Thermal modelling of linear friction welding

P. Jedrasiak a,b,⁎, H.R. Shercliff a, A.R. McAndrew b,c, P.A. Colegrove c

a Department of Engineering, University of Cambridge, Trumpington St, CB2 1PZ, UK
b TWI, Granta Park, Cambridge CB21 6AL, UK
c Welding Engineering Research Centre, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

HIGHLIGHTS

• A finite element thermal model for linear friction welding (LFW) was applied to an instrumented weld in Ti6Al4V.
• Expulsion of flash was included using a sequential step-wise technique, removing interface elements at discrete intervals.
• Predicted power was reasonably consistent with the experimental power, inferred from transverse load and displacement data.
• The thermal model sufficiently accurate for application to a new thermomechanical modelling approach for LFW.

GRAPHICAL ABSTRACT

Iterative axial shortening in the thermal model:

1. Introduction

Linear friction welding (Fig. 1) is a solid state joining method, in which a component subjected to reciprocating transverse motion is pressed against a stationary component. Frictional heat and plasticity at the interface soften the material, much of which is expelled as flash, and the joint is consolidated under axial load. Vairis and Frost [2] divided the process into four distinct stages. In the initial phase, oscillatory motion and pressure reach a stable value. Heat is generated through friction, and no axial ‘burn-off’ (shortening) is produced. In the transition phase, the temperature of the surrounding material rises and a layer of plasticised material is formed, along with initial flash. Subsequently, in the equilibrium phase, plasticised matter continues to be expelled as a constant rate, and the thermomechanically affected zone...
(TMAZ) and heat-affected zone (HAZ) expand. Finally, in the deceleration and forge phase, the amplitude is rapidly reduced to zero, and the forging pressure consolidates the joint.

Like other solid state friction welding methods, the process offers a number of advantages over fusion welding, including avoiding solidification and grain boundary liquation cracking, porosity, microsegregation and grain coarsening. The growth of intermetallic compounds can also be controlled when welding dissimilar metals. It is fast and repeatable, easy to automate, energy efficient, and requires little preparation of joined surfaces with no shielding gas or consumables [3, 4].

The process was developed mostly for joining aero-engine compressor blades to compressor disks, and for materials such as titanium alloys, titanium aluminides and nickel superalloys [5]. Titanium is particularly well-suited for LFW because of its mechanical properties and low thermal conductivity, which confines heat to the welded interface. An important aspect of LFw of titanium is the phase transition between modified HCP α-titanium at low temperatures, and BCC β-titanium (at about 1000 °C for Ti-6Al-4V [6]). The impact of the α-β transition on the deformation in linear friction welds of Ti is discussed in the subsequent paper on modelling of the heat generation [1].

1.1. Overview of experimental work

The early work by Vairis and Frost [2, 7] concentrated on Ti-6Al-4V and the three main welding parameters (oscillation amplitude, frequency and pressure) and their influence on the impact strength of the weld. The analysis covered several characteristics related to weld integrity: heat input, the size of the HAZ, extrusion of flash, and interface alignment. Ti–6Al–4V, nickel superalloy and steel LFw welds were characterized by Li et al. [8–13], who studied flash formation, axial shortening, temperature distribution, microstructure and fatigue performance. The strength, ductility and resistance to fatigue cracking of defect-free LFw welds can be comparable to the parent material. In Ti–6Al–4V, the weld strength may exceed that of the parent material, due to a refined microstructure resulting from high strains and rapid cooling of fully β transformed material. Non-recrystallised deformed grains form a TMAZ around the weld zone, but no prominent HAZ is observed in titanium [5, 13, 17, 19].

Jahazi et al. [14] provided further insight into the influence of welding parameters on tensile strength in Ti–6Al–4V LFw, identifying axial shortening as a further important parameter. Their study showed different microstructures in the weld zone and the TMAZ, suggesting temperatures below the Ti–6Al–4V β-transus in the TMAZ. Investigation of the influence of welding parameters on the interface temperature, which controls β grain growth, has identified power input and flash expulsion as key factors controlling the temperature [5, 14].

Karadge et al. studied microstructures in Ti–6Al–4V and nickel superalloy LFw, and showed that strong texture develops within ± 100 μm of the weld interface, though its evolution and influence on the properties are not fully understood. [15, 16] Using SEM and EBSD they showed that some of the observed texture could be attributed to the β → α transformation during cooling. As in most welding processes, residual stresses in LFw can have a negative impact on the weld integrity, as a consequence of the local thermal histories and constrained plasticity during welding [5].

1.2. Overview of numerical work

Linear friction welding has been extensively modelled with numerical methods, especially for Ti–6Al–4V [10, 17–25], other titanium alloys [12, 26], and their dissimilar combinations [27]. Turner et al. [20] and Schröder et al. [24, 25] studied peak temperatures and variation of the upset and flash shape, but also strain-rates, HAZ size, the onset of plasticity, self-cleaning of the joined surfaces, and the influence of weld size. In later work, Turner et al. [19] and Song et al. [28], successfully predicted residual stress, and verified them experimentally with X-ray diffraction. McAndrew et al. [22, 23, 29] explored weld characteristics (temperatures, strain-rates, burn-off rates, TMAZ thickness, welding forces and power) and their dependence on oscillation amplitude, frequency and pressure. Temperature distribution and heat generation, as well as residual stresses, were also studied by Buhr et al. [30, 31].

To date, numerical modelling of LFw has been dominated by fully coupled thermomechanical finite element analysis with explicit time integration. Finite element thermal models are used only to supplement coupled thermomechanical models, e.g. for predicting the temperature field after the dry friction stage, to provide the initial temperature of a fully-coupled model of the equilibrium stage (McAndrew et al. [22, 23, 29]). Buffa et al. [32] also used a heat transfer model to complement a fully-coupled thermomechanical model for predicting shear yield stress. In contrast, Vairis and Frost [18] separated the process into a thermal and a mechanical problem, solved in parallel during every time step, and coupled via transferring the mechanical work and temperature field. The current work picks up on this semi-coupled approach, seeking improvements in the computational efficiency of the analysis.

1.3. Small strain method for modelling LFw

Fully coupled thermomechanical models commonly handle the large deformations in LFw by an Arbitrary Lagrangian–Eulerian (ALE) kinematic description or other remeshing techniques [10–13, 17, 19,
shows the workpiece dimensions and the locations of four alloy at TWI [22, 23] with the following parameters: frequency 50 Hz, amplitude 2.7 mm, downforce 100 kN, and burn-off 3 mm.

Fig. 2. Workpiece dimensions for both parts (dimensions in mm). Thermocouples were positioned in the holes shown in sections A-A, B-B, at distances of 0, 0.5, 1.5 and 2 mm from the interface.

This new method uses independent thermal and deformation models, coupled via transfer of the thermal load and temperature field. The deformation model takes intermittent small strain “snapshots” during the continuous thermomechanical process, in order to capture enough of the material flow behaviour to compute realistic values of heat generation. There are two sources of computational efficiency – firstly, the deformation model runs for only a small fraction of the total process time; and secondly, the plastic strain and mesh distortions are small, so that demanding remeshing techniques are avoided.

This paper presents the first part of this new approach – an implicit FE thermal model of LFW, to predict the temperature field for the entire weld duration. For the thermal model, the important characteristics are the spatial distribution and time evolution of the heat input, and the ability to handle the significant axial shortening and heat loss to the flash. The integration of this thermal model with the small-strain thermomechanical model is presented in a subsequent paper [1]. Further potential applications of the thermal model include time-efficient parametric studies, and providing thermal histories for modelling and interpretation of weld microstructures and properties, as demonstrated in other friction welding contexts – for example, the post-weld hardness in precipitation-hardened aluminium alloys [39] or the formation of intermetallic compounds at the interface of dissimilar welds [39, 40].

2. Experimental work

A single instrumented weld was produced in Ti-6Al-4V titanium alloy at TWI [22, 23] with the following parameters: frequency 50 Hz, amplitude 2.7 mm, downforce 100 kN, and burn-off 3 mm. Fig. 2 shows the workpiece dimensions and the locations of four k-type thermocouples, fixed in the drilled holes with epoxy resin with a precision of approximately 0.1 mm.

Fig. 4(a) shows the upper and lower envelopes enclosing the peaks of the 50 Hz sinusoidal oscillation in the measured total transverse force, \( F_{\text{total}} \) (black curves). Ofem et al. [41] showed how to account for the inertial contribution to this total force, enabling the interface force \( F_{\text{int}} \) to be inferred. Assuming sinusoidal oscillation of a mass \( m \) for the chuck and workpiece (=280 kg in this setup):

\[
\begin{align*}
F_{\text{int}} &= F_{\text{total}} - ma \\
\dot{a} &= -A(2\pi f)^2 \sin(2\pi ft) = -x(2\pi f)^2
\end{align*}
\]

where \( a \), \( f \) and \( A \) are the oscillation acceleration, frequency and amplitude respectively. Inferring the interface force from experimental data needs care however. Fig. 4(b) shows sample cycles from the equilibrium stage of welding. The total force and displacement curves are not entirely smooth, particularly at the peaks, due to the data sampling frequency being too low. Acceleration was calculated by double numerical differentiation of the displacement data, giving the inertia force, \( (ma) \), which is 180° out-of-phase with the displacement (Fig. 4b). Subtracting the inertia force from the total force gives the interface force \( F_{\text{int}} \). Note however that the inertia force leads the experimentally measured total force by a small phase shift, and the magnitude of the resulting sinusoidal variation in \( F_{\text{int}} \) is sensitive to this phase shift. The accuracy of the data acquisition system was therefore checked by examining the predicted interface force before contact was established, when \( F_{\text{int}} \) should be zero. This revealed a small time offset between the force and displacement data, which was therefore eliminated as far as possible before predicting \( F_{\text{int}} \). The envelopes of the resulting predicted cycles of inertia and interface forces are also shown in Fig. 4(a). At first contact at \( t = 0 \), the small residual non-zero amplitude in \( F_{\text{int}} \) can be seen – this was minimised by adjusting the time offset in the raw data to 2.6 ms.

The interface force and velocity from this analysis are then used to estimate the power input to the weld, since the instantaneous power \( Q \) is given by:

\[ \dot{Q} = F_{\text{int}} \nu \]

Note that both \( F_{\text{int}} \) and \( \nu \) are approximately sinusoidal, but out-of-phase (by a little over 180°). In order to minimize the windowing
error associated with the relatively low data sampling frequency, the interface force and velocity curves were fitted to sinusoidal functions with constant frequency, but time-dependent amplitude and phase shift. This adjustment was necessary, as the force amplitude varies substantially throughout the process, while the power output is very sensitive to the phase difference between force and velocity. The instantaneous power of plastic dissipation was then averaged by integration over each cycle, to obtain the average power history $Q_{\text{avg}}$ over the full welding time:

$$Q_{\text{avg}}(t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} Q(t) \, dt$$

Fig. 5 shows the inferred variation of $Q_{\text{avg}}$ through the weld cycle. Due to the cumulative uncertainty in the data and the analysis, this experimental power history is not sufficiently accurate to be used as a model input. Obtaining accurate experimental power histories in friction processing is often difficult, or omitted altogether. The data manipulation needed here would not be suitable for routine measurement, but is important here to give the best available semi-quantitative validation of the power history inferred from thermocouple data and the thermal model.

3. Thermal model

A thermal model should ideally use an input power history that is validated independently by direct experimental measurement of power. But in this case the uncertainty in inferring input power to the workpiece (discussed above) necessitates a different approach, reverse-engineering the power input using thermocouple data. The spatial distribution of this power requires a physical basis, for which a simple analytical model is proposed. Finally the evolving geometry, due to burnoff, is managed via a sequence of independent simulations with small step changes in the geometry.

3.1. Geometry, boundary conditions and mesh

Linear friction welding can be modelled in two dimensions, without compromising the quality of the results, for several reasons. First, heat flow is practically one-dimensional, as the horizontal cross-section (parallel to the interface) is uniform, heat losses to the air can be neglected for a short cycle, and the heat generation can be assumed uniform for most of the interface. Second, most of the plastic deformation and expulsion of flash takes place through shearing in the plane of the welded interface, parallel to the oscillations. Some flash is extruded in the out-of-plane direction, but as shown later, axial shortening and the associated heat loss can be accounted for without modelling flash formation in full. Modelling in 2D provides a substantial gain in the efficiency of computation.

Temperature-dependent thermal properties and density of Ti-6Al-4V were used [42–46]. Due to the low thermal conductivity of titanium and the short welding cycle (of order 1 s), the heat flow distance during the weld cycle is limited in extent – for example, the temperature rise at a distance of 10 mm from the interface was below 1 °C. Hence the initial
dimensions of the workpieces were limited to 10 mm (including an allowance of 3 mm of burn-off), and remaining parts of the workpieces and clamping were neglected. All external boundaries were therefore defined as thermally insulated (Fig. 6). Significant amounts of heat in LFW are, however, lost from the vicinity of the interface via mass convection into the flash, and this burn-off is accounted for in the model, as discussed later.

The two workpieces are the same material and of equal size, with rapid sinusoidal oscillations about the position of complete overlap. Hence the heat conduction problem is symmetrical about the interface, with identical heat flow in both workpieces, and no heat transfer across this plane. The thermal contact conductance at the interface between the workpieces therefore has no influence on the thermal field and may be assumed to be perfect for simplicity. This was recognised by Vairis and Frost [18] and Li et al. [10], who reduced the problem to a single workpiece, treating the second as a rigid surface. Alternatively, after full contact is established leading into the equilibrium stage, the workpieces can be modelled as a single body, as in the work of McAndrew et al. [22] and Turner et al. [20].

The mesh (Fig. 6) was graded from 40 μm to 1.5 mm in the far field to limit the total number of degrees of freedom in the model. A sequential step-wise technique was used for handling burn-off, and this defined the size of the smallest elements perpendicular to the interface. Four-node linear heat transfer quadrilateral elements were used in the thermal model.

3.2. Thermal loads

Temperature data is used as an input for the thermal model in an iterative procedure, where the net heat input is reverse-engineered until the temperature predictions match the experimental data. The time history of power is obtained in a piecewise fashion, with the power being adjusted to match the temperature history from the thermocouple closest to the interface. As burn-off reaches the location of each thermocouple, it is moved and potentially damaged, so calibration then switches to the next closest thermocouple, and so on. The fit to the full thermal cycles of all thermocouples is then checked using the inferred power history. Inaccuracies in temperature data, and thus inferred power, are mostly related to thermocouple locations. As a result of the low thermal conductance of Ti-6Al-4 V, temperature gradients around the interface can reach as much as 500 °C/mm, requiring precise positioning of thermocouples.

The heat input is assumed to occur only at the welded interface. This is obviously accurate in the initial dry friction stage, but during the equilibrium stage, heat is generated in a layer of the bulk material through plastic dissipation. In the subsequent paper [1], it is shown that the deformation takes place in a thin layer, less than 1 mm thick, so it is a reasonable assumption to apply all of the heat input at the interface.

The lateral distribution of heat input at the interface can be established with a simple analytical model, first for the initial frictional stage. Consider two rigid blocks sliding against each other, with a constant pressure p, a coefficient of friction μ, and shear stress μp at the interface (Fig. 7a). One block is oscillating with a frequency f and displacement x in a coordinate system fixed with the stationary block, while the peak-to-peak amplitude is 2A. Contact is lost for part of the cycle in a region of width A at the edge of each block (Fig. 7a). The rate of heat generation per unit area is proportional to the distance over which contact is maintained throughout a cycle. In the central region (x > A), contact is maintained over a sliding distance of 4A per cycle, hence the average heat generation rate per unit area is equal to μp (4A)f. At the edges, heat will not be generated when the blocks are not overlapping. The edge of one workpiece is in contact with the other workpiece for exactly half a cycle, halving the average power density. This gives a linear fall-off in power at the edge, over a distance A:

\[
\begin{align*}
q_{avg} &= \mu p f 4A = q_{max} \\
q_{avg} &= \mu p f 2/2A - (A-x_s) = q_{max} / 2 (1 - x_s / A)
\end{align*}
\]

Due to symmetry, analogous expressions can be derived for the right-hand side of the interface, resulting in the final distribution in Fig. 7b.

After full contact of the joint develops, heat is generated via plastic dissipation, by shearing of the material close to the interface. It is then assumed that power is generated continuously over the entire interface, regardless of the overlap between the blocks, which are now physically connected. And if the temperatures and strain-rates are also uniform over the interface, the flow stress and resulting heat input distribution can be assumed uniform across the entire interface (Fig. 8). The results of the thermal model show that it is reasonable to assume a uniform

\[
\begin{align*}
q_{max} = \mu p f 4A = q_{avg}
\end{align*}
\]
temperature, while uniform strain-rates are a natural consequence of the constrained geometry of the plasticized layer.

3.3. Axial shortening

One particularly challenging aspect of modelling LFW with the proposed small-strain framework is the need to handle the change of geometry associated with burn-off. This is achieved in the thermal model by “deleting” layers of elements at the interface at intervals, with each layer of elements corresponding to an equivalent volume of material extruded to flash (that is, discretising further what is in reality a cycle-by-cycle event). The procedure is illustrated in Fig. 9. All the elements in the weld zone are arranged in regular layers of constant thickness. The rate of removal of element layers is controlled using the experimentally recorded axial displacement (which determines the rate of flash expulsion). After each operation of “deleting” a layer of elements, a new model is assembled by closing the gap vacated by the “deleted” layer. The initial thermal field is imported from the previous model on an element-by-element basis, excluding the deleted layer. The new thermal model is then run for a time corresponding to burn-off of the next single layer of elements, and the resulting thermal field is again exported to the next iteration.

Since the non-linearities associated with temperature-dependent material properties are mild, the solution can be obtained in one time-step in each implicit thermal model (corresponding to a single step in burn-off). Consequently, CPU time associated with each model can be assumed constant, and the overall computational time is proportional to the total number of burn-off steps:

\[ \text{Computational time} = \text{Burn-off steps} = \frac{\text{Total burn-off thickness}}{\text{Element layer thickness}} \]

As the cycle-by-cycle removal of interface material is simulated in larger steps, most of the time the volume of material in the model is slightly different than in the real experiment. Furthermore, the distance between the heat input and the thermocouple used for its calibration also changes in discrete steps. As a result, the element thickness has a comparable effect on the accuracy of the calibrated heat input as does the lack of precision in thermocouple positioning. As a compromise between computational efficiency and accuracy, an element thickness of 40 μm was adopted.

4. Results

Full temperature cycle predictions, obtained with the resulting inferred power history, are compared with the thermocouple data in Fig. 10. Prediction at each location finishes when burn-off reaches that point, except for the 2 mm case, where the data are truncated at the end of the heating stage. A good match was obtained for temperature predictions at each of the thermocouple positions.

![Fig. 9. Iterative axial shortening in the thermal model.](image)

![Fig. 10. Predicted temperature histories (dashed lines) and thermocouple data (solid lines), for given (initial) distances from the interface.](image)
phase, and identified potential sources of the significant discrepancy observed, which were up to 50%. These included the heat lost to flash expelled in the direction perpendicular to oscillation, which was neglected in their two-dimensional model, and uncertainty in the material constitutive data used to predict the heat generation rate. In the current work, the step-wise technique of handling burn-off captures the heat lost to flash in all directions, while our thermal model is calibrated to thermocouple data as an intermediate step before developing a thermomechanical model of heat generation.

Uncertainty in the measured power is related to the assumption of sinusoidal displacement, windowing errors due to an inadequate frequency of data acquisition, and possible contributions of forces other than the interface force and inertia. Furthermore, if the welded surfaces are not perfectly aligned at the start of the process, the initial heating and temperature rise would be concentrated towards one edge of the sample while the contact beds in this. This could explain the more gradual initial rise in measured power than was inferred using the thermal model. So in spite of the considerable experimental uncertainties, the two independent measures of the power history provide sufficient confidence in the thermal model for it to be applied in the subsequent small-strain deformation model [1].

Fig. 12 shows the predicted temperature distributions at the end of the equilibrium stage. The influence of the low thermal conductance is clear, as even at the end of the equilibrium stage the region of elevated temperature is limited to the vicinity of the interface.

5. Conclusions

A two-dimensional thermal model was developed for linear friction welding and applied to an instrumented weld in Ti6Al4V, with multiple embedded thermocouples, and force and displacement measurement in the transverse and axial directions. The cyclic force at the interface could be estimated by accounting for the machine inertia, and combined with the velocity cycle to give an independent estimate of the net plastic dissipation rate. The power history was reverse-engineered from the thermocouple data, and the lateral distribution of heat input at the interface was estimated with a simple analytical model. The expulsion of material to flash and associated change in geometry were successfully accounted for in the thermal model with a sequential step-wise technique, deleting layers of elements at the interface at discrete time intervals, matching the experimental burn-off rate. The temperature predictions corresponded well with the thermocouple data. The region of elevated temperature was found to be limited to a thin layer close to the interface, due to the low thermal conductivity of Ti alloys. Comparison with experimental power showed a 20% discrepancy throughout the equilibrium stage. This was attributed to uncertainty in the power inferred from force and displacement data, and sensitivity to the precision of locating the thermocouples. The thermal model is computationally efficient, and is sufficiently accurate for application to the proposed small-strain thermomechanical modelling approach developed in a subsequent paper [1].

Declarations of interest

None.

Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time due to technical or time limitations.

Acknowledgements

The work described in this paper was funded by EPSRC through the University of Cambridge Doctoral Training Account, with additional CASE award funding provided by TWI, Grant Park, Cambridge, UK. In addition, funding for the experiments was provided by the Boeing Corporation and EPSRC. The authors gratefully acknowledge the technical discussions and experimental assistance from Dr. Mike Russell, Dr. Kathryn Beamish and Bertrand Flipo, TWI.

References


