding a complex situation, can be used to estimate peak temperature in the work-piece using available numerically computed and experimentally measured thermal cycles for different alloys [102]. The non-dimensional peak temperature, $T^*$, and heat input, $Q^*$, are given by

$$T^* = \frac{T_p - T_a}{T_S - T_a} \quad \text{and} \quad Q^* = \frac{f \sigma_b A \omega C_p}{kU^2}$$

where $T_a$, and $T_S$ are the peak, ambient and solidus temperatures of the work-piece, respectively, $f$ represents the fraction of heat that is transported into the work-piece, $\sigma_b$ is the yield strength value at a temperature which is 0.8 times the solidus temperature and $A$ is the surface area of the shoulder. The thermal properties at the mean temperature, defined as $(T_S + T_a)/2$, are used in the calculations. Note that the non-dimensional heat input, $Q^*$, is similar to a recovery factor, or the ratio of actual to theoretical temperature recovery [103] for surface of a body undergoing adiabatic viscous heating:

$$Q^* = \frac{f \sigma_b A \omega C_p}{kU^2}$$

A correlation between dimensionless peak temperature, $T^*$, and dimensionless heat input, $Q^*$, valid for a range of $Q^*$ between $5 \times 10^3$ and $5 \times 10^5$, is obtained from experimentally measured and computed results (Fig. 14):

$$T^* = 0.131 \log(Q^*) + 0.196$$

In principle, the yield stress of the material drops dramatically as the solidus temperature is approached. If the yield stress decreases, so does the rate of heat generation, reaching zero in the limiting case of melting. The decrease in temperature allows the material to recover its strength, thus permitting a steady temperature to be established below the solidus [104]. However, in practice it is possible that localised melting can occur under non-steady conditions. The unsteady condition is
possible because heat generation is instantaneous while heat transfer takes time. Incipient melting has been reported in the nugget zone at high travel speeds in an aluminium alloy (AA 7100) [105], and locally melted films were observed during the FSW of cast magnesium alloy AZ91 [106].

Cooling rate determines the precipitation sequence in age hardenable alloys. It also affects the grain-size. In fact, friction-stir processing, combined with large cooling rates obtained at higher rotational speed, has been found to be useful for grain refinement and has been suggested as a technique for producing bulk nano-crystalline materials [107].

In high carbon steels, for example S70C (0.72 wt.% C), FSW can result in martensite formation [108]. To keep the volume fraction of martensite low, either the peak temperature should be lower than $A_1$, critical temperature or the cooling rate should be slower than the critical cooling rate. The peak temperature decreases with increase in welding speed or decrease in rotational speed. The cooling rate decreases with decrease in welding speed and with decrease in peak temperature. Low welding speed combined with low rotational speed produce slow cooling rate and prevent martensite formation [109].

2.5. Welding variables

The welding speed, the tool rotational speed, the vertical pressure on the tool, the tilt angle of the tool and the tool design are the main independent variables that are used to control the FSW process. The heat generation rate, temperature field, cooling rate, $x$-direction force, torque, and the power depend on these variables. The effects of several of the independent variables on the peak temperature have been discussed in the previous section. In short, peak temperature increases with increasing rotational speed and decreases slightly with welding speed. Peak temperature also increases with increase in the axial pressure. Fig. 15 shows significant increase in peak temperature with increase in rotational speed.

During FSW, the torque depends on several variables such as the applied vertical pressure, tool design, the tilt angle, local shear stress at the tool material interface, the friction coefficient and the extent of slip between the tool and the material. Measured torque values can provide some idea about the average flow stress near the tool and the extent of slip between the tool and the work-piece for certain conditions of welding, when other variables are kept constant.

The torque decreases with increase in the tool rotation speed due to increase in the heat generation rate and temperature when other variables are kept constant. It becomes easier for the material to flow at high temperatures and strain rates. However, torque is not significantly affected by the change in welding speed as shown in Fig. 16. The relative velocity between the tool and the material is influenced mainly by the rotational speed. Therefore, the heat generation rate is not significantly affected by the welding speed. High traverse speeds tend to reduce heat input and temperatures. The torque

![Fig. 15. Relationship between rotational speed and peak temperature in FS-welds of AA 6063 [110].](image-url)
increases only slightly with the increase in traverse speed because material flow becomes somewhat more difficult at slightly lower temperatures. The torque on the tool can be used to calculate the power required from \( P = \omega M \), where \( M \) is the total torque on the tool.

Excessive \( x \)-direction force can be an important indicator of potential for tool erosion and, in extreme cases, tool breakage. Axial pressure also affects the quality of the weld. Very high pressures lead to overheating and thinning of the joint while very low pressures lead to insufficient heating and voids. Power requirement also increases with the increase in axial pressure.

2.6. Tool design

Tool design influences heat generation, plastic flow, the power required, and the uniformity of the welded joint. The shoulder generates most of the heat and prevents the plasticized material from escaping from the work-piece, while both the shoulder and the tool pin affect the material flow. In recent years several new features have been introduced in the design of tools. Several tools designed at TWI are shown in Table 2. The Whorl and MX-Triflute have smaller pin volumes than the tools with cylindrical pins [112,113]. The tapered threads in the whorl design induce a vertical component of velocity that facilitates plastic flow. The flute in the MX-Triflute also increases the interfacial area between tool and the work-piece, leading to increased heat generation rates, softening and flow of material. Consequently, more intense stirring reduces both the traversing force for the forward tool motion and the welding torque [112,113].

Although cylindrical, Whorl and Triflute designs are suitable for butt welding, they are not useful for lap welding, where excessive thinning of the upper plate can occur together with the trapping of adherent oxide between the overlapping surfaces. Flared-Triflute and A-skew tools were developed to ensure fragmentation of the interfacial oxide layer and a wider weld than is usual for butt welding [114]. The Flared-Triflute tool is similar to MX-Triflute with an expanded flute, while A-skew\textsuperscript{™} tool is a threaded tapered tool with its axis inclined to that of the machine spindle. Both of these tools increase the swept volume relative to that of the pin, thus expanding the stir region and resulting in a wider weld and successful lap joints.

Motion due to rotation and translation of the tool induces asymmetry in the material flow and heating across the tool pin. It has been demonstrated that during FSW, material flows primarily on the retreating side [40,41,74,49,77]. To overcome this problem, TWI devised a new tool, Re-stir, which applies periodic reversal of tool rotation. This reversal of rotation eliminates most problems associated with inherent asymmetry of conventional FSW.

With the exception of FSW with Re-stir tool, material flow is essentially asymmetric about joint interface. Understanding the asymmetry in material flow is important for optimal tool design.
Colegrove and Shercliff [44,47] calculated material flow for the FSW of AA 7075 using computational fluid dynamics. They compared the mechanical efficiency of a triangular tool with convex surfaces (Trivex) with that of a conventional tool (Triflute) by examining the streamlines around these tools. They suggested that the Triflute tool produced a strong auguring action, thus increasing the downward force (Fig. 17).

Zhao et al. [115] studied the effect of tool pin design on the weldability and mechanical properties of welded 2014 Al plates. Cylindrical and tapered tool pins did not ensure effective mixing in the vertical direction leading to wormholes at the base of the TMAZ. However, when tapered tools with threads were used, defect free welds were obtained. Other studies have also confirmed that tools with screw threads generate more heat and improve flow of the softer material by exerting a downward force. Since the material flows mainly on the retreating side, insufficient plasticity and material flow results in a “wormhole” on the advancing side [36,76]. This effect becomes more prominent at low temperatures due to sluggish flow of materials. The choice of pin angle, which is the angle between the conical surface of the pin and its axis, is another important parameter which influences the FSW process; increasing the angle leads to a more uniform temperature distribution along the vertical direction, which helps in reducing distortion [68]. An angle of $\sim 40^\circ$ is thought to be optimum for 7xxx aluminium alloys [68].

Design of tools based on a quantitative understanding of material flow is just beginning. Colegrove and Shercliff [44,47] used a thermal model based on FLUENT to design a tool to minimise the traversing force during FSW of aluminium alloy. They examined several tool geometries including Trivex, which is triangular in shape with convex surfaces, and MX-Trivex which had similar shape with threads. They observed that the traversing force and downward forces were considerably lower for the Trivex tools relative to those for Triflute, especially at lower applied shear stress where considerable slip occurs between the tool and the work-piece. They suggested that Trivex with its convex surfaces avoids sticking to the material that reduces the shear force at the tool–metal interface and consequently reduces the traversing force. Triflute, on the other hand, has features that impede flow and the tool sticks to the material even at low applied shear stress. The entrapped material in the tool leads to a large shearing effect causing correspondingly greater tool forces. Interestingly,

Table 2
A selection of tools designed at TWI [112,16]

<table>
<thead>
<tr>
<th>Tool</th>
<th>Cylindrical</th>
<th>Whori™</th>
<th>MX triflute™</th>
<th>Flared triflute™</th>
<th>A-skew™</th>
<th>Re-stir™</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tool pin shape</td>
<td>Cylindrical with threads</td>
<td>Tapered with threads</td>
<td>Threaded, tapered with three flutes</td>
<td>Tri-flute with flute ends flared out</td>
<td>Inclined cylindrical with threads</td>
<td>Tapered with threads</td>
</tr>
<tr>
<td>Ratio of pin volume to cylindrical pin volume</td>
<td>1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Swept volume to pin volume ratio</td>
<td>1.1</td>
<td>1.8</td>
<td>2.6</td>
<td>2.6</td>
<td>Depends on pin angle</td>
<td>1.8</td>
</tr>
<tr>
<td>Rotary reversal</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Application</td>
<td>Butt welding; fails in lap welding</td>
<td>Butt welding with lower welding torque</td>
<td>Butt welding with further lower welding torque</td>
<td>Lap welding with lower thinning of upper plate</td>
<td>Lap welding with lower thinning of upper plate</td>
<td>When minimum asymmetry in weld property is desired</td>
</tr>
</tbody>
</table>
they [44,47] observed no change in terms of heat input or power requirement by different tool design.

The quantitative prediction of the forces that the tool experiences is particularly useful in designing tools for hard materials such as steel. Such predictions should ideally include a consideration of plastic flow in the thermal model [39–41]. The errors resulting from ignoring plastic flow for high thermal conductivity alloys are smaller than for low thermal conductivity materials such as steel. The errors are greatest near the tool where an accurate knowledge of temperature is important for the smooth operation of the FSW process.

Tool wear is an important concern for the FSW process, particularly for high temperature harder alloys. Titanium, which has a melting point of 1941 K, has been successfully friction-stir welded using sintered TiC welding tools, which require a water cooling arrangement to extract excessive heat from the tool. Mandal et al. [116] proposed a technique to reduce tool wear. Heat sources were introduced in front of the tool to preheat the work-piece, creating a hot channel for the tool to move in, thus reducing the tool wear. The model was tested using a modified Rosenthal equation for a combination of three heat sources, one corresponding to the tool and two for the preheat ahead of the tool. Though the concept is novel, the computational modelling ignored material flow, the wear on the tool was not quantified and the conclusion about reduction in wear was intuitive. There is a need for the development of reliable models for the wear of the FSW tools.

Fig. 17. Streamlines for isothermal model [47] that used a limiting shear stress of 40 MPa: (a) Triflute tool, (b) single streamline for Triflute tool showing vertical movement and (c) Trivex tool. Tool rotation 457 rpm and translation at 457 mm min⁻¹.
2.7. Defects

Common defects in friction-stir welds include porosity and surface defects. At a constant rotational speed, an increase in the travel speed leads to wormhole initiation near the bottom of the weld. Furthermore, the size of the wormholes increases with the travel speed [117] because of inadequate material flow towards the bottom of the weld. There are indications that the travel speed to rotational speed ratio is an important variable in the formation of the wormhole defect [118]. For the same material and tool geometry, a high ratio tends to favour the formation of wormhole defects [119].

Most of the heat generation occurs at the interface between the tool shoulder and the work-piece. Significant heterogeneity in heat generation at that interface can lead to defect formation in the form of excess flash due to surface overheating [117,120].

The propensity for cracks or voids increases with the welding speed although there is an alloy-dependence [121]. For example, defects dominated in AA 5083-O and AA 2024-T3 but not in AA 6063-T6 in which there is a significant drop in hardness within the TMAZ. The defects tend to occur on the advancing side where an abrupt microstructural transition occurs from the highly refined nugget zone to the TMAZ while the transition was gradual on the relatively defect free retreating side.

Cast Al–12 wt.% Si alloys contain coarse particles of silicon which can be used to reveal the flow and defect evolution during FSW [122]. Insufficient heating causes the brittle Si particles to crumble so that the observation of fine Si particles after FSW is indicative of defect formation due to limited material flow. Therefore, it is not surprising that finer particles were detected on the advancing side near the bottom of the tool pin where inadequate flow is associated with void formation.

It has been suggested that the force on the tool in the x-direction, which can be measured using a dynamometer, could be used to predict defect formation in FSW on the assumption that large forces indicate sluggish flow [123]. Maps showing the fraction of stirred material which is at a given flow stress and temperature, with superimposed strain rate contours (Fig. 18) can also indicate the tendency for defect generation [124]. Referring to Fig. 18, at 80 rpm a greater amount of material is being deformed at a higher strain rate with both the Triflat and Trivex tools than with the cylindrical tool. At 200 rpm, greater amount of material was deformed at moderate temperatures (725–735 K) and strain rates (1–50 s⁻¹) with the Trivex tool which shows largest deformation domain. It was concluded [124] that optimum welds correspond to maps with large regions of moderate to high strain rates and temperatures 30–50 K below the solidus temperature. Higher temperatures will lead to softening of the material and reduce the size of the deformation region.

Tool design and welding variables affect materials flow patterns. However, no specific character of the material flow has been related with the porosity formation and no unified mechanism of porosity formation exists.

Elangovan et al. [125] examined the effects of rotational speed and tool pin design on defect formation in friction-stir processing of AA 2219. Five pin profiles (straight cylindrical, tapered cylindrical, threaded cylindrical, triangular and square) were used to fabricate joints at various tool rotational speeds. The square tool pin profile produced the least defect content in the weld as the flat faces produced a pulsating action which led to more effective stirring. Also, a square tool has higher eccentricity which is defined as the ratio of the volume swept by the tool to the static volume of the tool.

In FSW lap joints of AA5083 and SS400, the size of the voids formed along the weld interface increases with increase in diameter of the tool pin and tool tilt angle [126]. Large diameter pin produces (>5 mm) more heat and forms intermetallic compound FeAl3 and Fe2Al5 instead of FeAl formed at lower temperature. Aluminum rich FeAl3 and Fe2Al5 are harder and more brittle than FeAl and they reduce joint shear strength. Increasing the tool tilt angle (>1°) also increases the heat generation rate, forms aluminum rich precipitates and decreases joint strength [126].

One advantage of the FEM model [36] is its ability to predict void formation because the arbitrary-Lagrangian–Eulerian formulation allows for large material deformation and for the grid to track the material so that separation can occur between the work-piece and tool. Fig. 19 shows void formation at the lower advancing side, near the trailing edge of the pin/work-piece interface, due to incomplete deposition of plastic material.
Fig. 18. Material condition maps for (a) Triflat, (b) Trivex and (c) cylindrical tool pin at 80 and 200 rpm [124].

Fig. 19. Void formation at the lower advancing side due to incomplete filling [36] modeled using ALE formulation of FEM. Temperature contours are shown in °C.
2.8. Residual stress

The presence of residual stress in a weld plate affects its distortion behaviour and ability to sustain applied loads while maintaining structural integrity [127–129]. While compressive stresses can in some circumstances be beneficial [130], tensile stresses can cause crack initiation and aid its propagation leading to catastrophic failure.

Residual stresses in the weld can be measured by using: (1) diffraction studies using X-ray or neutron sources without destroying the weld and (2) destructive hole-drilling methods [127,128,131]. Since diffraction is expensive and conventional hole-drilling is only suited to uniform plane stress around the hole, Ya et al. [132] used Moiré interferometry incremental-hole-drilling method to assess residual stress in a friction-stir weld. Synchrotron diffraction has also been used to measure residual stress in such welds in the context of fatigue behaviour [133].

Using neutron-diffraction, Peel et al. [134] showed that longitudinal stress increases as traverse speed increases due to steeper thermal gradients and reduced time for stress-relaxation. Fig. 20 shows that transverse stresses do not display a direct dependence on the rotational speed. It also shows that the weld zone is under tension and the work-piece material is under compression.

As in ordinary welds, residual stresses develop in constrained assemblies during FSW due to expansion during heating and contraction during cooling; a feature unique to FSW is the additional stress caused by the rotational and translational components of the tool so that the welding parameters of FSW must affect the final state of stress [135]. The stirring action of the tool is believed to relieve some of the stresses within the thermomechanically affected zone [136]. A finite element analysis of clamped FS-welded AA 6061 samples has shown that lateral and longitudinal residual stresses decrease with an increase in the rotational speed [135]. A higher welding speed enhances the
longitudinal residual stress but reduces it along the lateral direction. The analysis also showed that the maximum temperature gradients in the sample are located just beyond the edge of the tool shoulder. As might be expected, the residual stress distribution is dramatically altered on unclamping the samples after friction-stir welding and this must be taken into account in any modeling effort.

The role of plasticity during the friction-stir welding process is known to be important in the calculation of residual stress; a large overestimation of the magnitudes of the residual stress [136] can result when this effect is ignored. Nevertheless, the outcome that the longitudinal stress along the weld centreline is tensile and larger than all the other component stresses is believed to be correct [138]. Fig. 21 shows some detail at a point where the FSW tool is halfway along the length [137]; the longitudinal stress is lower in front of the tool (marked by A) as it has not been affected by thermal stress or by structural loading. Behind the tool, marked by D, compressive stresses exist. Transverse stress shows similar trend though it is spread over a very larger area behind the tool compared to that in longitudinal stress as the end of work-piece right behind the tool is not constrained leading to free thermal expansion. The stress in the vertical direction is negligible along the edges of the work-piece (region B) while it is compressive in the remaining regions.

Fig. 21. Normal stresses in MPa in (a) the longitudinal x-direction, (b) the lateral y-direction and (c) the vertical z-direction when the tool is mid-way between the two ends of AA 6061 weld plates [137].
Tensile stresses present in FS-welded samples lead to poor mechanical properties. Applying external tensioning during welding [139] induces compressive stresses which have the benefit of inhibiting crack propagation. Models predict that increasing tensioning levels to values higher than 50% of the room temperature yield stress in AA2024 aluminium alloys, leads to tension in the weld being replaced by desirable compressive stresses [140]. This is an important finding and can be used by welders as a guiding principle to produce high-quality welds. Not surprisingly, crack propagation rates are known to correlate strongly with the state of stress [141].

There has been systematic work on the resistance to fatigue of friction-stir welds in aluminium and titanium alloys [142], which demonstrates the key role of residual stresses in controlling crack growth within the HAZ along the welding direction [143]. Although the levels of residual stress can be smaller than in conventional welding scenarios, they have a large effect on near-threshold crack propagation rates. In AA 7050-T7451, fatigue crack propagation in the weld nugget is inferior, while it is superior to that of base alloy in the HAZ [143]. In HAZ, compressive residual stresses dominate crack growth [144]. For the weld nugget region, microstructure and intergranular failure mechanism dominate crack growth. In precipitation hardened aluminium alloys the particles in the HAZ are much coarser and less coherent than the base material. This should in general reduce the fatigue crack growth rate in the HAZ relative to the base material, but only if tensile residual stresses are not present within the HAZ [142]. Post-weld solution heat-treatment and aging also restores the strengths of the nugget to the same level as the base material [83]. In AA 7050 grain boundaries are sensitized during FSW [145], leading to intergranular fracture when fatigued in a 3.5% NaCl solution.

3. Microstructure and properties of friction-stir welded alloys

3.1. Aluminium alloys

Many aluminium alloys are strong by virtue of precipitation hardening through natural or artificial ageing from the solution-treated condition. The heat associated with welding changes the microstructure of the material. In Fig. 22, HV\textsubscript{min} and HV\textsubscript{max} represent the hardness in the solution-treated and precipitation hardened states. The effect of welding is to cause a drop in hardness from HV\textsubscript{max} towards HV\textsubscript{min} as the peak temperature experienced increases, curve (a), Fig. 22. This is because precipitates will coarsen and reduce in number density in regions remote from the heat source, and will re-enter solution when the peak temperature is sufficiently high [100]. Some re-precipitation may occur during the cooling part of the thermal cycle, resulting in a hardness value beyond HV\textsubscript{min}, curve (b), Fig. 22. The

![Fig. 22](image-url) A schematic diagram showing dissolution and re-precipitation in age-hardenable aluminium alloys. HV denotes the Vickers hardness number. Adapted from [146].
ultimate result is the continuous line with a minimum in hardness somewhere in the heat-affected zone, due to the competing effects of dissolution and re-precipitation [146]. But in contrast to age hardenable AA 6082, where a minimum hardness occurs in the HAZ, FSW of non-hardenable AA 5082 results in uniform hardness across the weld [147].

This general scenario may be complicated by the effects of deformation in FSW as described for the specific example of AA 2219. AA 2219 is a copper precipitation-strengthened alloy containing about 6.3 wt.% Cu, which because of its strength and toughness at low temperatures, is used for containing liquefied gases for rockets of various kinds. It is frequently supplied in the T87 condition, meaning that it has been solution-treated, cold-worked (10% reduction in rolling) and artificially aged. It can be welded using arc processes but this results in a reduction in the cross-weld strength because the proof strength of the fusion zone decreases to about 140 MPa compared with the 370 MPa of the plate [148]. The former can be increased to between 220 and 275 MPa using pulsed GTAW or pulsed electron beam welding techniques because this promotes finer grains in the fusion zone [148].

Friction-stir welding does not seem to have an advantage over arc welding with respect to the strength of the fusion zone, as illustrated in Fig. 23. Notice that the TMAZ is somewhat softer than the fusion zone because the latter dynamically recrystallises into a fine grain structure. It is the coarsening of the Al2Cu precipitates in the TMAZ that is partly responsible for its softening [150]. Some transmission electron micrographs across the weld are illustrated in Fig. 24; these show clearly the huge changes due to the heat from the process. The formation of coarse precipitates at the grain boundaries, and their associated precipitate-free zones, are common detrimental features in the microstructure. The fact that FSW welds in precipitation hardened alloys lead to a weak zone is not surprising given that the majority of strengthening in most strong alloys comes from precipitates [151]. There is some evidence that manipulation of pin profiles and FSW parameters may help improve slightly, the hardness in the central region [125] or indeed, in the HAZ [152]. Theoretical work has also been done to see if cryogenic cooling after the tool pass can help retain alloying elements in solution after the peak temperature is experienced, so that the alloy can then naturally age and not develop the coarsened microstructures typical of slow cooling from the peak temperature [153]. However, computations indicate that the advantage in so doing is likely to be minimal [153].

Fig. 25a shows that the grain-size increases with increase in peak temperature caused by increase in rotational speed. Here grain-size is related to peak temperature by assuming static grain-growth of dynamically recrystallised grains, during the cooling of the thermal cycle [110]:

$$D^2 = D_0^2 + Ate^{-Q/RT_p}$$

(23)

where $D_0$ and $D$ are the initial and recrystallized grain-size, $A$ is a constant, $t$ is the time to cool to 448 K, $Q$ is the activation energy for grain-growth, $T_p$ is the peak temperature. Assumption of isothermal condition leads to over-prediction of grain-sizes [110]. Precipitate free zones form near the grain boundaries because grain boundaries act as sinks for nearby dislocations, reducing nucleation sites for

---

**Fig. 23.** Open circles are data from friction-stir weld, plotted relative to the centreline of the FSW for alloy AA2219 [149]; similar data have been reported by Paglia and Buchheit [150]. Filled circles from a variety of arc and electron-beam welds [148]. In the latter case all values are plotted at zero on the horizontal axis to represent the fusion zone.
precipitates and also as precipitation sites, effectively reducing the solute content around them. As grain-size increase, assuming constant width of PFZs, their volume fraction decreases with increase peak temperature as shown in Fig. 25b [110].

Not all alloys of aluminium are precipitation hardened. In the 2000 series alloys, the strength depends more on grain-size \((d)\), which has been expressed in terms of the Zener–Holloman parameter [154]:

\[
\log d = a + b \log Z
\]

Fig. 24. Microstructure variations across aluminium alloy AA2219 joined using friction-stir welding. (a) Unaffected parent material in the precipitation hardened condition; (b) the nugget; (c) the heat-affected zone. Micrographs courtesy of Dr. Christian Paglia.

Fig. 25. Relationship between (a) peak temperature and grain-size in AA 6063 [110] and (b) grain-size and volume fraction of PFZs.
where $a$ and $b$ are empirical constants based on data from extrusion experiments [155] and the hardness is then related to $d$ using a form typical of the Hall–Petch type equation:

$$HV = HV_0 + c/\sqrt{d}$$  \hspace{1cm} (25)

where $c$ is a constant.

In the cast Al–Si alloys, friction-stir welding breaks up the large silicon particles in the nugget and the TMAZ, Fig. 26; as a consequence, the fracture is located in the base plate during cross-weld tensile tests because in this case, it is the coarse silicon particles which control failure [156,5]. FSW can also heal casting defects such as porosity [157].

---

Fig. 26. (a) A section through a friction-stir weld made in an Al–Si casting alloy. There are pores indicated in the base metal (BM). HAZ represents the heat-affected zone, TMAZ the thermomechanically affected zone, and SN the stir nugget. Optical micrographs showing the microstructure in (b) the base metal; (c) heat-affected zone; (d) the thermomechanically affected zone, where considerable refinement of the silicon has occurred. Optical micrographs of regions (e–g) of the stir nugget as identified in macroscopic section presented above. Micrographs courtesy of Professor H. Fujii.
Corrosion studies indicate that the weld zones produced by friction-stir welding have comparable environmentally assisted cracking susceptibility as the unaffected parent [150].

3.2. Magnesium alloys

Magnesium alloys, normally produced by casting, may find significant applications in the automotive and aerospace industries with rapid growth particularly in die-cast vehicle components because of their better mass-equivalent properties [158–161]. They are used for light-weight parts which operate at high speeds. The motivation for using FSW for magnesium alloys is that arc welding results in large volumes of non-toxic fumes [162]. On the other hand, solid-state FSW does not result in solute loss by evaporation or segregation during solidification, resulting in homogeneous distribution of solutes in the weld [163]. Also, many magnesium alloys in the cast condition contain porosity which can be healed during FSW.

The hardness and strength can be retained after friction-stir welding [164]. There is no significant precipitation hardening in the alloy studied (AZ31, \(\approx\) Mg–3Al–1Zn wt.% wrought) and the net variation in hardness over the entire joint was within the range 45–65 HV, with the lower value corresponding to the base plate [164–166]. In the same system, a higher starting hardness of 70 HV leads to a substantially lower hardness in the nugget (50–60 HV); the variations in hardness appear to be consistent with measured variations in grain-size in accordance with the form of the Hall–Petch relationship [165,167]. The grains in both the nugget and TMAZ tend to be in a recrystallised form, and tend to be finer when the net heat input is smaller (for example at higher welding speeds).

In Mg–Zr alloys with Zr-containing particles, FSW leads to a considerable refinement of the grain structure and sound welds can be produced in thin sheets over a wide range of welding conditions; for sheets thicker than about 3 mm, the welds contained defects associated with an inability to supply sufficient heat during welding [168]. There are, however, contradictory results [169] showing successful welds in 6 mm thick Mg–Zn–Y–Zr plates so it is unlikely that these results are generic to magnesium alloys.

3.3. Copper alloys

Copper which has much higher thermal diffusivity than steel cannot easily be welded by conventional fusion welding techniques. Heat input required for copper is much higher than conventional FSW because of the greater dissipation of heat through the work-piece. Still, FSW has been successfully used to weld very thick (50 mm) thick copper canisters for containment of nuclear waste [170]. FSW in copper alloys have all the typical zones found in other materials: the nugget, TMAZ, HAZ and base structure. The nugget has equiaxed recrystallised small grains and its hardness may be higher or lower than the base material depending on the grain-size of the base metal. When 4 mm thick copper plates with average grain-size of 210 \(\mu\)m were welded at high rpm (1250) and low welding speed (1.01 \(\text{mm s}^{-1}\)) [171], nugget had lower hardness (60–90 HV), compared to base metal (105–110 HV). Even though grain-size decreased from 210 to 100 \(\mu\)m, hardness decreased slightly due to reduction in dislocation density relative to base metal [171]. Similar decrease in dislocation density in the nugget zone compared to parent metal has been observed for AA 7075 [11] and AA 6061 [12].

On the other hand, when 2 mm thick copper plates with average grain-size of 30 \(\mu\)m were welded at 1000 rpm and 0.5 \(\text{mm s}^{-1}\) low welding speed, nugget (128–136 HV) was harder than the base metal (106–111 HV) due to reduction in average grain-size to 11 \(\mu\)m [172]. Flores et al. have also shown that as-cast AA 7073 showed that weld nugget was harder than base metal while the 50% cold-rolled alloy showed reduced hardness in the nugget [173].

3.4. Titanium alloys

By far the most dominant of titanium alloys is Ti–6Al–4V, which in its commercial condition has a mixed microstructure consisting of hexagonal-close packed \(\alpha\) and body-centred cubic \(\beta\) phases, which is the stable phase at high temperatures. The usual microstructure is produced by cooling from the \(\alpha\) phase field to produce Widmanstätten \(\beta\) with about 50% of the \(\alpha\) being retained in the microstructure.
This alloy, which accounts for about half of all the titanium that is produced, is popular because of its strength (1100 MPa), creep resistance at 300 °C, fatigue resistance and castability.

Friction-stir welding must clearly disrupt the base microstructure both through the thermal and deformation components of the process [174] but the consequences of this on performance during fabrication and service need investigation. General investigations on fatigue performance indicate that the crack growth rate in the HAZ can be higher or lower than the base material depending on specimen geometry, microstructure and residual stress levels [142].

Experiments have also been conducted on a fully β-titanium alloy in thin sheet form, primarily to prove that the crystallographic texture observed corresponds to one generated by shear deformation [175], consistent with similar observations in aluminium alloys.

Pure titanium in its hexagonal close-packed α-form is interesting because there is also a tendency for deformation by mechanical twinning during friction-stir welding [6]. The nugget region of an FSW joint is found to contain a large density of dislocations and mechanical twins, with transmission microscopy showing an elongated fine-structure, but the overall grain shape seem to remain equiaxed on the scale of optical microscopy. It is speculated that recrystallisation must have occurred during welding but was followed by a small amount of plastic deformation. The HAZ simply revealed grain-growth, a consequential lower hardness, and hence was the location of fracture in cross-weld tests. There was no clearly defined TMAZ as is typical in aluminium FS-welds.

3.5. Steels

The friction-stir welding of steels has not progressed as rapidly as for aluminium for important reasons. First, the material from which the tool is made has to survive much more strenuous conditions because of the strength of steel. Second, there are also numerous ways in which steel can be satisfactorily and reliably welded. Third, the consequences of phase transformations accompanying FSW have not been studied in sufficient depth. Finally, the variety of steels available is much larger than for any other alloy system, requiring considerable experiments to optimise the weld for a required set of properties. Early optimism [176,177] that FSW will become a commercially attractive method for the fabrication of ships, pipes, trucks, railway wagons and hot plate has not yet come to fruition. That the application of FSW to steels is premature is emphasised by the fact that with few exceptions, only elementary mechanical properties have been characterised; most reports are limited to simple bend, tensile and hardness tests. For serious structural applications of the type proposed above it would be necessary to assess fracture toughness and other complex properties in greater depth. There are indications that elongation suffers following FSW [178].

A typical temperature profile behind a friction-stir weld on steel is illustrated in Fig. 27 [176]. The maximum temperature reached is less than 1200 °C [176,48] and the time $\Delta t_{8.5}$ taken to cool over the range 800–500 °C is about 11 s. This is comparable to for example, $\Delta t_{8.5}$ for a manual metal arc weld

![Temperature profile](image-url)
with a heat input of about 1.3 kJ mm\(^{-1}\). Therefore, the metallurgical transformations expected on the basis of cooling rates alone are not expected to be remarkably different from ordinary welds.

However, there are certain major qualifications to this statement. Since the peak temperatures reached are much lower than in arc welds, the part of the HAZ which becomes fully austenitic will be smaller in dimension and with a finer austenite grain-size. The latter helps avoid detrimental transformations, for example to untempered martensite, by reducing the hardenability of the steel. It should therefore be possible to weld high carbon-equivalent steels using FSW than would be possible for arc welding processes. Indeed, examination of the data for the mild steel designated Alloy 2 in Table 3 shows that the hardness of the HAZ is well below that expected for a fully martensitic microstructure in this alloy, estimated at about 335 °C using the method described in Ref. [179].

The TMAZ should become fully austenitic at some stage of its thermomechanical history. It is likely that the severe deformation in this zone causes the austenite to recrystallise, perhaps repeatedly, prior to its transformation during subsequent cooling. This can result in a finer and consequently somewhat harder microstructure than the HAZ, consistent with the hardness behaviour listed for Alloy 2 in Table 3.

The situation for the FSW ferritic stainless steel (Alloy 1, Table 3) is different, in that the heat-affected zone is harder than the TMAZ. The reason for this is not clear from the limited microstructural evidence presented in [176].

### 3.5.1. Role of thermomechanical processing

The microstructure of the thermomechanically processed zone shows trends which are similar to those observed in the control-rolling of steels in which the austenite recrystallises prior to transformation, Fig. 28. That the austenite recrystallises is obvious to see from the equiaxed prior austenite grain structures seen in the TMAZ [48]. In these circumstances, the only manifestation of the deformation component of the processing is in the recrystallised austenite grain-size, which may be finer in those regions which experienced greater strains and strain rates [48]. However, the differences in the sizes of these grains are not very large. As far as the authors are aware, there have been no reports of a ‘pan-caked’ austenite grain structure of the type typical in many control-rolled steels [184].

There would be two significant consequences if the austenite was left in its deformed state prior to transformation. Deformation leads to a large increase in the amount of austenite grain surface and grain edge per unit volume [185–191,18]. This would lead to a corresponding large increase in the grain boundary nucleation rate of allotriomorphic ferrite. The austenite grain boundaries also become

### Table 3

Compositions of steels (wt.%) that have been studied using FSW

<table>
<thead>
<tr>
<th>Alloy</th>
<th>C</th>
<th>Si</th>
<th>Mn</th>
<th>Ni</th>
<th>Mo</th>
<th>Cr</th>
<th>V</th>
<th>Others</th>
<th>Hardness (HV)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.38</td>
<td>1.09</td>
<td>0.55</td>
<td>0.03</td>
<td>11.2</td>
<td>0.01</td>
<td>Nb 0.27</td>
<td>158</td>
<td>280</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.16</td>
<td>0.69</td>
<td>0.08</td>
<td>0.01</td>
<td>0.06</td>
<td>–</td>
<td>Ti 0.04</td>
<td>131</td>
<td>149</td>
</tr>
<tr>
<td>3</td>
<td>0.002</td>
<td>0.10</td>
<td>0.29</td>
<td>0.50</td>
<td>125</td>
<td>230 (500)</td>
<td></td>
<td></td>
<td>110</td>
<td>130 (160)</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
<td>0.21</td>
<td>0.69</td>
<td>Cu 0.01</td>
<td>155</td>
<td>280 (360)</td>
<td></td>
<td></td>
<td>200</td>
<td>360 (520)</td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.20</td>
<td>0.70</td>
<td>Cu 0.01</td>
<td>135</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.32</td>
<td>0.35</td>
<td>1.20</td>
<td>0.20</td>
<td>0.65</td>
<td>1.30</td>
<td>0.05</td>
<td>B 0.0025</td>
<td>250</td>
<td>300</td>
</tr>
<tr>
<td>7</td>
<td>0.32</td>
<td>0.35</td>
<td>1.20</td>
<td>0.20</td>
<td>0.65</td>
<td>1.30</td>
<td>0.05</td>
<td>B 0.0025</td>
<td>250</td>
<td>300</td>
</tr>
</tbody>
</table>

The Vickers hardness values represent approximate mean values within the zone concerned; more detail can be found in the original references. Peak hardness values are reported in brackets. In some cases the compositions are assumed from the published specification range.
rough during deformation and this adds to their potency as heterogeneous nucleation sites; shear bands are introduced which are additional locations for ferrite formation. In summary, the formation of ferrite should be greatly accelerated by transformation from deformed austenite. At the same time, displacive transformation such as Widmanstätten ferrite, acicular ferrite, bainite and martensite should be retarded by a phenomenon known as mechanical stabilization. This is because these transformations are propagated by the displacement of glissile interfaces which are hindered by the presence of deformation debris in the parent austenite.

The consequence of both these effects is to promote diffusional transformations (allotriomorphic ferrite, pearlite) over and above the displacive products. It would be interesting to conduct systematic experiments to reveal such effects.

The recrystallised austenite grains in the stir zone typically tend to transform into coarse microstructures such as allotriomorphic ferrite and Widmanstätten ferrite, resulting in Charpy impact-toughness values which are reasonable for some requirements, but not impressive. Fig. 29 shows toughness data for the stir zone of the weld, all of which meet a minimum requirement of 27 J at −18 °C; a later study revealed larger values of toughness but those data are on half-size Charpy specimens and hence not suitable for direct comparison. The Charpy test itself is recognised to be an empirical measure of toughness, of use in quality control but not as an aid to design. The mechanism of fracture is sensitive to sample dimensions – the effects of plastic constraints are well-documented. Data generated from identical specimens might however be useful in making comparisons between different welding scenarios; thus, Konkol and Mruczek, using sub-sized Charpy specimens have demonstrated that the toughness obtained from FSW welds on a particular steel is similar to that from submerged-arc welds. However, the particular submerged arc weld data

Fig. 28. Typical microstructure of an FSW weld in a C–Mn steel (micrographs courtesy of Threadgill and Johnson). (a) Parent steel, with bands of allotriomorphic ferrite and pearlite; (b) intercritically heated region of HAZ; (c) TMAZ (fine-grained region near stir zone); (d) stir zone.
are not representative of the strength and flux used; typical all-weld metal values for full sized Charpy samples tested at −30 °C using SAW would be in the range 70–130 J depending on the choice of wire and required weld metal strength. There have been reports of significant scatter in Charpy test data, the cause of which is not yet established [201,177]. What is also lacking, is a clear indication of major advantages associated with friction-stir welding of steel compared with conventional processing, which would give the FSW process a lead in the context of steels.

3.5.2. Corrosion

Given that the TMAZ of steel welds does not contain a grossly deformed microstructure, there should be no detrimental corrosion property associated with friction-stir welding. There may even be an advantage since the chemical composition of the weld region is identical to that of the plates, i.e. a homogeneous electrochemical potential. For example, salt spray corrosion tests showed no significant differences in weight loss or the tendency to pit between the weld and plate in friction-stir joined HSLA65 [181]. The problem is relatively severe with precipitation hardened aluminium alloys where sensitisation of the heat-affected zone following friction-stir welding increases corrosion susceptibility [202]. Also, corrosion pits may form [203], shortening the fatigue crack initiation life and decreasing the fatigue crack initiation threshold, with fatigue cracks always initiating from these pits.

3.5.3. FSW of martensite in mild steel

Friction-stir welding has been done on mild steels with an untempered martensitic microstructure, and with the latter following a 35% cold-rolling deformation (Alloy 11, Table 3, [182]), using a variety of tool rotation speeds.

In all cases, FSW causes a large reduction in hardness within the stir zone Fig. 30. A slower tool rotation speed leads to a lesser reduction and a narrower weld. Furthermore, the final grain-size in the stir zone is always coarser and more isotropic than that in the ultrafine starting microstructure.

The research discussed above is in the context of initial microstructures which are experimental (as-quenched martensite and rolled martensite). Of interest is the friction-stir welding of classical quenched and tempered martensitic steels. These show the same general behaviour, i.e. softening in the HAZ and stir zone relative to the initial hard state [204,205]. The microstructure illustrated in Fig. 31 correlates with the hardness data; considerable softening occurs in the HAZ, presumably because the cooling rate is not sufficient to reproduce a fully martensitic microstructure. The hardness within the stir zone is strongly varies violently because of the presence of bands which are
predominantly martensitic and others which are not. Neither the impact toughness nor the ductility obtained was adequate.

3.5.4. FSW of steels at $T < Ac_1$

Not all steels undergo austenitisation during FSW. Alloys which are lean in carbon tend to have a high $Ac_1$ temperature. Furthermore, because they tend to be mechanically weak, the temperatures reached during friction-stir welding are not particularly high.

Experiments on interstitial-free steels (Alloy 3, Table 3) friction-stir welded in the ferritic condition show that the scale of the final microstructure in the TMAZ depends on that of the starting microstructure. Originally coarse grains become much finer in the TMAZ. However, ultrafine grains in a strong initial microstructure become coarser following FSW; this is because the heat generated is greater when a stronger substrate is welded [122]. This can be verified from the model, as both $\tau$ and $p$, in Eq. (2) are larger for harder work-piece material. This is illustrated in Table 4; the results show that FSW will not be useful in the welding of materials with a grain-size of less than $\approx 1 \mu m$, if the objective is to preserve or refine the scale of the microstructure. Some of the fine-grained structures for which FSW has been attempted show grain coarsening in the HAZ with a loss of properties [206,207].

---

$Ac_1$ represents the lowest temperature at which austenite begins to form during heating at a specific rate – the corresponding equilibrium temperature is designated $Ae_1$. 
3.5.5. Hardness of stir zone

The data on the hardness of the stir zone as listed in Table 3, have been analysed here as a function of the IIW carbon-equivalent of the steel, where the latter is defined as

$$CE = C + \frac{Mn + Si}{6} + \frac{Ni + Cu}{15} + \frac{Cr + Mo + V}{5}$$  \hspace{1cm} (26)$$

where the element symbols refer to their concentrations in weight percent. Fig. 32 shows that there is a strong correlation of hardness versus the calculated carbon-equivalent. The stainless steel (Alloy 1, Table 3) is the red point, but with the Cr concentration set to zero. This is because the IIW equation is not appropriate for stainless steels; it is remarkable that when the Cr is neglected, the hardness falls on the same trend as the mild steels.
The highest of hardness values correspond to microstructures which contain substantial quantities of martensite and bainite in carbon-containing steels (see for example, [183]).

4. Outlook and remarks

Friction-stir welding technology has been a major boon to industry advanced since its inception. In spite of its short history, it has found widespread applications in diverse industries. Hard materials such as steel and other important engineering alloys can now be welded efficiently using this process. Significant progress has also been made in the fundamental understanding of both the welding process and the structure and properties of the welded joints. The understanding has been useful in reducing defects and improving uniformity of weld properties and, at the same time, expanding the applicability of FSW to new engineering alloys. With better quantitative understanding of the underlying principles of heat transfer, material flow, tool–work-piece contact conditions and effects of various process parameters, efficient tools have been devised. At the current pace of development, FSW is likely to be more widely applied in the future.

Several important key problems and issues remain to be addressed. First, the fundamental knowledge of the FSW process and the knowledge of the evolution of the structure and properties needs to be combined to build intelligent process control models with a goal to achieve, defect free, structurally sound and reliable welds. Tailoring weld structure and properties based on fundamental knowledge still remains an important milestone in FSW. Attainment of this important goal would require new, more reliable and efficient process sub-models and reliable sub-models to describe the evolution of structure and properties of the welded joints. Current FSW process sub-models are complex, time consuming, and cannot be used in real time. Furthermore, they all suffer from lack of reliability of the predicted results because the underlying physics is highly complex and the current phenomenological models do not contain any model component designed to ensure compliance with experimental results. Recent work in the fusion welding suggests that the lack of reliability of the phenomenological models may be contributed, at least to a large extent, by the uncertainty in several input parameters. In FSW, these uncertain parameters include the friction coefficient, the extent of slip between the tool and the work-piece, the heat transfer coefficient at several work-piece surfaces, partitioning of the heat between the work-piece and the tool at the tool–work-piece interface, and the computed values of non-Newtonian viscosity based on the available constitutive models. Current phenomenological models of FSW do not have any built in mechanism to address these uncertainties. This problem

![Fig. 32. The Vickers hardness of the TMAZ as a function of the IIW carbon equivalent of the steel. The >2.5 value of carbon equivalent corresponds to Alloy 1, Table 3. The red point corresponds to Alloy 1 (Table 3) with Cr neglected in calculating the carbon equivalent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image-url)
can be solved by combining a rigorous phenomenological process sub-model with an appropriate multivariable optimisation scheme to determine optimised values of the uncertain variables from a limited volume of experimental data. Such an approach would result in agreement between the phenomenological models and the experimental results with greater degree of certainty [54–57].

However, the lack of reliability is not the only obstacle for tailoring weld attributes based on fundamental scientific principles. An important difficulty is that the existing process models are unidirectional. In other words, the process sub-models require as input welding parameters, thermophysical properties, tool and work-piece geometry and provide, as output, the temperature and velocity fields and the cooling rates at various locations. In contrast, engineers should be able to specify cooling rates, the geometry of the stir zone and/or other attributes of the FSW as input and obtain as output several alternative sets of welding parameters involving combinations of the welding speed, rotational speed, tool dimensions and other variables as output. Unless the models have a reliability component built into them and are bi-directional in nature, their use is likely to be restricted to researchers with only limited use in manufacturing and process control. Since the quantitative knowledge base in FSW is embodied in process, structure and property sub-models, these sub-models must be made more useful to manufacturing. Previous work in fusion welding show that such restructuring of the process, structure and property sub-models is both necessary and achievable [208,209,57]. Success of such an undertaking will ensure availability of practically the entire quantitative knowledge base of FSW to the whole FSW community for the purpose of tailoring FSW weld attributes and fabricating defect free, structurally sound, and reliable welds. Attainment of this milestone is well within the reach of the welding community within the next 10 years.

Acknowledgements

We would like to thank Dr. G.G. Roy for many helpful discussions and his interest in this work. The research was supported by a grant from the Materials Division, Office of Naval Research, Dr. J. Christodoulu and Dr. J. DeLoach, contract monitors. One of the authors, Rituraj Nandan, is thankful to American Welding Society for a fellowship.

References


