CHARACTERISATION AND FATIGUE OF FRICTION STIR WELDING

Aidy Ali* and Omar Suliman Zaroog

Received: Mar 5, 2008; Revised: Jun 3, 2008; Accepted: Jun 4, 2008

Abstract

Previous attempts on the characterisation of Friction Stir Welding (FSW) based on microstructures, hardness, and residual stress distribution have been reviewed. The role of these parameters on fatigue damage of FSW is then discussed. Relevant conclusions have been drawn to demonstrate the current issues and the future research potential of these joints.

Keywords: Friction stir welding (FSW), fatigue

Introduction

Friction stir welding (FSW) is a solid state welding process that has received the worldwide attention, particularly for joining aluminium alloys (Mendez and Eagar, 2001; Schmidt et al., 2001). In FSW a profiled pin attached to a rotating cylindrical shouldered tool is inserted into the joint line between two pieces of material and translated along the joint to complete the joining process. During the joining operation, the rotating tool is forced down into the metal plates and moved relative to them, causing frictional heating and mechanical deformation of the material being welded (Bussu and Irving, 2003). Generally in FSW the action of the tool produces five distinct microstructural zones (Bussu and Irving, 2001), namely the weld N (N), the shoulder contact zone or FA region (FA), the thermomechanically affected zone (TMAZ), the heat affected zone (HAZ), and the unaffected zone or parent plate (PP). Consequently the fatigue strength of FSW joints produced varies for each zone of the welds.

The Characteristic of Friction Stir Welds

Since FSW was invented in late 1991 by The Welding Institute (TWI), Cambridge, United Kingdom, many attempts have been made to investigate the strength of these joints for aircraft applications (Dalle Donne and Biallas, 1999; Ericsson and Sandstrom, 2000; Esparza et al., 2002). Although this process is considered a relatively new welding process, the process has received worldwide attention and many companies are using the technology...
in production, particularly for joining aluminium alloys (Williams, 2001). However, this joint is still in the research stage in aircraft applications. For the past thirteen years, the investigations of the fatigue strength of FSW can be classified in characterisation of macrostructure, microstructure, hardness, residual stress, and fatigue performance which will be presented in the following sections.

Macrostructure and Microstructure Characterisation

There have been numerous observations of microstructures in connection with the FSW of aluminium alloys such as 2024 (Dalle Donne and Biallas, 1999; Booth and Sinclair, 2002; Sutton et al., 2002; Bussu and Irving, 2003), 7075 (Mahoney et al., 1998), 7050 (Jata et al., 2000), 6061 (Murr et al., 1998), 6013 (Heinz et al., 2000), 6063 (Sato and Kokawa, 2001), 1050 (Kwan et al., 2002), 1100 (Flores et al., 1998; Murr et al., 1998), 1080 and 5083 (Sato et al., 2001). Figure 1 illustrates the typical overall FSW structure. The FSW weld zone is V-shaped and widens near the top surface due to the close contact between the shoulder of the tool and the upper surface (Heinz et al., 2000). Regardless of the specific aluminium alloy type, composition, and manufacturing attributes, generally FSW joints have a core that can be defined as N which is located at the centre of the joint.

The N region can be characterised by the appearance of a distinct annular banded (or ‘onion ring’) structure as reported within the literature (Mahoney et al., 1998; Murr et al., 1998; Strombeck et al., 2001; Booth and Sinclair, 2002). The N dimensions and density vary depending on the process parameters (Bussu, 2000; Williams, 2001). The onion rings extend following a three dimensional pattern within the entire N. Recent study showed that no variation in microstructure or hardness has been observed within the rings (Sutton et al., 2002). Within this region, the grain size was fine, typically equal to or smaller than 10 mm (Jata, 2000; Booth and Sinclair, 2002; Bussu and Irving, 2003), which was much finer than the PP microstructure. The intermetallic particle distribution in the N was seen to be refined in relative to the PP, with particle dimensions of the order of 1 - 2 mm in the N as opposed to large clusters of up to 20 mm in the PP microstructure (Booth and Sinclair, 2002).

On top of the N it is also possible to identify a FA region between the N and the top surface on which welding was carried out (Booth and Sinclair, 2002). It was reported that the FA region has a fine equiaxed grain structure up to 10 mm, the same microstructure size as in the N region. It is worth noting that the FA and N region sizes depend on the process parameters. In some FSW joints the FA size is huge and dominant at the centre of the plate joint. Consequently the N region seems to be very small or unclear in the joint (Murr et al., 1998) (Figure 2).
Similarly in some FSW joints when the N size was huge and dominant the FA region might be unclear (Benavides et al., 1999; Bussu and Irving, 2003). Sato et al. (2001) pointed out that the shape of the weld zone depended on the welding parameters and the material used. Dalle Donne and Biallas (1999) showed that with proper FSW tooling and welding parameter control a drop of only 20% compared to the base material values for the joint ultimate strength and S-N fatigue endurance can be achieved. They performed such tests to see the effect of tool and rotation weld speed as shown in Figure 3.

It is evident that the macrostructure of FSW joints depends highly on the process parameters such as pin size, tool speed, and the thickness of the plate (Thomas et al., 2002). Nevertheless the microstructure features of the N and FA have been found to be similar with respect to (a) the fine grain dimension and intermetallic particle distribution, and (b) the very low void/defect fraction that would appear to be conducive to good fatigue life performance (Booth and Sinclair, 2002).

Further out from the weld line is the thermo-mechanically affected zone (TMAZ) (Figure 1). The size of this region is wide on the top welding surface and narrows down throughout the thickness of the plate. Studies have indicated the presence of an elongated grain structure suggesting that severe plastic deformation takes place during welding (Bussu, 2000) (Figure 4). The bending of the grains in the TMAZ regions also suggests that the stirring action of the FSW tool causes the flat grains of the PP metal to be drawn into the weld N zone (Jata et al., 2000).

In this region the grain size is slightly bigger or similar to the PP material. Bussu and Irving (2003) found the grain size of the TMAZ was about 50 - 100 mm compared with the PP that was about 50 mm. However Sato et al. (2001) found the grains in the TMAZ region of 1080 Al alloy had deformed and contained a sub-grain structure. They found that the average grain size of the deformed grains was about 80 mm and the sub-grain size was approximately 1.47 mm. The anomaly of the grain size in the TMAZ depends on the material and process parameters of the FSW joint. As a result, the grain and sub-grain structure size in the TMAZ region will affect the yield and flow stresses of this joint (Sato and Kokawa, 2001; Sato et al., 2001).

Figure 3. Comparison of transverse cross section of an as-welded FA W 2024 T351 with different tool and weld speed (Dalle Donne and Biallas, 1999)

Figure 4. Typical microstructure in TMZ of an Al-Si-Mg FSW on a plane normal to the weld direction (Bussu, 2000)
Next to the TMAZ region is the HAZ. This region is not affected by the mechanical action of the tool, but the mechanical properties are influenced by the heat flow from the weld (Williams, 2001; Thomas et al., 2002). The size of this region varies depending on the process parameters. The grain size in this region is retained (Jata et al., 2000; Sato and Kokawa, 2001; Bussu and Irving, 2003). However the hardness deteriorated in this zone suggested that strengthening precipitates were coarsened by a factor of 5 (Heinz et al., 2000; Jata et al., 2000; Sato and Kokawa, 2001). In contrast, the strengthening precipitates in the N, FA and TMAZ are dissolved even though in these regions they are coarsened considerably (Heinz et al., 2000).

Furthermore, studies observed that precipitates in the HAZ region overage and transform into b precipitates (Heinz et al., 2000). These precipitates were not yet identified but due to their size and the studies published on the precipitation process in 6xxx alloys it was assumed that these precipitates are B (Heinz et al., 2000; Sato and Kokawa, 2001). Although the grain size of the HAZ is retained as in the PP zone, this region behaves differently, mainly due to the heat causing overaging of the strengthening precipitates (Murr et al., 1998; Sato et al., 2001).

**Hardness Characterisation**

There have been numerous attempts on hardness characterisation in connection with the micro-structure in FSW joints. For instance, Booth and Sinclair (2002) found that the hardness values for the top and the bottom surfaces of 2024-T351 FSW alloy were different, especially in the weld region that covers the N, while Heinz et al. (2000) reported for 6013 Al alloy that the top showed slightly higher hardness numbers compared with the bottom surfaces.

Figures 5 and 6 show that the weld zone is considerably softer than the base material. Going from the PP towards the weld line, both traces for bottom and top surfaces exhibit an initial rise from the PP hardness level (HAZ region), which is followed by a gradual decline in hardness to a minimum either in the TMAZ or the HAZ boundary. There is then a steep rise to a central value of intermediate hardness that is located at the N or FA regions.

This hardness profile is not similar for all Al alloy FSW. Jata et al. (2000) reported that for the 7050 Al alloy the hardness of the top side is lower than the bottom side (Figure 6). They suggested that this was due to the fact that this side of the plate was in full contact with the FSW tool shoulder, and thus experienced direct heat from the rapidly rotating tool shoulder. The back side, on the other hand, was in contact with a back plate that acted as a heat sink and rapidly draws away heat.

Such differences in the hardness profiles between the top and the back side of FSW joint plates have also been observed for FSW of Al-Li-Cu alloy by Jata and Rioja (1998). These differences can be attributed to the through thickness variation in the extent of precipitate dissolution in the N region and also the post weld room temperature (Jata et al., 2000).

Comparing the hardness between the different regions or zones in FSW joints, the hardness within the N varies depending on the alloy and its initial heat treatment. For 2024-T351, 7050-7745 and 6061-T6 alloys, hardness profiles in the weld N showed a local maximum value at the plate joint line or center of the N (Murr et al., 1998; Booth and Sinclair, 2002; Bussu and Irving, 2003). For cast 1100 and 5083 Al alloys, the hardness profile is almost uniform or on average the same hardness compared with the PP hardness (Figure 7).

For 6063 Al alloy, the hardness profiles in the weld N showed a minimum value among other regions. These differences in hardness value within the N have been correlated with the size of the precipitates present in the region (Flores et al., 1998; Jata et al., 2000; Sato and Kokawa, 2001; Sato et al., 2001). The 2xxx and 7xxx series showed the hardness minima within the TMAZ region in FSW joints (Rhodes et al., 1997; Dalle Donne and Biallas, 1999; Bussu, 2000; Jata et al., 2000; Bussu and Irving, 2001; Booth and Sinclair, 2002; Bussu and Irving, 2003). This indicates
that complete overaging is responsible for the lowest hardness value (Jata et al., 2000).

However, for 1080 Al alloy the maximum hardness was located in the TMAZ region by Sato et al. (2001). They suggested that the stir zone consisted of recrystallised fine grains, while the TMAZ had a recovered grain structure. They found that there were many small Al$_6$ (MnFe) particles which were detected in all the grains in this region. The hardness was mainly affected by the distribution of small particles in FSW of Al alloys (Sato et al., 2001).

Within the HAZ region, hardness profiles show an increase in the direction of the PP. Hardness maxima and minima have been observed almost next to the PP and next to the TMAZ region respectively in 2024-T351, 7050-T7451 and 6013 (Heinz et al., 2000; Jata, 2000; Jata et al., 2000; Booth and Sinclair, 2002; Bussu and Irving, 2003). A recent study suggested that the hardness variation, when the hardness minima were found within this region, was due to coarsening of the precipitate distribution. No detailed studies have been carried out to date in order to explain hardness maxima in the HAZ (Jata et al., 2000).

**Residual Stress Characterisation**

Residual stress fields are widely believed to have a significant effect on fatigue crack growth. Therefore many studies have been done to investigate the effects that residual stresses have on the fatigue crack growth rate in FSW (Mahoney et al., 1998; Dalle Donne and Biallas, 1999; Ericsson and Sandstrom, 2000; Esparza et al., 2002; Sutton et al., 2002). The residual stress distribution varies between the zones and the type of material (Mahoney et al., 1998; Dalle Donne and Biallas, 1999; Ericsson and Sandstrom, 2000; Esparza et al., 2002; Sutton et al., 2002).
It is worth noting that the longitudinal residual stress reveals an “M”-like shape for most types of material even though it has different levels of residual stress, residual tension, or residual compression (Figure 8). This shape was also found in a fusion welded plate joint by Gatolo and Lanciotti (1997). This shape is formed due to different cooling rates along the FSW joints.

Webster et al. (2001) measured the residual stress using a Synchrotron X-ray technique for AA 7108-T79 and found tensile residual stress in the N zone. This phenomenon was also reported by Bussu and Irving (2003) and Oosterkamp et al. (2000) for AA7108-T79 and for 2024-T351, respectively. Nevertheless, Jata et al. (2000) and Dalle Donne et al. (2001) found a small compressive residual stress located at the plate joint line (PJL) in the N zone for 7050-T7451, Al-Li-Cu, and 6013-T6, respectively.

Dalle Donne et al. (2001) used Neutron, Synchrotron and X-ray Cu-ka techniques to extract the distribution of the longitudinal residual stress profile in FSW of 6013-T6 alloys. The results obtained from these techniques showed good quantitative and qualitative agreement. Generally, the result of the residual stress analysis reveals that the residual stress distribution is inhomogeneous in the longitudinal and transverse direction of the weld as well as across the thickness of the weld, as shown in Figure 8 for the longitudinal direction.

The maximum values of the tensile residual stresses are located in the HAZ. From these various values the tensile residual stresses decrease in the PP material adjacent to the HAZ as well as the weld seam, which contains small compressive residual stresses. With increasing distance from the weld, the residual stress then gradually changes into the initial stress state of the sheet material.

Sources for Fatigue Problems in Friction Stir Welds

Although FSW is a solid phase process, changes in the microstructure, a softened heat affected zone, highly tensile residual stress, and defects are also produced, as in the traditional fusion weld (Dawes, 1995; Dawes and Thomas, 1996; Nicholas and Thomas, 1998; Bussu, 2000). The limitations of the FSW process are being reduced by intensive research and development. The following section presents the problems faced from the manufacturing and engineering points of view in FSW joints.

Manufacturing Defects in FSW

The main limitation of the FSW process at present is the welding speeds, which are moderately slower than those of some fusion welding processes (up to 750 mm/min for welding 5 mm thick 6,000 series aluminium alloy on commercially available machines). However, the more obvious challenge is to maintain the weld quality by reducing the defect level or to eliminate the defects. Recent characterisation exhibits the typical manufacturing defects in FSW to be severe surface irregularities, internal defects, porosity, and oxide inclusions.

It is worth noting that, although the fine grain structure is formed in the N region, the defects such as voids, inclusions, and surface cracks are dominant in this region due to the stir process. The defects have been found using scanning electron microscopy (SEM) techniques by many researchers (Flores et al., 1998; Bussu, 2000; Dalle Donne et al., 2000; Webster et al., 2001), with examples shown in Figures 8 and 9.

Fatigue failure process is sensitive to defect. As a rule of thumb, any parameter that will increase the local stress concentration will also be degrading for fatigue performance. Therefore the defect level in FSW needs to be controlled or eliminated, in order to improve the joint strength.

Residual Stress Induced in FSW

Self-stresses exist in most manufactured parts. Self-stresses have also been called self-equalizing stresses within a part, without any external load. They have been called residual stresses because they may be left over from a previous manufacturing operation. Residual stresses exist in many manufactured components as a result of
the mechanical and thermal processing. From these manufacturing processes material flaws or small cracks can be generated which may start to grow under cyclic loading during service. Therefore, the understanding of fatigue crack growth under any type of loading and residual stress conditions is an important dilemma in engineering with respect to the life or safety of certain engineering components.

It is interesting to note that a crack may initiate its own residual stress field under cyclic loading, and it was believed to have a significant affect on fatigue crack growth (Meguid and Coufopanos, 1986). A lot of research found that, in regions of compressive residual stresses, fatigue crack growth was retarded or cracks were arrested, while tensile residual stress regions generated opposite effects (Sun and Sehitoglu, 1992; Beghini et al., 1994; Fitzpatrick and Edwards, 1998; Mohshier and Hillberry, 1999; Wang et al., 1999). This means that tensile residual stresses that produce on opening movement have a detrimental effect on fatigue life. On the other hand, compressive residual stresses that produce a clamping movement on the crack faces can be beneficial.

From the previous mention of residual stress characterisation, it is evident that the existing highly tensile residual stress in the HAZ is due to thermal heating and cooling processes during FSW. The main problem is a challenge to manufacturers to reduce the tensile residual stress level, whilst a second challenge to engineers is to overcome the effect of these tensile residual stresses in promoting fatigue crack propagation.

At this stage there is no clear picture about the residual stress distribution generated, depending on the type of material used, welding speed, plate thickness, and plate clamping forces during joining in FSW. Nevertheless, several attempts have been made to simulate the residual profile induced after the welding process in three dimensional representations which are still under development (Oosterkamp et al., 2000; Ulysse, 2002). Once the residual stress distribution is clear and well controlled, perhaps the next consideration is how the residual stress in FSW will be relaxed during cyclic loading.

**Understanding of Fatigue Behaviour Concerning FSW**

Fatigue stages involve nucleation or initiation of a crack, growth of the fatigue shear crack (Stage I), tensile crack propagation (Stage II), high growth rate or crack coalescence (Stage III), and...
fracture or failure. Once initiation begins and forms a microcrack, this crack may be arrested by a microstructural barrier or may propagate until reaching a critical size, causing the final failure (Navarro and de los Rios, 1992; Miller and O’Donnell, 1999). The fatigue life is defined as the sum of the number of cycles to initiate a fatigue crack and the number of cycles to propagate it subcritically to some final crack size (Suresh, 1998). Therefore, to have a good fatigue life of FSW joints, one needs a good control of initiation and propagation stages.

In FSW, there has been no systematic attempt to investigate the natural crack initiation site so far. The question about where is the crack initiation origin in FSW remains unclear. A theoretical answer is needed to justify why the crack is not initiated at the hardness minima as reported by Bussu and Irving (2003), why the crack can initiate in the finest grain region, and what is the role of macrostructure, microstructure, hardness, and residual stress, which may influence initiation behaviour in FSW.

The parameters that govern the crack initiation site in FSW seem to be complicated to investigate. Perhaps the early stage of initiation was a process of the competition of the inherent microcracks and defects in FSW. The issues of crack initiation in FSW need to be resolved very soon because the initiation is extremely important to determine an accurate fatigue life of these joints.

In terms of fatigue crack propagation, current work on fatigue crack propagation of FSW has been done by introducing a semicircular flaw by electric discharge machining (EDM) at each zone (Bussu, 2000; Bussu and Irving, 2003) or by using compact tension specimens (Dalle Donne and Biallas, 1999; Jata, 2000; Jata et al., 2000; Strombeck et al., 2001). As a result, under fatigue cycling a crack was initiated and propagated from the EDM defect of the notch. It is evident that the fatigue crack growth rates (FCGR) in the N and TMAZ are approximately 10 times higher than those of the PP as shown in Figure 10. Here the TMAZ is located around 11mm from the PIL. The TMAZ and N exhibited the lowest threshold stress intensity factor.

Bussu and Irving (2003) suggested that

Figure 10. Crack growth data, 2024-T351 for crack starting from PJL (N), 11 mm (TMAZ) and 25 mm from PJL (HAZ) (Bussu and Irving, 2003)
the residual stress plays the dominant role of accelerating and decelerating the propagation rate which is shown by comparing the propagation data of welded and stretched specimens as shown in Figure 11. With 2% pre-strain, the residual stress field is wiped out by plastic deformation. Hence Figure 11 shows the growth rate in the absence of residual stress. However it is noted here that, in the case of the HAZ, crack growth rates are marginally lower and, for the N, they are faster than the TMAZ, indicating that residual stress does not entirely account for all the observed trends in FCGR. The role of microstructure should also be considered.

The coarse microstructure in the HAZ region should deteriorate the FCGR. However the compressive residual stress components override the microstructural role and produces beneficial FCGR in the HAZ (Jata, 2000). There is no attempt so far to model the fatigue damage of FSW joints. The existence of inhomogeneity of the microstructure, hardness, and residual stress distribution in FSW could lead to complications in the process of development of the fatigue model.

However before commencing with modelling one should gather complete information about the mechanical properties of the weld. If there is clear distinction in the microstructure and the properties of each zone, should the model treat them as bimaterial cases? In fatigue, many attempts have been made to model the FCGR, from the simplest empirical model by Paris and Erdogan (1963) to advanced and contemporary models such as the Weertman (1979) strain energy model, Tomkins (1981) high strain model, Coffin (1950) plastic strain model, Hobson-Brown model (Hobson et al., 1986) and Navarro- E de los Rios (de los Rios, and Navarro, 1988; Navarro and de los Rios., 1988a, 1988b) N-R models.

The question is, which models are appropriate and would they be able to model the FCGR in FSW by fully incorporating the microstructure, hardening, residual stress, or mean stress effects? Ideally a robust model is needed to handle the situation of the crack propagating from one region to another experiencing different residual stress, hardness, and closure effects, and whether or not the residual stress will relax in FSW joints. Are there any phenomena of crack retardation or acceleration while propagating from one zone to another in FSW?

In simulating the process, the models should be developed based on fatigue design criteria. Therefore in terms of structural integrity these are issues concerning fatigue design philosophies i.e. infinite life, safe life, fail safe, and damage tolerant design for FSW, which need to be resolved quickly.

Conclusions

1. The macrostructure and microstructure of a FSW joint are associated with the welding process itself. The microstructure features of the N and FA were the finest grain dimension and intermetallic particle distribution, whilst these of the TMAZ and HAZ are coarsened.

2. The hardness in FSW is governed by the distribution of the strengthening precipitates affected by the heat that causes overaging, rather than the grain size.

3. Residual stress distribution is inhomogeneous and gradient in the longitudinal and transverse direction of the weld as well as across the thickness.

Figure 11. Crack growth data, 2024-T351 for samples FSW strained 2% parallel to weld line (Bussu and Irving, 2003)
4. The fatigue life of FSW could be determined accurately by addressing the role of microstructure, hardness and residual stress wisely from the initiation stage to fracture.

5. A robust model needs to be developed that can incorporate the effect of residual stress relaxation, hardening, closure, and mean stress on fatigue of FSW. To obtain an accurate fatigue life prediction, the role of microstructure, hardness, and residual stress on initiation and crack propagation must be well addressed.

References:


