INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 9 Model Predictive Control and Economic Optimization

CHAPTER 3

Friction Stir Welding Simulator for Regime-Dependent Tuning and Predictive Control

Pages 49–75

Edited by John D. Hedengren

How to cite this chapter. "Friction Stir Welding Simulator for Regime-Dependent Tuning and Predictive Control," in *Model Predictive Control and Economic Optimization*, John D. Hedengren, Ed., vol. 9 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026, ch. 3, pp. 49–75.

The page numbers in this file are those of the printed volume.

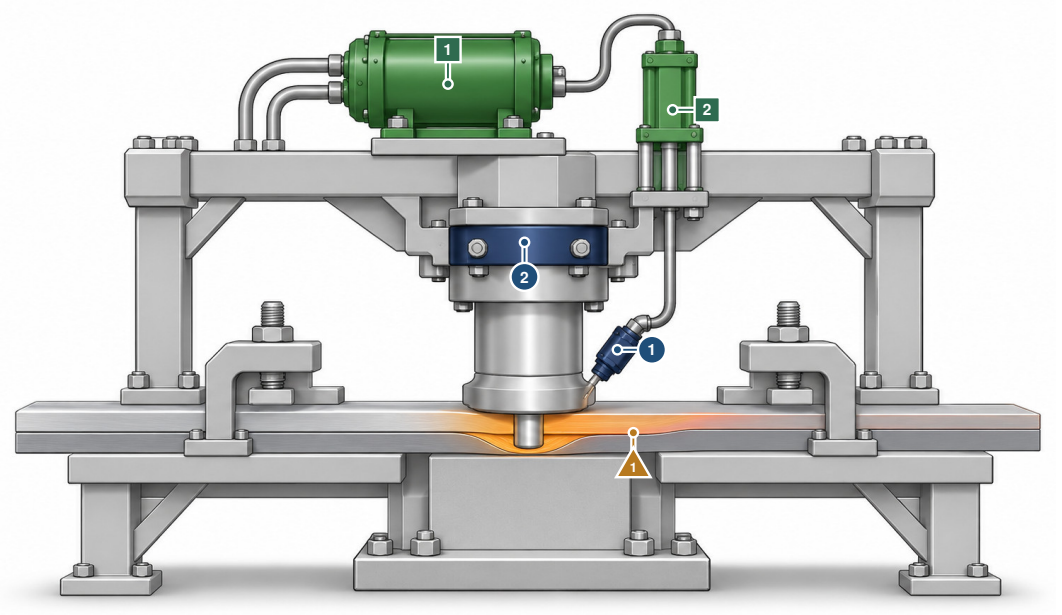
The app for this chapter
https://apmonitor.com/apps/friction-stir-welding/

The volume online: contents, other chapters and citation records
https://apmonitor.com/apps/books/v09-mpc-optimization/#friction-stir-welding

Copyright © 2026 Advanced Process Solutions, LLC. All rights reserved. Produced by Advanced Process Solutions, LLC.

CHAPTER 3

Friction Stir Welding Simulator for Regime-Dependent Tuning and Predictive Control



The controller moves 1 spindle drive (spindle speed) 2 axial axis drive (plunge depth)

The controller reads 1 tool temperature thermocouple 2 axial load cell

What disturbs it 1 thin patch in the plate

Notice that the spindle speed sets the heat, and that the weld answers it as one plant while the tool plunges and as another once it travels. Illustration, not to scale.

IN PLAIN TERMS Welding Metal Without Melting It

The main idea. A friction stir weld is made by a spinning tool that is pushed into the seam between two plates and dragged along it. Rubbing heats the metal until it is soft enough to be stirred shut. The spinning speed sets how hot the weld runs.

The challenge. While the tool sinks in, the weld answers a change of speed fast and strongly. Once the tool moves along, it answers slowly and weakly. An automatic speed controller that learns the weld at one of these times is wrong at the other.

Why it matters. The wrong choice leaves hollow or squeezed-out metal that must be cut away.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact in Section 3.1, the history in Section 3.2, the two try-it segments and the second fact in Section 3.6, and the key points
Study	55 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the results (Section 3.6), Technology in Practice (Section 3.7), Limitations (Section 3.8) and the open challenges

The control question A step test in quasi-steady welding returns a gain 1.7 times smaller and a time constant ten times longer than the same test through the plunge of the same weld. Which test should fill the dynamic matrix, and what does the wrong regime cost in millimeters of defect?

Learning objectives

- Explain from the flow stress why heat generation is self-limiting and why the plunge is a different plant.
- Identify a first-order-plus-dead-time model and compute ITAE servo and regulator tunings from it.
- Compare servo and regulator tuning phase by phase, and explain how the entry into Auto changes the result.
- Predict how a step-response controller whose model gain is too high responds, and why move suppression does not rescue it.
- Choose position or force control of the axial axis from the root clearance as well as the heat.

Prerequisites Sections 1.4.2 to 1.4.5 (step response, gain matrix, move suppression, output bias), 1.6.2 and 1.6.4, and Chapter 2.

The app https://apmonitor.com/apps/friction-stir-welding/

See also Volume 7, Chapter 2 for heat that grows with temperature and Volume 7, Chapter 13 for a limit set by the spindle speed. Volume 1, Chapter 2 for a model that follows the operating point and Volume 1, Chapter 12 for identification from records.

3.1 Introduction

A friction stir weld is made by a rotating shouldered tool that is pushed into the joint line between two plates and then dragged along it. Nothing melts. Friction under the shoulder and plastic shear around the pin bring the aluminum to roughly four fifths of its absolute melting temperature, where the metal is soft enough to be extruded around the pin and forged shut behind it [1, 2]. The quantity that decides whether the joint is any good is the tool temperature. Below the window the metal does not flow fast enough to close the wake behind the pin and the weld carries a continuous tunnel void, and above it the metal extrudes sideways from under the shoulder as flash while the root softens and the tool wears [1]. In the simulator of this chapter that window is 435 to 505 °C at 120 mm/min and both of its edges move with the traverse speed.

DID YOU KNOW? Half a Mile of Weld on Every Tank

Each Space Shuttle External Tank carried about half a mile of welds. About 700 feet of it, eight longitudinal joints on the liquid hydrogen barrel and four on the liquid oxygen barrel, was made by friction stir welding, and a plasma arc process made the rest [3].

A block diagram of that loop is a first-order lag with dead time and a single manipulated variable, and it hides the one property that makes the weld hard to control. Heat generation is proportional to spindle speed times interface shear stress, and that shear stress follows the flow stress of the shear layer, which falls by a factor of eighteen between room temperature and 500 °C [4, 5]. Heat generation is therefore self-limiting, which is why a friction stir weld does not run away and melt the joint and why the linearization of the loop moves with the operating point. The exothermic reactor of Volume 7, Chapter 2 is the opposite case, a heat source that grows with temperature until the vessel can ignite. The misconception the simulator is built to correct is that a single well-tuned controller is the goal. Measured on the shipped model, the same loop returns 0.1425 °C/rpm with a 12.2 s time constant in quasi-steady welding and 0.240 °C/rpm with a 1.2 s time constant during the plunge of the same weld, so the controller that wins one phase is mistuned for the other and the choice is which phase to be wrong in.

The class of problem is general. It is the plant whose linearization changes materially between operating phases inside one batch, so that a model-based controller inherits the operating point of the

experiment that identified it. Volume 1, Chapter 2 meets the same inheritance on a tank whose gain and time constant both follow the level. A batch reactor between charging and reaction, a furnace between light-off and soak, an extruder between purge and running, and a crystallizer between seeding and growth all have that shape, and the standard answers are gain scheduling, a controller identified over several operating points, adaptive identification, or a nonlinear predictive formulation [6, 7]. Surveys of industrial predictive control report that the identification experiment, and not the optimizer, is where most of the commissioning effort and most of the later trouble sit [8]. What a simulator adds is the ability to run that experiment twice on the same plant, in two phases of one weld, and then run both resulting controllers over a weld and count the defect each one leaves.

This chapter works with a friction stir welding simulator whose two dynamic matrix controllers come from two step tests taken on the same weld in different phases, so that the price of identifying a model-based controller in the wrong regime is measured in defect length rather than asserted. Section 3.2 traces how the machine moved from position to force control and then to closed-loop temperature control. Section 3.3 sets out the nine-state thermal and mechanical model, with its instrument chain, its force-limited axial axis and its quality window, completely enough to be rebuilt from the text alone. Section 3.4 identifies three plants inside one weld, and Section 3.5 describes the five architectures that share the one actuator: the operator, two ITAE-tuned PI loops and two step-response predictive controllers. The predictor, the gain matrix and the output bias are those of Sections 1.4.2 to 1.4.5. Section 3.6 runs the scenario sequence, from a weld made by hand to a graded challenge in which force control on the axial axis is a trap. Section 3.7 describes the sensors, drives and controller platforms of real welding machines. Sections 3.8 and 3.9 state what the model cannot support and which methods it could carry next.

3.2 Friction Stir Welding and the Arrival of Temperature Control

The process was invented at The Welding Institute in Cambridge in 1991 and patented there, and the first published account aimed at industry set out what the solid-state route buys, which is a joint in alloys that cannot be fusion welded at all, with no porosity, no solidification cracking and very little distortion [9]. Adoption followed the alloys. Shipbuilding and rail took extruded aluminum panels first, because a long straight butt joint in 6xxx extrusion is the easiest weld the process makes, and aerospace followed with the 2xxx and 7xxx alloys whose fusion weldability is poor [1, 9]. By the time the process was reviewed at length it had moved from a laboratory curiosity to a production method for aircraft floor panels, rail carriage sides, fuel tanks and, later, the cooling plates under electric-vehicle batteries [1, 2].

> ⧖ HISTORY **1995: A Laboratory Curiosity Reaches a Rocket Tank**
>
> Friction stir welding was conceived in the United Kingdom, and when NASA took it up in 1994 it was little more than a laboratory curiosity. A laboratory system made the agency's first welds in the fall of 1995, and development began on a five-axis horizontal boring mill given over entirely to the work. Its engineers found that the process needs the control of only three variables, shoulder plunge depth, tool rotational speed and travel speed, and that once optimized they repeat weld after weld. A simulated hydrogen tank barrel 27.5 feet in diameter was welded in 1998, and in 2001 the process entered production of the Space Shuttle External Tank, for the longitudinal welds of its hydrogen and oxygen barrels [3]. Those three variables are the recipe that this chapter's loops act on.

Automation of the machine came before automation of the process. The first generation of production equipment held the tool at a programmed height above the machine table, because a milling machine already did that and because position is what a numerical control program naturally commands. Position control on a plate of varying thickness costs shoulder engagement directly, so the axial force and the heat input follow the rolled stock rather than the recipe, and the industry moved to force control on the axial axis for that reason [10]. The requirements that make force control

work against finite machine stiffness and a compliant fixture, chiefly the load-cell bandwidth and the axial stiffness of the head, were identified explicitly for robotic equipment where the compliance is largest [11]. Force control is the production default now, normally interlocked against a position limit rather than run alone, because a force loop answers a soft patch of plate by sinking a pin that has only a fraction of a millimeter of clearance to the backing anvil [10]. A robot that pushes a peg into a hole faces the same choice once contact fixes the part, as Volume 20, Chapter 13 shows.

Neither position nor force is the variable that sets weld quality. Temperature is, and closing a loop on it required a temperature measurement from inside a rotating tool. That measurement became practical with a thermocouple embedded a few millimeters behind the shoulder and telemetry or a slip ring off the spindle, and the first closed-loop demonstrations used it to hold a tool temperature by manipulating spindle speed, reporting that the loop removed the sensitivity to the fixturing and the plate condition that an open-loop recipe carries [12]. Model-based control of the same loop followed. A first-order-plus-dead-time model identified from a step in spindle speed supports a predictive controller that outperforms a PID loop on the transient parts of a weld, and the comparison has been reported for the initial traverse and for quasi-steady welding separately [13], and more recently with a distributed thermal model inside the controller [14]. Qualification practice has not moved with it. A friction stir weld is still qualified as a recipe, which is a spindle speed, a traverse speed and a plunge depth run open loop [10], and the scrap that practice produces is concentrated at the two ends of every weld, where the recipe is wrong.

3.3 Process Model

The simulated joint is a 6.35 mm AA6061-T6 butt weld 245 mm long in a 300 mm plate clamped to a steel backing anvil. The tool has a flat 12 mm shoulder and a cylindrical 4 mm pin 6.00 mm long, so a 0.15 mm shoulder plunge leaves 0.20 mm of unwelded parent metal under the pin. The machine program is a plunge to depth, a 12 s dwell, the traverse, a run-out at the plate edge and a retract. Figure 3.1 shows the arrangement, the four thermal nodes, the conductances between them and every symbol used below. The controlled variable is the tool temperature T_t as reported by the rotating thermocouple, and the manipulated variable is spindle speed ω in every architecture the application offers.

What the loops interact with is classified in Table 3.1. The plant is written in the units a welding engineer uses, temperature in °C, spindle speed in rpm, traverse speed in mm/min, depth in mm, force in kN, torque in N m and time in seconds, with SI substituted inside the stress and heat expressions and converted at the boundary. Nine states are integrated. Four of them are temperatures, two are drive states, one is the axial position of the shoulder above the machine table, one is the distance welded and one is cumulative tool wear.

3.3.1 Contact Geometry

Everything mechanical follows from the tool height, so the geometry is written first, as shown in Equation (3.1).

$$\begin{aligned} z &= t + \Delta t - h_s, \qquad c = h_s - H_p \\ f_s &= \mathrm{clamp}(z/z_{eng},\,0,\,1), \quad L_p = \mathrm{clamp}(z + H_p,\,0,\,H_p) \\ f_{tip} &= \mathrm{clamp}((z + H_p)/0.2\,\mathrm{mm},\,0,\,1) \\ A_{sh} &= \pi\left(R_s^2 - R_p^2\right) f_s, \qquad A_{tip} = \pi R_p^2 f_{tip} \end{aligned} \tag{3.1}$$

where z is the shoulder plunge below the plate surface in mm, c is the unwelded parent metal left under the pin, t is the nominal plate thickness, Δt is the thickness deviation of the rolled stock, h_s is the height of the shoulder face above the table, H_p is the pin length, f_s is the fraction of the shoulder in contact, L_p is the buried length of the pin and f_{tip} is the fraction of the pin face in contact. The shoulder radius is derated by wear, $R_s \leftarrow R_s(1 - 0.04W)$, which is one of the two ways wear reaches the heat balance.

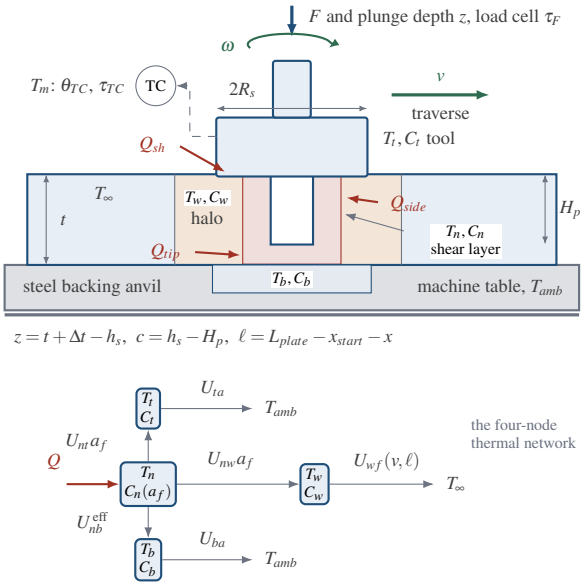


Figure 3.1 Section through the weld, above, and the four-node thermal network it reduces to, below. The rotating tool generates Q_{sh} under the shoulder, Q_{side} along the buried pin and Q_{tip} at the pin face, all into the shear layer at T_n, whose capacity and whose three conductances scale with the engaged contact fraction a_f. Heat leaves that node up the tool to T_t, sideways into the heated halo at T_w and down into the anvil interface at T_b, and leaves the halo into cold plate at T_∞ through a conductance that changes with traverse speed v and with the remaining distance ℓ to the plate edge. The controller reads T_m from a thermocouple in the rotating tool and manipulates ω.

Table 3.1 Variables the Loops Interact With, Read at the Design Weld of 900 rpm, 120 mm/min and 0.15 mm of Shoulder Plunge, 90 mm Along the Joint.

Symbol	Role	Nominal	Range
T_t	CV, measured	471.4 °C	380–560 SP
T_m	measurement of T_t	471.4 °C	delayed, lagged
ω	MV	900 rpm	200–1500
z	MV in force mode	0.15 mm	0–0.45
F	inner CV, force mode	9.50 kN	2–18 SP
v	operator setpoint	120 mm/min	20–300
Δt	DV, unmeasured	0 mm	±0.25
a_d anvil contact	DV, unmeasured	1.0	0.15–1.2
T_∞	DV, measured	22 °C	−20–80
W tool wear	DV, unmeasured	0	0–0.8 initial
T_n	state	516.6 °C	—
T_t	state	471.4 °C	—
T_w	state	395.6 °C	—
T_b	state	327.4 °C	—
ω_A, v_A	drive states	900, 120	—
h_s	shoulder height	6.20 mm	clamped
x	distance welded	90 mm	0–245
W	wear state	0	0–1

The engaged area also sets how large the shear layer is, as the fraction a_f of fully engaged contact

area in Equation (3.2).

$$a_f = \mathrm{clamp}\left(\frac{f_s\left(R_s^2 - R_p^2\right) + R_p^2 f_{tip} + 2R_p L_p}{R_s^2 + 2R_p H_p},\ 0,\ 1.2 \right) \tag{3.2}$$

The three conductances out of the shear layer and its own thermal capacity are all scaled by a_f, so a pin tip touching cold metal is a small fast node and a fully engaged tool is a large slow one. That scaling is what makes the plunge a different plant rather than the same plant at a different temperature.

3.3.2 Flow Stress and the Interface Shear Stress

The flow stress of the shear layer is the single expression from which the regime change follows, and it is written as a decaying exponential in temperature with a residual, collapsed linearly through the solidus to liquidus range, as shown in Equation (3.3).

$$\begin{aligned} \sigma(T) &= \max(0.2\,\mathrm{MPa},\ \sigma_b(T)\,\phi(T)) \\ \sigma_b(T) &= \sigma_0 \mathrm{e}^{-T/T_\sigma} + \sigma_{res} \\ \phi(T) &= \mathrm{clamp}\left(\frac{T_{liq} - T}{T_{liq} - T_{sol}}, 0, 1 \right) \end{aligned} \tag{3.3}$$

where σ_0 is a pre-exponential in MPa, T_σ is the decay temperature in °C and σ_{res} is the residual strength near the solidus. The constants are fitted to hot-working flow stress for 6061 at the strain rates a friction stir shear layer sees, of order 100/s, rather than to the quasi-static yield stress [4]. Equation (3.3) returns 1165 MPa at 22 °C, 209 MPa at 300 °C, 115 MPa at 400 °C, 65 MPa at 500 °C and 42 MPa at the solidus; from room temperature to 500 °C it falls by a factor of eighteen.

The interface carries either the sliding shear stress or the shear yield of the metal, whichever is smaller, and the transition between the two is the plunge to weld handover. Equation (3.4) writes it as a smooth minimum so that no corner appears in the Jacobian at the crossover.

$$\begin{aligned} \tau_{stick} &= \frac{\sigma(T_n)}{\sqrt{3}} \left(\frac{\max(1, \omega)}{\omega_{ref}} \right)^m \\ \mathrm{sm}(a,b) &= \left(a^{-n_{sm}} + b^{-n_{sm}}\right)^{-1/n_{sm}} \\ \tau_{sh} &= (1 - w_{slip} W)\,\mathrm{sm}(\mu p_{sh},\ \tau_{stick}) \\ \tau_{pin} &= (1 - w_{slip} W)\,\mathrm{sm}(\mu p_{pin},\ \tau_{stick}) \end{aligned} \tag{3.4}$$

where τ_{stick} is the von Mises shear yield with a strain-rate exponent m carried on spindle speed, $\mu = \mu_0(1 - 0.35W)$ is the sliding friction coefficient, p_{sh} and p_{pin} are the contact pressures of Equation (3.6), n_{sm} sharpens the smooth minimum and w_{slip} derates both branches for tool wear. Cold metal is strong and the indentation pressure under the shoulder is large, so μp_{sh} is the smaller term and the contact slides. Once the weld is running the metal is soft and the shear yield is the smaller term, so the contact sticks [5]. Both regimes are logged, and at the design point the sticking branch is the smaller of the two, with $\tau_{stick} = 34.76$ MPa against $\mu p_{sh} = 35.16$ MPa, so that Equation (3.4) returns $\tau_{sh} = 29.40$ MPa under the shoulder.

3.3.3 Heat Generation and Spindle Torque

Frictional and plastic work is the Schmidt integral of surface velocity against interface shear stress over the three contact surfaces of a flat shoulder with a cylindrical pin, as shown in Equation (3.5) [5, 15].

$$\begin{aligned} Q_{sh} &= \tfrac{2}{3}\pi\,\tau_{sh}\,\Omega\left(R_s^3 - R_p^3\right) f_s \\ Q_{tip} &= \tfrac{2}{3}\pi\,\tau_{pin}\,\Omega\,R_p^3\,f_{tip} \\ Q_{side} &= 2\pi\,\tau_{pin}\,\Omega\,R_p^2\,L_p \\ Q &= Q_{sh} + Q_{tip} + Q_{side}, \qquad \mathscr{T} = Q/\Omega \end{aligned} \tag{3.5}$$

where $\Omega = 2\pi\omega/60$ is the spindle speed in rad/s and $\mathscr{T}$ is the spindle torque in N m. At the design point Equation (3.5) gives 1755 W and 18.62 N m, of which the shoulder makes 68.8 %, the pin side 28.1 % and the pin tip 3.1 %, which is 877 J into every millimeter of joint at 120 mm/min. Because τ in (3.5) falls with temperature through (3.3), the response to spindle speed is strongly sublinear. Doubling the spindle from 450 to 900 rpm raises the steady heat generation only from 1405 to 1755 W, an increase of 24.9 %, while the flow stress of the shear layer falls from 103.5 to 59.4 MPa and the interface shear stress under the shoulder from 47.0 to 29.4 MPa.

3.3.4 Contact Pressure, Axial Force and the Stalled Axis

The axial force is an indentation problem, and an indentation is resisted by the metal a few millimeters out and a few millimeters down rather than by the thin shear layer at the interface. The load-bearing temperature is therefore taken mostly from the halo, as shown in Equation (3.6).

$$\begin{aligned} T_{bulk} &= w_b T_w + (1 - w_b) T_n \\ p_{sh} &= C_{ind}\, \sigma(T_{bulk}) \left(1 + \kappa \frac{\max(0, z)}{z_0} \right) \\ p_{pin} &= C_{pin}\, \sigma(T_{bulk}) \\ F &= p_{sh} A_{sh} + p_{pin} A_{tip} \end{aligned} \tag{3.6}$$

where w_b is the weight of the halo temperature in the load-bearing metal, C_{ind} and C_{pin} are indentation pressures as multiples of the flow stress, κ is the rise in shoulder pressure per reference plunge depth z_0, and F is the axial force. When F exceeds the machine limit F_{max} both pressures are scaled by F_{max}/F, the force is held at F_{max} and the downward motion of the axis is blocked. Otherwise the axis is throttled by the remaining force margin through the stall factor of Equation (3.7).

$$s = \operatorname{clamp}\left(\frac{F_{max} - F}{F_{stall}}, 0, 1 \right) \tag{3.7}$$

Equations (3.6) and (3.7) are what make the plunge a different plant. Cold 6061 under a 12 mm shoulder needs more than the 20 kN the machine has, so the axis stalls against its limit and sinks only as fast as the metal softens. A plunge commanded at 120 mm/min, three seconds of travel, takes 12.2 s with the axial force pinned at 19.99 kN and a peak spindle torque of 35.9 N m. Through those twelve seconds an extra revolution per minute buys depth as well as heat, and that positive feedback is the source of the extra plunge gain of Section 3.4. Setting the adjustable weight w_b to zero moves the load onto the shear layer and removes the force-limited plunge and the regime change with it.

3.3.5 The Four Thermal Nodes

Energy balances on the shear layer, the tool, the halo and the anvil interface are Equations (3.8) to (3.11).

$$C_n(a_f) \frac{\mathrm{d}T_n}{\mathrm{d}t} = Q - q_t - q_w - q_b \tag{3.8}$$

$$C_t \frac{\mathrm{d}T_t}{\mathrm{d}t} = q_t - U_{ta} (T_t - T_{amb}) \tag{3.9}$$

$$C_w \frac{\mathrm{d}T_w}{\mathrm{d}t} = q_w - U_{wf} (T_w - T_\infty) \tag{3.10}$$

$$C_b \frac{\mathrm{d}T_b}{\mathrm{d}t} = q_b - U_{ba} (T_b - T_{amb}) \tag{3.11}$$

where the three heat flows out of the shear layer are $q_t = U_{nt} a_f (T_n - T_t)$, $q_w = U_{nw} a_f (T_n - T_w)$ and $q_b = U_{nb}^{\mathrm{eff}} (T_n - T_b)$, the capacities are in J/K and the conductances in W/K. Both sides of each equation carry W. The shear-layer capacity grows with engagement, $C_n(a_f) = C_{n,\min} + C_{n,stir}\, a_f$, so

it is 0.70 J/K with the pin tip alone in contact and 3.64 J/K at full engagement. At the design point the 1755 W generated leaves as 69.6 % into the plate, 22.6 % into the anvil and 7.7 % up the tool.

The two conductances that are not constant carry the moving heat source and the clamp-up, as shown in Equation (3.12).

$$\begin{aligned} U_{wf} &= U_{wf,0}\, g_v \left(1 - d_e \mathrm{e}^{-\ell / L_e}\right), \\ g_v &= \mathrm{clamp}\left(1 + c_V \left(\frac{v}{v_{ref}} - 1\right), 0.3, 3\right) \\ U_{nb}^{\mathrm{eff}} &= U_{nb}\, a_d\, \mathrm{clamp}\left(1 - \frac{c}{1.5\,\mathrm{mm}}, 0, 1\right) \mathrm{clamp}(a_f, 0, 1.2) \end{aligned} \tag{3.12}$$

where $\ell = \max(0, L_{plate} - x_{start} - x)$ is the distance from the tool to the trailing edge of the plate, c_V is the rise in plate conductance per unit of traverse speed, d_e is the loss of conductance at the edge and L_e is the distance over which the edge is felt. The first factor is the moving source. A faster tool spreads its heat over more cold metal per millimeter of weld, so going faster makes the weld colder, and the steady tool temperature falls from 490.5 °C at 60 mm/min to 446.9 °C at 240 mm/min at a fixed 900 rpm. The second factor is the run-out. Between the middle of the joint, 155 mm from the plate edge, and the last millimeter of the weld, 25 mm from it, the plate conductance falls from 3.31 to 1.70 W/K and the steady tool temperature rises from 470.8 to 502.4 °C with nothing on the machine having changed. Scaling the anvil path by a_f matters more than it appears, because the anvil takes almost a quarter of the heat in a running weld and leaving that path at full strength during the plunge would make the plunge a much stiffer plant than it is.

3.3.6 Machine Axes, Wear and the Instrument Chain

The spindle and traverse drives are first order, and the axial axis is a first order servo with separate limits in each direction and the stall factor of (3.7) on the way down, as shown in Equation (3.13).

$$\begin{aligned} \tau_\omega \frac{\mathrm{d}\omega_A}{\mathrm{d}t} &= \omega_{cmd} - \omega_A, \qquad \tau_v \frac{\mathrm{d}v_A}{\mathrm{d}t} = v_{cmd} - v_A \\ \dot{h}^* &= (h_{cmd} - h_s)/\tau_z \\ \frac{\mathrm{d}h_s}{\mathrm{d}t} &= \begin{cases} -s\,\mathrm{clamp}\left(-\dot{h}^*, 0, \dot{z}_{max}\right), & \dot{h}^* < 0 \\ \mathrm{clamp}\left(\dot{h}^*, 0, \dot{z}_{ret}\right), & \dot{h}^* \geq 0 \end{cases} \\ \frac{\mathrm{d}x}{\mathrm{d}t} &= \begin{cases} v_A/60, & x < L_{weld} \\ 0, & \text{otherwise} \end{cases} \\ \frac{\mathrm{d}W}{\mathrm{d}t} &= k_W \frac{\omega}{\omega_{ref}} \frac{\max\left(0,\ T_n - 450\,^\circ\mathrm{C}\right)}{100\,\mathrm{K}} \end{aligned} \tag{3.13}$$

where $\dot{z}_{max}$ and $\dot{z}_{ret}$ are the commanded plunge and retract rates in mm/s and k_W is the wear rate. Wear accrues only above 450 °C in the shear layer, so running hot is expensive twice, and at half life the steady tool temperature falls from 471.4 to 436.7 °C, two degrees above the cold edge of the window.

The thermocouple sits in a rotating tool a few millimeters behind the shoulder, so its reading reaches the controller through a transport delay and its own bead time constant, as shown in Equation (3.14).

$$\tau_{TC} \frac{\mathrm{d}T_m}{\mathrm{d}t} = T_t(t - \theta_{TC}) + n(t) - T_m \tag{3.14}$$

where θ_{TC} is the telemetry delay off the spindle, τ_{TC} is the bead time constant and n is an optional noise sequence made by filtering a Gaussian with a 0.4 s time constant and rescaling it so the stationary standard deviation matches the selected value. The chain is stepped at each execution of the regulatory controllers, so the reading is held for 0.2 s between executions. The lag uses the exact

hold expression $T_{m,k} = \alpha T_{m,k-1} + (1-\alpha)(\cdot)$, in which $\alpha = \exp(-0.2\,\mathrm{s}/\tau_{TC})$, and the transport delay is a ring buffer of those samples read by linear interpolation, which applies the 0.60 s delay exactly on that grid. Together the two account for most of the apparent dead time in the identified models, and both sit inside the loop the reader tunes. The axial load cell is a first order lag of time constant τ_F on F with no delay.

Pause. Try Exercise 3.2 before reading on.

3.3.7 Weld Quality

The quality window is a band in temperature whose edges move with traverse speed, and the defect classification is a set of thresholds on the tool temperature and the root clearance, as shown in Equation (3.15).

$$\begin{aligned} T_{lo}(v) &= T_{lo,0} + s_{lo}\left(v - v_{ref}\right) \\ T_{hi}(v) &= \min\left(T_{sol} - 15,\; T_{hi,0} + s_{hi}(v - v_{ref})\right) \end{aligned} \tag{3.15}$$

Faster travel needs hotter metal to keep the flow closing behind the pin, so s_{lo} is four times s_{hi} and the window narrows as the production rate rises, from 70 °C wide at 120 mm/min to 52 °C at 240. A millimeter of weld is scored as sound when $T_{lo} < T_t < T_{hi}$ and the root clearance c lies between the anvil strike limit of 0.05 mm and the lack-of-penetration limit of 0.45 mm. Otherwise it is recorded as a wormhole, as flash, as a root flaw or as an anvil strike, accumulated by length along the joint and drawn on the plan view in the color of that defect. The record does not heal, because the metal that should have filled the void is already behind the tool.

3.3.8 Parameters, Initialization and Numerics

The constants are collected in Table 3.2. Every one of them is exposed in the model panel and is adjustable, apart from the bulk aluminum properties, the sharpness of the smooth minimum and the mid-weld start position, which are held back as reference values.

A cold scenario starts with the tool clear of the plate at $h_s = t + H_p + 0.15\,\mathrm{mm}$ and every node at plate temperature, and the machine program runs from the plunge. A mid-weld scenario starts at the quasi-steady condition 90 mm along the joint, which is obtained by solving the shear-layer energy residual rather than by marching to it. That residual is strictly decreasing in T_n, because heat generation falls with temperature through (3.3) while every loss term in (3.8) rises, so the root is unique and sixty bisections of the interval between ambient and the liquidus locate it to a nanokelvin from any starting point. A damped fixed point was tried first and it oscillated above about 1200 rpm, where the solidus collapse in (3.3) makes the map non-contractive.

The engine of Section 1.6.1 advances the states with fourth-order Runge-Kutta on a fixed 0.02 s step. The regulatory controllers execute every 0.2 s and the predictive controllers every 1.0 s, and ten samples per simulated second are logged. The exported file holds the true tool temperature beside the measured one, both identified models, the predictive controller's own bias, the defect record along the joint and the measured cost per control execution. The noise seed is 20260906.

3.3.9 Assumptions

Detail is spent where the physics forces a control decision and withheld everywhere else, and every simplification below fails somewhere.

Four thermal nodes, not a moving-source field solution. The travelling thermal field appears only through the traverse-dependent conductance in (3.12), calibrated so that the steady temperature falls the right amount with traverse speed. There is no isotherm map, no heat-affected-zone width and no advancing to retreating asymmetry.

Heat generation is the flat-shoulder integral. Equation (3.5) assumes a flat shoulder and a cylindrical pin sweeping a full circle at the local surface velocity. Real tools are threaded or scrolled, and the effective slip at the interface is the largest single uncertainty in every published thermal model of the process [5, 15]. No slip factor and no shear-layer thickness appear here.

Table 3.2 Model Parameters. All Values Are Adjustable in the Application Except the Three Marked Hidden, Which Are Never Displayed.

Symbol	Meaning	Value	Unit	Visibility
R_s, R_p, H_p	shoulder radius, pin radius, pin length	6.0, 2.0, 6.00	mm	adjustable
t	nominal plate thickness	6.35	mm	adjustable
L_{plate}, L_{weld}	plate and programmed weld length	300, 245	mm	adjustable
x_{start}	plunge from the leading edge	30	mm	adjustable
σ_0, T_σ, σ_{res}	flow stress: factor, decay, residual	1329, 158.8, 8	MPa	adjustable
T_{sol}, T_{liq}	solidus and liquidus	582, 652	°C	adjustable
m, ω_{ref}	rate exponent, reference spindle	0.12, 800	rpm	adjustable
μ_0, w_{slip}	sliding friction, shear loss at wear	0.52, 0.30	—	adjustable
w_b	halo weight in the load-bearing metal	0.80	—	adjustable
C_{ind}, κ, z_0	shoulder pressure, rise, reference depth	0.265, 1.49, 0.15	mm	adjustable
C_{pin}, z_{eng}	pin pressure, full shoulder contact	2.10, 0.080	mm	adjustable
U_{nt}, U_{ta}	shear layer to tool, tool to spindle	3.00, 0.302	W/K	adjustable
U_{nw}, $U_{wf,0}$	shear layer to halo, halo to plate	10.1, 3.37	W/K	adjustable
U_{nb}, U_{ba}	shear layer to anvil, anvil to table	2.42, 1.30	W/K	adjustable
c_V	plate conductance per traverse speed	0.66	—	adjustable
d_e, L_e	edge conductance loss and length scale	0.78, 55	mm	adjustable
$C_{n,\min}$, $C_{n,stir}$	pin-tip capacity, added at contact	0.70, 2.94	J/K	adjustable
C_t, C_w, C_b	tool, halo and anvil capacity	27.6, 30, 69	J/K	adjustable
F_{max}, F_{stall}	machine force limit, stall margin	20.0, 4.0	kN	adjustable
τ_ω, τ_v, τ_z	spindle, traverse and z-axis lags	0.25, 0.40, 0.15	s	adjustable
$\dot{z}_{max}$, $\dot{z}_{ret}$	commanded plunge and retract rate	120, 300	mm/min	adjustable
z_{target}, t_{dwell}	programmed plunge, dwell	0.15, 12	mm, s	adjustable
k_W	wear rate above 450 °C	6×10^{-5}	1/s	adjustable
$T_{lo,0}$, $T_{hi,0}$	quality window at v_{ref}	435, 505	°C	adjustable
s_{lo}, s_{hi}	window edge slopes	0.20, 0.05	°C min/mm	adjustable
v_{ref}	reference traverse speed	120	mm/min	adjustable
τ_{TC}, θ_{TC}	thermocouple bead, telemetry delay	1.20, 0.60	s	adjustable
τ_F	axial load cell	0.15	s	adjustable
ρ, c_p, k	aluminum properties	2700, 896, 167	SI	hidden
n_{sm}	smooth-minimum order	4	—	hidden
x_{run}	mid-weld start	90	mm	hidden

The flow stress is two parameters and a residual. Equation (3.3) is not a Zener-Hollomon or Johnson-Cook law. There is no strain or strain-rate history state, and the entire rate dependence is the single exponent m in (3.4). The fit is to hot-working data at one order of magnitude of strain rate [4], and it is extrapolated outside it.

The load-sharing split is a modeling decision. Taking the axial load from T_{bulk} in (3.6) and the interface shear from T_n in (3.4) is defensible, because indentation is resisted by bulk material and shear by the interface layer, but the weight w_b is calibrated rather than measured, and it produces the plunge force spike and most of the extra plunge gain.

The quality window is an empirical band with straight edges. Real defect formation depends on material flow around the pin, thread pitch and direction, and tilt angle, none of which is represented [1]. The window is calibrated to temperature ranges reported for 6xxx alloys and is a teaching device rather than a process map a weld could be qualified against.

Tool wear is one scalar. It shrinks the shoulder and derates the interface shear stress, and it accrues linearly above a threshold. Real wear in aluminum is slow, geometry dependent and dominated by loss of the pin thread.

The numerics are more repeatable than a machine. One trajectory comes back on every computer, whereas the controller in a real machine carries jitter, a coarser measurement resolution and a scan order that moves.

Table 3.3 Identified First-Order-Plus-Dead-Time Models. The Top Four Rows Are Each a 100 rpm Step Differenced Against a Held Run. The Last Two Are the Models the Controllers Carry, Which Are Baselined Differently. The Residual Is the Root Mean Square Fit Error in the Same Units as K_p.

Identified at	K_p (°C/rpm)	τ_p (s)	θ_p (s)	residual
Traverse, +100 rpm	0.1425	12.24	2.31	0.0006
Traverse, −100 rpm	0.1569	12.55	2.34	0.0006
Settled dwell, 700 rpm	0.1782	15.81	2.45	0.0008
Plunge, shoulder contact	0.2401	1.17	0.00	0.0179
Traverse model in the controller	0.1356	11.78	2.68	—
Plunge model in the controller	0.2552	1.92	2.10	—

3.4 Three Plants in One Weld

Identification is the whole subject of the application, so the experiments are described as experiments rather than as linearizations. Each one is an open-loop step test in spindle speed on the shipped nonlinear model, made in Manual, differenced against a second run held at the base speed so that the slow baseline drift cancels, normalized by the 100 rpm step size and fitted by least squares to the first-order-plus-dead-time form of Equation (3.16). The traverse steps are taken on the long plate of scenario 4 switched to Manual, stepped at 20 s and fitted over 300 s, which reproduces the step of scenario 3 to within 0.002 °C. The dwell step goes from 700 to 800 rpm after 120 s of the 240 s dwell of scenario 3b and is fitted to the end of the dwell. The plunge experiment compares two scenario 2 plunges at 900 and 1000 rpm, the faster speed set before the run so that its drive turns at 1000 rpm from the start, over the window from shoulder contact, 8.56 s into the slower run, to the start of the faster run's traverse 14.8 s later. The thermocouple reading is sampled every 0.04 s, and time is counted from the controller execution at which the new speed reached the drive.

$$\frac{\Delta T_m(s)}{\Delta \omega(s)} = \frac{K_p\, \mathrm{e}^{-\theta_p s}}{\tau_p s + 1} \tag{3.16}$$

The fit minimizes the sum of squared residuals over K_p, τ_p and θ_p jointly, with K_p solved analytically inside a search over the other two. Results for four operating points are collected in Table 3.3 and the two extreme responses are plotted in Figure 3.2.

Three results in Table 3.3 matter for the controller. The first is the size of the change between the plunge and the traverse. The gain rises from 0.1425 to 0.2401 °C/rpm, a factor of 1.68, and the time constant falls from 12.24 to 1.17 s, a factor of 10.5, between two phases of one weld. The gain change comes from the stalled axis of (3.7), where an extra revolution per minute buys depth and therefore contact area, and the time-constant change from the capacity term $C_n(a_f)$ in (3.8), a fifth of its fully engaged value while the pin is still entering.

The second is the asymmetry within one operating point. Stepping the same condition down instead of up returns 0.1569 against 0.1425 °C/rpm, ten percent more gain for a step of the same size in the other direction, which is the plant reporting that it is not linear in either direction. The settled dwell at 700 rpm is a third point again, 0.1782 °C/rpm with a 15.81 s time constant, higher gain and a slower response than the weld because with no traverse to carry heat into cold metal the sink conductance in (3.12) is smaller and every watt is worth more degrees. That third measurement is worth having because it shows the fast 1.17 s time constant belongs to the plunge transient rather than to a stationary tool.

The third is the residual. The three fits taken at a settled operating point have residuals of 0.0006 to 0.0008 °C/rpm, and the plunge fit has 0.0179, twenty to thirty times larger. Its dead time is held at zero because the unconstrained optimum, −0.11 s from shoulder contact, is not physical. Figure 3.2 shows why. The differenced plunge response is not a monotone first-order rise. It climbs to 0.2610 °C/rpm

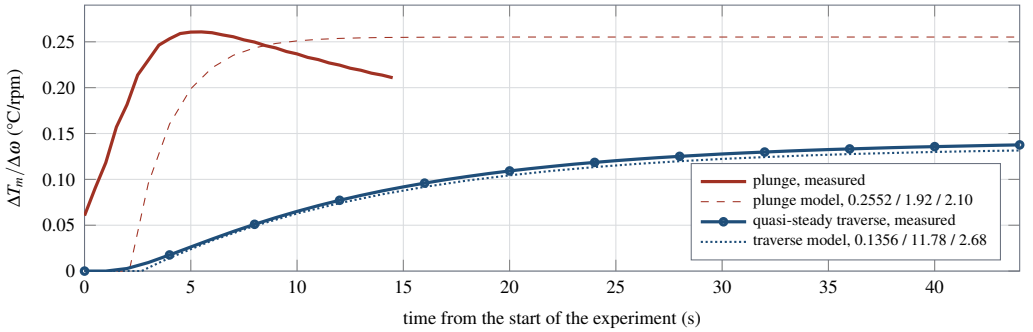


Figure 3.2 The two step tests behind the predictive controllers. Each curve is the difference between a run stepped by one hundred revolutions per minute and a run held at the base speed, divided by the step size. The traverse experiment is timed from the step and the plunge experiment from shoulder contact. The measured plunge response is the thick solid line and the measured traverse response the thick line with circles. The two models whose step responses fill the dynamic matrices are the thin dashed line (plunge) and the dotted line (traverse).

5.3 s after the shoulder engages and then decays to 0.2098 by the end of the dwell as the halo warms and takes the extra pressure back out of (3.6). Fitting a first-order model to that shape is what a commissioning engineer would do with the trend in front of them, and it means the plunge-identified controller is handicapped by model structure as well as by operating point.

The two models the predictive controllers actually carry are the last two rows of Table 3.3, and the two rows are not baselined the same way. The traverse model comes from the same experiment fitted *without* the held-run difference, over a window long enough that the held run's own drift of –0.53 °C per 300 s biases the gain low and the apparent dead time high, which returns 0.1382 °C/rpm, 11.10 s and 2.60 s, within two percent of the stored gain. The plunge model cannot be read that way at all, because a plunge from a cold plate is a large transient with or without the step, so its stored gain of 0.2552 °C/rpm matches the peak of the differenced response in Figure 3.2, 0.2610, to within three percent rather than the least-squares asymptote of 0.2401. Both sets are defensible readings of the same trends and they differ by five percent on the traverse gain and six percent on the plunge gain, which changes nothing in the argument because the ratio between the two regimes is 1.9 on the stored constants and 1.7 on the differenced fits. The stored values are quoted wherever a controller's internal model is meant and the differenced values wherever the plant itself is meant.

3.5 Controllers

Tool temperature is the controlled variable and spindle speed is the manipulated variable in all five architectures, so that the comparison is between designs rather than between actuators. Traverse speed belongs to the operator, because it is also the production rate, and the axial axis is a second loop of its own. A friction stir weld is a sequence rather than an operating point, so the temperature program holds a soak setpoint of 330 °C through the plunge and the dwell and hands over to the welding setpoint of 470 °C when the tool starts moving. Without that program the plunge would be twelve seconds of saturated actuator and the two identified models could not be compared in the regime where they disagree.

🕮 **Pause.** Try Exercise 3.1 before reading on.

3.5.1 The Two PI Loops

Both PID controllers are the velocity-form algorithm of Section 1.6.2, with the derivative taken on the filtered measurement and anti-windup by back-calculation on the integral term alone. Equation (3.17)

writes it in the symbols of this loop, where s_k is the running sum S_k of Equation (1.15) and c_k is the output before limiting, v_k of Equation (1.16).

$$\begin{aligned} e_k &= T_{\mathrm{SP},k} - T_{m,k}, \qquad P_k = K_c e_k \\ D_k &= -K_c \tau_D \left. \frac{\mathrm{d}T_m^f}{\mathrm{d}t} \right|_k \\ s_k &= s_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ c_k &= s_k + I_{k-1}, \qquad u_k = \mathrm{clamp}(c_k,\ 200,\ 1500) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - c_k}{T_t} \right] \Delta t_c \end{aligned} \tag{3.17}$$

where u is spindle speed in rpm, $\Delta t_c = 0.2\,\mathrm{s}$ is the controller period, c_k is the speed the algorithm demands, I is the integral state and $T_t = \tau_I$ is the tracking time of the back-calculation, so a clipped command sheds $\Delta t_c / \tau_I$ of its excess at each execution. There is no bias term and no slew limit, and the clamp acts on the emitted command only. Both loops are reverse acting in the instrument sense, with the sign convention set by a direction flag of $+1$, since a tool below its setpoint calls for a faster spindle on a plant of positive gain. The derivative estimate uses a first-order filter of 2.0 s on the measurement and both defaults set $\tau_D = 0$, so derivative action is available but off. Under manual operation I_k is instead back-solved as $u_k - s_k$, which pins the output to the operator's spindle speed and removes the need for any separate transfer logic in either direction. The loop that is not driving the spindle is fed the applied speed the same way every execution, so switching architecture mid-weld costs nothing. A reset is different. It seeds both loops with $s_0 = 0$ and $I_0 = 900\,\mathrm{rpm}$, the manual station value, so a weld started already in Auto from a cold plate adds the whole proportional term $K_c e_0$ at the first execution, with $e_0 = 308\,^{\circ}\mathrm{C}$ against the soak setpoint, and the command saturates at 1500 rpm. A loop that enters Auto from Manual has been tracking the station and takes over from 900 rpm without a bump, provided the switch is made after the run has started, since a switch made before the run is part of the starting conditions and is a start in Auto. Section 3.6 measures what that difference costs.

The two tunings are the ITAE setpoint and ITAE disturbance correlations for a first-order-plus-dead-time plant [16], applied to the stored traverse model of Table 3.3, as shown in Equations (3.18) and (3.19).

$$K_c K_p = 0.586 \left(\frac{\theta_p}{\tau_p} \right)^{-0.916}, \quad \tau_I = \frac{\tau_p}{1.03 - 0.165\,\theta_p/\tau_p} \tag{3.18}$$

$$K_c K_p = 0.859 \left(\frac{\theta_p}{\tau_p} \right)^{-0.977}, \quad \tau_I = \frac{\tau_p}{1.484} \left(\frac{\theta_p}{\tau_p} \right)^{0.680} \tag{3.19}$$

With $K_p = 0.1356\,^{\circ}\mathrm{C/rpm}$, $\tau_p = 11.78\,\mathrm{s}$ and $\theta_p = 2.68\,\mathrm{s}$, so that $\theta_p/\tau_p = 0.2275$, Equation (3.18) returns $K_c = 16.77\,\mathrm{rpm}/^{\circ}\mathrm{C}$ and $\tau_I = 11.87\,\mathrm{s}$, and Equation (3.19) returns $K_c = 26.91\,\mathrm{rpm}/^{\circ}\mathrm{C}$ and $\tau_I = 2.90\,\mathrm{s}$. The application ships those rounded to 16.8 with 11.9 s and to 26.9 with 2.9 s. The disturbance tuning carries sixty percent more proportional gain and four times the reset action, on the same plant and from the same step test, which is the whole of the difference between the two entries.

The axial force loop is a PI controller of the same form whose output is the commanded shoulder plunge depth, with $K_c = 0.012\,\mathrm{mm/kN}$, $\tau_I = 1.6\,\mathrm{s}$ and an output range of 0 to 0.45 mm. It is reverse acting for the same reason the temperature loop is, because more depth means more force. In position mode the depth command is the programmed value referenced to the machine table and the nominal plate thickness, which is what a position-controlled machine actually holds, and the force controller is tracked so that either mode can be selected without a bump.

3.5.2 Step-Response Predictive Control

Both predictive entries are dynamic matrix control in its original step-response form [7, 17], with the gain matrix formed once and reused. The two entries share every line of code, every weight and the same horizon. Only the three constants of the step response differ. Section 1.4.2 builds the same predictor from a step response, and the gain matrix of Equation (3.21) below is Equation (1.5) with A in place of G, λ in place of R and the columns of A set by move blocking.

Three quantities are built when the controller is created. Because the identified model is (3.16), its unit step response is written down rather than integrated, as shown in Equation (3.20).

$$a_k = \begin{cases} K_p\left(1 - \mathrm{e}^{-(kt_m - \theta_p)/\tau_p}\right), & kt_m > \theta_p \\ 0, & \text{otherwise} \end{cases} \qquad a_0 = 0 \tag{3.20}$$

where $t_m = 1.0\,\mathrm{s}$ is the execution period and k runs from 0 to $P+1$. The prediction horizon is $P = 40$ points, which is 40 s, and the move horizon is $M = 3$ after blocking the moves at samples $n_j \in \{0, 2, 6\}$, so three decisions cover forty samples [7]. The dynamic matrix and the constant gain matrix of Equation (3.21) follow.

$$A_{ij} = \begin{cases} a_{(i+1)-n_j}, & i+1 > n_j \\ 0, & \text{otherwise} \end{cases}$$
$$K = \left(A^{\mathsf{T}} Q A + \lambda I\right)^{-1} A^{\mathsf{T}} Q \tag{3.21}$$

where A is 40×3, the diagonal of Q carries unit weight over the first thirty points and a terminal weight $w_T = 3.0$ over the last ten, and K is 3×40. Inverting the 3×3 takes a hand-coded Gauss-Jordan pass and is the only linear algebra anywhere in the application. One revolution per minute is worth a seventh to a quarter of a degree, so the diagonal of $A^{\mathsf{T}} Q A$ runs from 0.58 to 0.71 for the traverse model and from 3.2 to 3.6 for the plunge model, five times larger. The move suppression is therefore scaled against the tracking term, as shown in Equation (3.22).

$$\lambda = \lambda_{rel} \frac{1}{M} \operatorname{tr}\left(A^{\mathsf{T}} Q A\right) \tag{3.22}$$

so that one slider position means the same aggressiveness for both models, which is what makes the comparison between them fair. The default $\lambda_{rel} = 0.08$ evaluates to $\lambda = 0.0523$ on the traverse model and $\lambda = 0.2754$ on the plunge model. Equation (3.21) is formed again only if λ_{rel} or w_T is moved, and Equation (3.20) is never recomputed.

Each execution begins by charging the prediction with the speed change the spindle received rather than the change that was requested, as shown in Equation (3.23).

$$Y_i \leftarrow Y_i + a_i\left(\omega_{cmd} - \omega_{cmd}^{prev}\right), \qquad i = 0 \ldots P \tag{3.23}$$

A move cut short by the clamp, one the rate limit has not finished delivering, or one that never occurred at all because another architecture held the spindle, enters Equation (3.23) at whatever size it really had, and nothing is ever charged in advance of being delivered. That single substitution supplies all of the anti-windup this controller has, and it also keeps the inactive model anchored to the plant, so either model can be selected in mid-weld without a bump. The prediction is then shifted one sample and its tail held at the last value, $Y_i \leftarrow Y_{i+1}$ for $i < P$, which is the correct extension for a self-regulating output and would be wrong for an integrating one. The plant correction and the reference trajectory are Equation (3.24).

$$\begin{aligned} b_k &= T_m - Y_0, & Y_i &\leftarrow Y_i + b_k \\ r_0 &= T_m, & r_{i+1} &= T_{\mathrm{SP}} + (r_i - T_{\mathrm{SP}})\,\mathrm{e}^{-t_m/\tau_{ref}} \\ e_i &= r_{i+1} - Y_{i+1}, & i &= 0 \ldots P-1 \end{aligned} \tag{3.24}$$

where b_k is a constant output-disturbance estimate added across the whole horizon and $\tau_{ref} = 2.0\,\mathrm{s}$ shapes the approach to a new setpoint. The constant b_k is what removes steady offset, as in Section 1.4.5, and it does for this controller what reset action does for the PI loops. On an integrating output the same constant bias leaves an offset, which Chapter 2 measures on a balloon's altitude. It is also the reason a controller carrying a gain error of nearly a factor of two produces a slowly corrected offset instead of ringing, and that robustness is measured in Section 3.6.

What remains is one matrix-vector product, $\Delta u = Ke$. The first blocked move is rate limited to $\Delta u_{max} = 250\,\mathrm{rpm}$ per execution, clamped to the drive envelope of 200 to 1500 rpm and applied, which projects the unconstrained move onto the limits rather than solving the constrained program of Section 1.4.3. Nothing is written back into Y because the next execution reads the delivered change through (3.23). Both models execute on every cycle whatever architecture is selected. Timed over 100 000 executions on an Apple M-series laptop the online path costs 0.272 µs per execution for the traverse model and 0.274 µs for the plunge model. The application repeats that measurement at every reset and reports it on the scorecard, so the figure the reader sees is measured on the machine in front of them.

3.6 Results

Thirteen scenarios form a laboratory sequence, from a weld made by hand in scenario 1 to the graded challenge, with a second step test held in the dwell numbered 3b. Each of them presets the starting condition, the architecture, the three tunings, the disturbances and whatever timed events it needs, asks for a prediction before anything runs, and carries its own completion rule, as Section 1.6.4 describes. The challenge is judged on the case as given, so a run with a changed model parameter or disturbance does not count. Reset restarts a case from $t = 0$ and keeps every setting the reader changed, while reloading the case restores its shipped settings, so a setting chosen before the run and the same setting carried through Reset give one run.

Scenario 1 sets the problem before a controller is involved. Held at a fixed 900 rpm for the whole job the weld comes out 89.8 % sound with 24.9 mm of wormhole, all of it in the first 25 mm of the joint, because the tool leaves the dwell at 372 °C, 63 °C below the cold edge of the window, and takes 12.9 s of traverse to climb into it. A spindle speed that suits the plunge does not suit the weld, and the two identified models of Table 3.3 say by how much.

> **TRY IT IN THE APP** **1 · Weld It by Hand (MANUAL)**
>
> Load the scenario and press Run without touching the spindle. Watch the weld track drawn behind the tool. The tool leaves the dwell at 372 °C, below the window, and needs 12.9 s of traverse to climb into it, so all 24.9 mm of wormhole lies in the first 25 mm of the joint.

3.6.1 Five Architectures Over a Complete Weld

Table 3.4 runs all five architectures over the identical job, a cold plate, a plunge, a 12 s dwell, 245 mm of traverse at 120 mm/min and the run-out at the edge, and scores each phase separately. In the top five rows each architecture is selected and switched to Auto from the 900 rpm manual station of scenario 1 at the first controller execution after the run starts, 0.2 s in. A switch made before the run is part of the starting conditions and gives the start in Auto of the last two rows. The soak measure is integrated absolute error between the tool and the soak setpoint over the settled part of the dwell, from eight seconds after the axis reaches depth to the start of the traverse, which is a window of about four seconds. The traverse and run-out measures are integrated absolute error over those phases, 108 and 15 s long. The defect lengths and the sound fraction are read at the end of the weld and need no window.

The servo and regulator tunings change places between two phases of one weld, on the same loop and the same plant, when each enters Auto from the manual station. The plunge and the dwell are

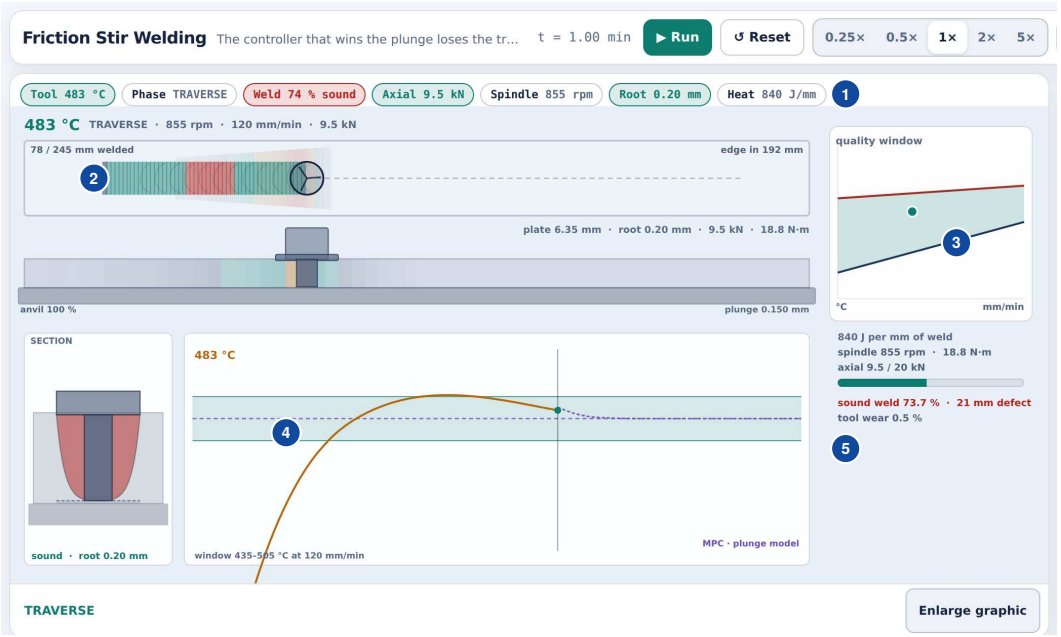


Figure 3.3 Scenario 9, the plunge-identified controller meeting the traverse, read 60 s into the run with the weld still being made, cropped to the header and the plant view. The tool stands at 483 °C against a setpoint of 470 and a window of 435 to 505 °C, and the controller is pulling the spindle back to 855 rpm, below the 891 rpm that holds 470 °C in quasi-steady welding. (1) The badges carry the tool temperature, the spindle speed, the root at 0.20 mm and a heat input of 840 J/mm. (2) The plan view records the consequence, 78 of 245 mm welded, with the flash drawn where the tool ran hot and the wormhole made as the traverse began, and (5) the readout under the inset gives 73.7 % of the weld sound and 21 mm of defect, of which the run's record puts 18 mm in flash and 2.5 mm in wormhole. (3) The inset places the operating point inside the quality window, and (4) the temperature plot shades the window behind the tool temperature. The scorecard and the trend strips, outside the crop, record a peak of 507 °C, a command that has spent 38.3 % of the run against a limit and accumulated 1316 rpm of travel, and an axial force of 9.55 kN against a 9.50 setpoint.

Table 3.4 Per-Phase Comparison Over One Complete Weld From a Cold Plate. In the Top Five Rows Each Architecture Enters Auto From the 900 rpm Manual Station Just After the Run Starts, and Each Column Has a Different Winner, Shown in Bold. The Last Two Rows Start the PI Loops Already in Auto, as Scenario 5 Ships.

Arch.	Soak IAE (°C s)	Into window (s)	Trav. IAE (°C s)	Run-out IAE (°C s)	Cold, hot (mm)	Sound weld (%)
Manual, 900 rpm	82	12.9	1453	186	24.9, 0	89.8
PI, servo tuned	**18**	8.1	817	34	15.4, 0	93.7
PI, regulator tuned	213	4.2	**777**	**17**	7.7, 0	96.9
DMC, traverse model	232	2.5	974	59	4.1, 0	**98.3**
DMC, plunge model	239	**0.0**	1311	74	**2.5**, 18.2	91.6
PI servo, Auto start	217	3.5	637	34	6.3, 0	97.4
PI reg., Auto start	212	4.2	773	17	7.6, 0	96.9

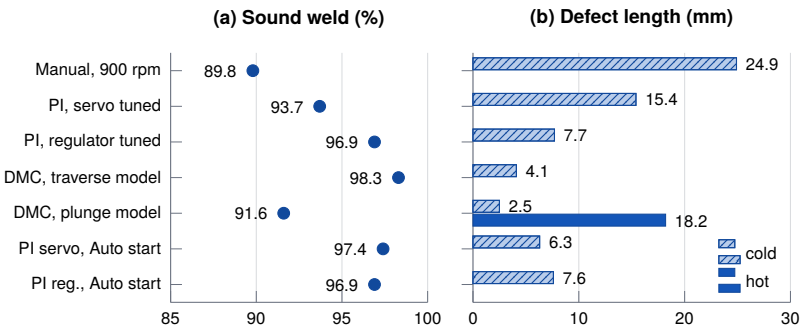


Figure 3.4 The weld-level columns of Table 3.4: (a) the sound fraction of the weld, as dots on an axis that starts at 85 %, and (b) the cold defect (hatched bars) and the hot defect (dark bars) of each architecture, in the table's row order. Notice that the plunge-model DMC leaves the shortest cold defect and the only hot one.

a setpoint change, and the ITAE setpoint tuning holds the soak to 18 °C s against the disturbance tuning's 213, a factor of twelve. The start of the traverse is a load change, because the tool moves into plate that has no thermal field in it, and there the disturbance tuning reaches the window in 4.2 s against 8.1, leaving 7.7 mm of cold defect against 15.4. A 0.08 mm thin patch in quasi-steady welding is a load change of the same kind. Set 15 s into scenario 4, it costs the regulator tuning 109 °C s of error over the following sixty seconds with a peak deviation of 6.1 °C, against the servo tuning's 160 and 8.3 °C. A setpoint step is not a load change, and stepping the setpoint from 470 to 500 °C at the same instant overshoots by 10.9 °C with the regulator tuning against 1.5 °C with the servo tuning. The disturbance tuning also pays in damping. After the step into the traverse it rings with a 16.7 s period, each swing about two thirds of the one before, from +17.2 °C at 34 s to ±2.7 °C by 109 s, inside the window throughout, while the servo tuning passes its setpoint once, by 1.3 °C. There is no single best tuning. If one PI loop has to be shipped, the one that is shipped is the one that is wrong in the phase that costs less, and that is a production decision rather than a control one. Figure 3.4 draws the sound fraction and the defect lengths of all seven rows.

How a loop enters Auto matters as much as how it is tuned. Scenario 5 starts the same weld already in Auto, so the first execution adds the whole proportional term on a 308 °C soak error to the 900 rpm station, as described with Equation (3.17), and both loops hold the spindle at 1500 rpm until about 16 s, while the lagging thermocouple catches up. Neither then holds the soak, at 217 and 212 °C s in the last two rows of Table 3.4. The tool leaves the dwell at 400 °C under servo tuning and 393 °C under regulator tuning, so the servo tuning now reaches the window first, in 3.5 s against 4.2, with 6.3 mm of cold defect against 7.6 and a weld 97.4 % sound against 96.9 %. That ordering comes from the hotter start rather than from better load rejection.

✦ DID YOU KNOW? **A Bad Start That Made a Better Weld**

From the simulator: started already in Auto, the servo-tuned PI loop saturates and spoils the soak, 217 °C s of error against 18 when it enters Auto after the run begins. Yet it leaves the sounder weld, 97.4 % against 93.7 % (Table 3.4).

3.6.2 The Controller That Wins the Plunge Loses the Traverse

The two DMC rows of Table 3.4 are the result the application exists for. The two controllers are the same code with three different numbers in Equation (3.20), and scenario 9 as shipped, which starts the plunge model already in Auto, reproduces its row to within 1.4 °C s and 0.2 mm.

The plunge-identified controller wins the plunge outright. It puts the tool inside the quality window at the very first millimeter of the traverse, which is 0.0 s against 2.5 for the traverse-identified controller, 4.2 for the regulator PI and 8.1 for the servo PI, and it leaves 2.5 mm of cold defect at the start against 4.1 and 15.4. For the first thirty millimeters of joint it is the best controller in the application.

It then loses the traverse, and it loses it as flash. Its internal model says the temperature will arrive where it was asked for in two seconds while the plant takes twelve, so it keeps adding spindle speed while very little appears to be happening on the thermocouple. Its integrated error through the traverse is 1311 °C s against 974, its peak tool temperature is 508 °C against 496, and it is the only one of the five architectures that makes a hot defect at all, 18.2 mm of it, laid down between 33 and 51 mm along the joint. That is not a metric that degraded. It is metal extruded out from under the shoulder and lost from the joint, visible on the plate and on the plan view of the animation in the color of the defect, as shown in Figure 3.3 sixty seconds into the run, and the weld finishes 91.6 % sound against 98.3 % for the identical controller carrying the other model.

The failure is a slowly corrected offset rather than an oscillation. Sampling the bias term of (3.24) at $t = 30$ s, nine seconds into the traverse, and at $t = 80$ s, the plunge model reads 5.54 °C and then −1.23 °C while the traverse model reads 4.91 and 0.15 °C. A constant output-disturbance estimate absorbs a gain error of 1.7 to 1.9 as an offset that decays over about a minute, which is the underrated robustness property of this controller form and also the reason a badly identified predictive controller can stay in service for years without drawing attention [8]. Every column of the top five rows of Table 3.4 has a different winner, and that is the argument for gain scheduling, for identification over several operating points, and for a nonlinear predictive formulation [6, 14].

🕮 **Pause.** Try Exercise 3.3 before reading on.

On the setpoint step of scenario 8, 470 to 500 °C at 15 s, the traverse-model controller is slower than the servo PI given the same step in scenario 4, and gentler. It overshoots by 0.8 °C against 1.5 °C and needs 21.2 s to stay inside ±1.5 °C against 8.5 s, and that ordering is not a fault. The servo PI answers the 30 °C step by throwing the spindle from 901 to 1458 rpm and pulling it back to 1122 rpm, while the traverse-model controller lifts it only to 1206 rpm and settles at 1123 rpm. On a machine where every excess revolution is tool wear above 450 °C, the distinction is worth having. The plunge-model controller overshoots the same step by 6.0 °C and takes 58 s to settle. The ordering turns against both predictive controllers on the thin patch, where they accumulate 315 and 410 °C s against the regulator PI's 109, because a reference trajectory shapes the approach to a setpoint and does nothing for a disturbance the model never saw.

3.6.3 Position, Force and the Root

Scenario 7 removes 0.10 mm of plate thickness with the temperature loop switched off, so that the disturbance can be seen directly. Under position control the axis holds its programmed height above the table and never learns that the plate got thinner. The shoulder plunge falls from 0.15 to 0.05 mm, the engaged shoulder area from 100 to 62 %, the heat generation from 1755 to 1470 W and the steady tool temperature from 471.4 to 414.3 °C, which is twenty-one degrees below the cold edge of the

window and a continuous wormhole. In the run itself the tool reaches 422.7 °C at 56 s before the approaching plate edge warms it again.

Force control holds the heat instead. The loop sinks the tool 0.098 mm to restore the contact pressure, so the steady heat generation is 1761 W and the tool sits at 470.2 °C, and in the run the tool dips only to 466.6 °C and is back at 471 °C by 40 s. What it spends is the root, and there is only 0.20 mm of clearance between the pin and the backing anvil. The clearance falls to 0.086 mm thirty seconds after the patch arrives and to 0.011 mm by sixty seconds. It passes the 0.05 mm anvil strike limit at 60.6 s and reaches zero at 74.3 s, with the tool at 491 °C, because the tool is entering the run-out where the metal is softer and more depth is needed to hold the force setpoint. Neither mode is safe alone, which is why production machines interlock one against the other [10, 11].

3.6.4 Disturbances Nobody Measures, and the Run-Out

A plate that lifts off its backing bar removes most of a path that carries almost a quarter of the heat leaving the shear layer. At an anvil contact of 0.30 the steady tool temperature rises from 471.4 to 485.7 °C at mid-plate, and at 0.15 it reaches 492.9 °C. Nothing on the machine measures it. At a contact of 0.30 the spindle speed that holds 470 °C is 88 rpm lower. The regulator PI gets there through a lightly damped swing, from 891 rpm at the event in scenario 10 to a trough of 616 rpm 39 s later and 803 rpm at 60 s, and the tool never passes 474.7 °C, against 507.4 °C and 5.4 mm of flash open loop at 900 rpm with the run-out on top of it. A temperature loop turns an unmeasured heat-sink failure into a visible actuator movement, which is an argument for closing the loop on a process that runs perfectly well open loop most of the time.

The run-out is the opposite kind of disturbance, because it is entirely predictable from a position the machine already knows. Every closed-loop architecture handles it. Over the complete weld of Table 3.4 the run-out costs 17 °C s under the regulator PI, 34 under the servo PI, 59 under the traverse-identified controller and 74 under the plunge-identified one, the reverse of the soak ordering for the two PI loops. Started at 170 mm in scenario 11, each loop begins in Auto with the tool 6 °C above the setpoint, and the regulator PI is still swinging by 4 °C from that start when the run-out begins at 22.6 s. There the same four score 45, 32, 53 and 49 °C s, and the servo PI is the best of them. Open loop at a fixed 900 rpm the tool reaches 487.6 °C at the last millimeter, which is still inside the window and leaves no margin for anything else, and adding the loose clamp to it puts flash into the end of the joint. Closed loop the margin can be spent instead of lost, and a scenario 11 run under the traverse-identified controller with its setpoint raised to 500 °C has the run-out carry the tool to 506.4 °C against a window top of 505, leaving 8.5 mm of flash over the last 8.5 mm of an otherwise sound joint.

3.6.5 The Graded Challenge

The challenge is a complete weld at 190 mm/min with a 0.09 mm thin patch arriving at 50 s, handed over with a servo loop detuned to $K_c = 2\,\text{rpm}/^\circ\text{C}$ and $\tau_I = 60\,\text{s}$. As handed it finishes 46.1 % sound with 132.1 mm of wormhole and fails by 48.9 points and 117.1 mm. The target is 95 % sound, under 15 mm of defect, and the pin still clear of the anvil.

The answer has three parts and only one of them is tuning. The traverse speed comes down, because at 190 mm/min the cold edge of the window has risen to 449 °C while the tool at a fixed spindle speed has fallen to 455.6 °C, which is 6.6 °C of margin against 31.8 °C at 130 mm/min, and the run-out still has to be survived at the top of the window. The temperature goes on the traverse-identified controller. With the axis in position it finishes 97.7 % sound with 5.6 mm of defect and a 0.200 mm root at 130 mm/min, passing by 2.7 points, 9.4 mm and 0.150 mm whether that configuration is revealed before the run or carried through Reset, while the regulator PI with its ITAE tuning at that speed reaches 96.4 % and 8.8 mm and also passes.

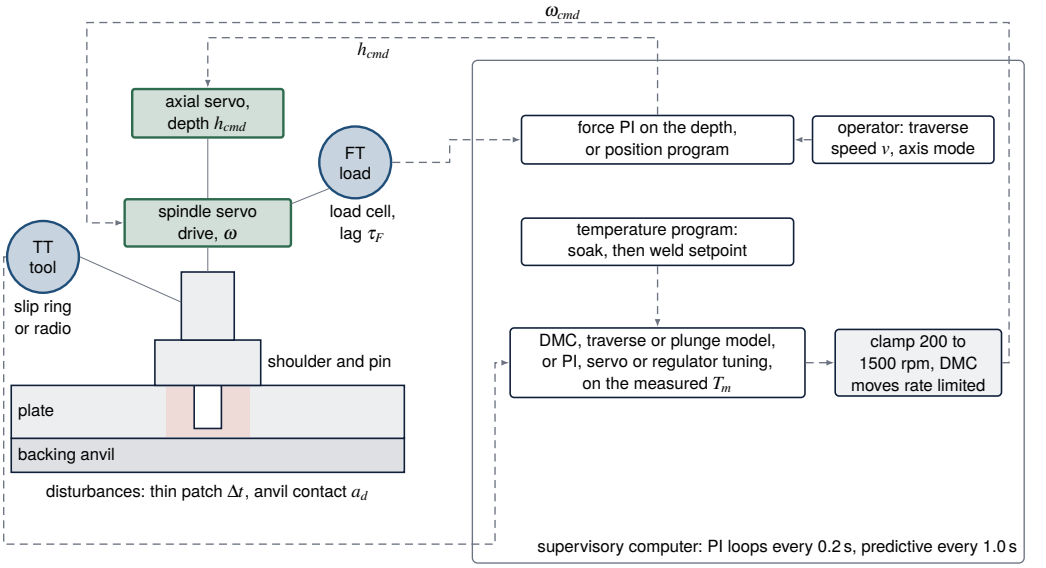


Figure 3.5 The chapter's loops on welding hardware. TT, a thermocouple in the rotating tool, feeds whichever temperature controller is selected, and its command sets the speed of the spindle servo drive. FT, the axial load cell, feeds the force loop, which moves the depth command of the axial servo, or the axis holds a programmed position. Notice that the thin patch and the anvil contact reach the controllers only through the tool temperature and the force.

> **TRY IT IN THE APP** **Challenge · A Complete Weld**
>
> Load the challenge and press Run as it is handed over. The detuned servo loop at 190 mm/min finishes 46.1 % sound with 132.1 mm of wormhole. Press Reset, apply the suggested configuration, which puts the traverse-identified predictive controller on the spindle at 130 mm/min with the axis in position, and run again: 97.7 % sound with 5.6 mm of defect.

The axial axis is the interesting choice, and the safe answer is to leave it in position. Force control is the obvious move, because it removes the thin patch as a heat disturbance entirely, and it finishes 96.2 % sound with 9.4 mm of defect, 5.7 mm of it at the end of the joint where the pin is inside the anvil strike limit. By the run-out, where the metal is softest and the force loop has to sink furthest to hold its setpoint, the root clearance has fallen to 0.040 mm against an anvil strike limit of 0.050. Force control fails the challenge on the root with a weld that would have passed every surface inspection.

3.7 Technology in Practice

What Measures It

Weld temperature is read in three ways. Thermocouples set in the workpiece give no signal a controller can use during the weld and cannot be reused. Thermocouples set in the tool can. The tool-workpiece thermocouple uses the voltage that arises between the steel tool and the aluminum workpiece and averages the temperature over the whole contact. Pyrometers and thermal cameras are troubled by the reflective aluminum and see only the edge of the shoulder [18]. After the weld, phased-array ultrasonics and eddy-current arrays look for the flaws that the chapter scores, wormholes and lack of penetration among them, and film radiography is kept for wormholes and larger flaws [3]. The chapter's TC is a thermocouple in the tool, and its defect record stands in for that inspection.

TECHNOLOGY **Sensor: A Thermocouple in the Rotating Tool**

The usual sensor is a thermocouple set in a hole drilled in the tool, in the shoulder or in the center of the pin, as close to the tool-workpiece interface as it can go so that it answers quickly. A slip ring or a wireless link carries its signal off the spinning tool. The steel around the junction and its ceramic coating make it read low: in one comparison a shoulder thermocouple read 125 °C and a pin thermocouple 70 °C below a reading averaged over the whole tool contact. The hole makes the tool costly, a small tool may have no room for it, and a junction that close to the interface can break during a weld [18]. The chapter's TC carries a 0.60 s telemetry delay and a 1.20 s bead lag.

What Moves It

The spindle sets the heat, and the axial axis sets how deep the shoulder rides. Depth was once held by roller assemblies or by actuators fed from external depth sensors, whose reading is only an indirect sign of depth. Feedback on the penetration force has proved a robust and reliable way: on a tapered panel, a force that varied by 890 N kept the depth to a variation of 0.1 mm [19]. On a compliant robot under force control, metal that grows too hot can collapse under the tool, a meltdown that a rigid position-controlled machine does not risk [20]. The chapter's axial axis holds a position or, through its PI loop, a force read by a load cell with a 0.15 s lag.

TECHNOLOGY **Final Element: The Spindle Servo Drive**

The tool turns on a spindle driven by an electric servo motor through a gearbox, and the servo drive sits on a fieldbus that carries its speed setpoint in and its measured speed out [20]. Conventional practice holds the spindle at constant speed by feedback control of the motor. Neglecting the losses through the pin and the spindle, the heat put into the work is the product of torque and rotational speed, so one proposal regulates the motor current to hold the torque constant and lets the speed vary [21]. With a small tool the torque changes little with temperature, which makes it a poor variable to control [20]. The chapter's spindle follows its command with a 0.25 s lag, and its torque is the heat divided by the spindle speed.

Where the Controller Runs

In one robotic cell the force loop runs inside the robot controller, and the temperature loop is a software PI controller on a separate measurement computer that sends a spindle-speed setpoint over a fieldbus. Its loop runs at 5 Hz, limited by the iteration speed of the robot program, and its first-order model came from a step in spindle speed during steady welding [20]. The chapter's PI loops execute every 0.2 s, the same period, and its predictive controllers every 1.0 s, a layer that would sit on such a supervisory computer. Figure 3.5 draws the loops on that hardware.

IN PRACTICE **Qualifying the Recipe, Not the Loop**

Welds in crewed spaceflight hardware are qualified by statistics. Once the optimal parameters are set for a material, a joint design and a thickness, a large population of test welds is made and tested, and the strength data must show with 95 percent confidence that 90 percent of the weld strengths fall within the structural design criteria. On the tank the friction stir welds earned a greater margin of safety than the arc welds before them, because their test data scattered less [3]. The Discussion of this chapter notes that a sacrificial run-out tab is cheaper than requalifying a recipe.

What Is Modernizing It

The tool-workpiece thermocouple needs no sensor inside the tool. A PI loop on its reading, moving spindle speed, has held the welding temperature within 10 °C on workpieces of changing shape and has started the traverse when the plunge reached its temperature, and its authors left more advanced models and control schemes to future work [20]. For the welds of a launch vehicle, a physics-based flow model predicts input power and wormhole defects, and a reinforcement-learning agent adjusts spindle speed, travel speed and load during the weld to hold input power in a target window [22]. The

self-reacting tool, with a second shoulder under the plate, removes the expensive tooling that reacts the forging load from below [3], which is the backing anvil against which this chapter measures its root clearance.

3.8 Limitations

The application is not a friction stir welding process model. With four thermal nodes, a calibrated load-sharing split and an empirical quality window, no claim about a real weld's peak temperature, its nugget geometry, its defect map or its mechanical properties can rest on it, and the effective slip at the tool interface, the largest uncertainty in published thermal models, is not represented at all [5]. What it is evidence for is regime-dependent plant behavior, the servo against regulator trade with numbers attached, step-response predictive control from identification through to measured execution cost, and what a model of the wrong operating point costs.

Three scope limits follow from the control design rather than from the physics. The predictive controllers are single input, manipulating spindle speed only, because traverse speed is the production rate and an optimizer free to slow the machine down needs an economic objective the application does not carry, so multivariable move allocation cannot be demonstrated. Both identified models are first order plus dead time while the real response has at least three thermal modes in it, and the plunge residual of Table 3.3 reports that mismatch, which means the plunge-model controller is handicapped by structure as well as by operating point. The axial axis offers position or force but not the interlocked combination that production machines run [10], which is why the challenge can be failed on the root. Section 3.3.9 states where each physical assumption breaks.

Three elements of the scenarios are stylized on purpose. A plate whose thickness changes by 0.10 mm in one instant is not something rolled stock does, but only a step gives a response that can be read straight off a trend. A tool that reaches half its wear life inside one weld wears far faster than a real tool in aluminum, which puts a drift a shop discovers over weeks inside a single run. The 330 °C soak setpoint turns a saturated actuator into a control problem, which is what lets the two identified models be compared where they disagree most.

> ⚠ **OPEN CHALLENGE** **No Agreed Target for a Sound Weld**
>
> Weld temperature is agreed to bear on quality, yet there is no agreement on the best weld temperature. Reported targets run from 60 to 90 percent of the solidus temperature in degrees Celsius, and a reading at one point of the tool is questioned because the temperature varies across it [18]. In the welds of a launch vehicle, abnormally low strengths have appeared sporadically, and an investigation tied quality to input power instead, which suggests holding a power window [22]. The chapter's window is a teaching device. Progress would be a quality window measured on production welds and stated in a variable a machine can hold.

3.9 Discussion

A friction stir weld runs open loop on a recipe because the middle of a long weld is genuinely forgiving, and Table 3.4 shows where that forgiveness ends. Started at quasi-steady conditions in scenario 4, each of the four automatic architectures holds the tool within 0.1 °C of setpoint, and over a complete weld every millimeter of defect any of the five leaves lies in the first 52 mm of the joint, so the whole spread between the best and the worst of them, 89.8 against 98.3 % sound weld, is made where the traverse begins. Two of the disturbances are unmeasured on a real machine and both are ordinary. Tool wear moves the steady tool temperature by 34.7 °C between a new tool and half life at the same spindle speed, which is why recipes are written per tool, and a plate lifting off the backing bar is worth 14.3 °C. Closing the temperature loop is how the two ends of the weld are recovered and how both drifts become visible in the manipulated variable, which is the case the closed-loop literature has been making since the rotating-tool measurement became practical [12, 13].

OPEN CHALLENGE When Is a Model Accurate Enough?

A survey of industrial predictive control in 2003 found model uncertainty poorly handled. Products could be detuned for robustness, but the trade against performance was not clear, and until that link is made, the survey concluded, it is not possible to tell when a model is accurate enough. It expected test signals to be designed to explore the regions where a model is inadequate for control [8]. The two step tests of Table 3.3 meet both questions inside one weld. Progress would be a test that finds where a model must be accurate.

Methods well beyond the five architectures implemented here fit the model of this chapter. The four-node balance is small enough for nonlinear predictive control with the shear-layer temperature as a state, which would remove the regime problem rather than trade between its two halves, and that is the direction recent friction stir welding work has taken [14]. A gain schedule on the engagement fraction a_f of (3.2) is a smaller step and suits a plant whose gain moves with a measured geometric quantity [6]. Moving horizon estimation on the same equations could recover the anvil contact and the tool wear from the tool temperature and the axial force together. The run-out is the easiest of all, because the machine knows the distance to the plate edge and the rise from (3.12) is a known function of it, so a feedforward spindle taper cancels most of 31.6 °C of predictable disturbance. Almost nobody implements it, because a sacrificial run-out tab is cheaper than requalifying a recipe.

3.10 Conclusion

This chapter studied a friction stir weld as a laboratory for plants whose linearization changes inside a single batch. The simulator integrates nine states covering the shear layer, the tool, the heated halo, the anvil interface, the two drives, the axial position, the distance welded and cumulative tool wear, with Schmidt heat generation over three contact surfaces, an interface shear stress taken as a smooth minimum of a sliding and a sticking branch, a flow stress falling by a factor of eighteen from room temperature to 500 °C, an axial force limited to 20 kN, and a plate conductance that changes with traverse speed and collapses at the plate edge. The design point is 471.4 °C at the tool with 9.50 kN on the axis and 1755 W of frictional and plastic work, which is 877 J into every millimeter of joint. A step test in quasi-steady welding identifies 0.1425 °C/rpm with a 12.24 s time constant, and the same experiment through the plunge returns 0.2401 °C/rpm with 1.17 s, ratios of 1.68 and 10.5 within one weld. Entering Auto from the manual station once the run is under way, ITAE setpoint tuning holds the plunge soak to 18 °C s against 213 for disturbance tuning and then overshoots a setpoint step by 1.5 °C against 10.9, while a loop started already in Auto saturates and holds neither soak. The predictive controller built on the traverse step test finishes a weld 98.3 % sound, and the identical controller built on the plunge step test enters the window fastest of the five and then finishes 91.6 % sound with 18 mm of flash, both at 0.27 µs per execution. A fuller model would add a second manipulated variable with an economic objective, a gain schedule on the engagement fraction, an interlocked axial axis, and a distributed thermal model against which the four-node network could be checked. On this weld the price of choosing the wrong regime for an identification experiment is paid in millimeters of scrap.

KEY POINTS

- Heat in a friction stir weld falls as the metal softens, so the weld does not run away.
- The plunge has 1.7 times the gain and a tenth of the time constant of quasi-steady welding.
- The traverse-model DMC finishes a weld 98.3 % sound, the plunge-model DMC 91.6 % with flash.
- Servo and regulator tunings change places between the soak and the traverse.
- Force control holds the heat against a thin plate and spends the root clearance.

Where else this applies. A tank in Volume 1, Chapter 2 also hands its operating point to the model, because its gain and time constant both follow the level.

References

[1] R. S. Mishra and Z. Y. Ma, "Friction stir welding and processing," *Materials Science and Engineering: R: Reports*, vol. 50, no. 1–2, pp. 1–78, 2005.

[2] P. L. Threadgill, A. J. Leonard, H. R. Shercliff, and P. J. Withers, "Friction stir welding of aluminium alloys," *International Materials Reviews*, vol. 54, no. 2, pp. 49–93, 2009.

[3] J. Ding, B. Carter, K. Lawless, A. Nunes, C. Russell, M. Suites, and J. Schneider, "A decade of friction stir welding R&D at NASA's Marshall Space Flight Center and a glance into the future," NASA Marshall Space Flight Center, NASA Technical Reports Server preprint 20080009619, 2006. [Online]. Available: https://ntrs.nasa.gov/citations/20080009619

[4] T. Sheppard and A. Jackson, "Constitutive equations for use in prediction of flow stress during extrusion of aluminium alloys," *Materials Science and Technology*, vol. 13, no. 3, pp. 203–209, 1997.

[5] H. Schmidt, J. Hattel, and J. Wert, "An analytical model for the heat generation in friction stir welding," *Modelling and Simulation in Materials Science and Engineering*, vol. 12, no. 1, pp. 143–157, 2004.

[6] D. J. Leith and W. E. Leithead, "Survey of gain-scheduling analysis and design," *International Journal of Control*, vol. 73, no. 11, pp. 1001–1025, 2000.

[7] E. F. Camacho and C. Bordons, *Model Predictive Control*, 2nd ed., ser. Advanced Textbooks in Control and Signal Processing. London: Springer, 2007.

[8] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003.

[9] W. M. Thomas and E. D. Nicholas, "Friction stir welding for the transportation industries," *Materials & Design*, vol. 18, no. 4–6, pp. 269–273, 1997.

[10] B. T. Gibson, D. H. Lammlein, T. J. Prater, W. R. Longhurst, C. D. Cox, M. C. Ballun, K. J. Dharmaraj, G. E. Cook, and A. M. Strauss, "Friction stir welding: Process, automation, and control," *Journal of Manufacturing Processes*, vol. 16, no. 1, pp. 56–73, 2014.

[11] W. R. Longhurst, A. M. Strauss, and G. E. Cook, "The identification of the key enablers for force control of robotic friction stir welding," *Journal of Manufacturing Science and Engineering*, vol. 133, no. 3, p. 031008, 2011.

[12] A. Fehrenbacher, N. A. Duffie, N. J. Ferrier, F. E. Pfefferkorn, and M. R. Zinn, "Toward automation of friction stir welding through temperature measurement and closed-loop control," *Journal of Manufacturing Science and Engineering*, vol. 133, no. 5, p. 051008, 2011.

[13] B. S. Taysom, C. D. Sorensen, and J. D. Hedengren, "A comparison of model predictive control and PID temperature control in friction stir welding," *Journal of Manufacturing Processes*, vol. 29, pp. 232–241, 2017.

[14] S. P. Meyer, S. Fuderer, and M. F. Zaeh, "A holistic, model-predictive process control for friction stir welding processes including a 1D FDM multi-layer temperature distribution model," *Metals*, vol. 11, no. 3, p. 502, 2021.

[15] Y. J. Chao, X. Qi, and W. Tang, "Heat transfer in friction stir welding—experimental and numerical studies," *Journal of Manufacturing Science and Engineering*, vol. 125, no. 1, pp. 138–145, 2003.

[16] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.

[17] C. E. Garcia and A. M. Morshedi, "Quadratic programming solution of dynamic matrix control (QDMC)," *Chemical Engineering Communications*, vol. 46, no. 1–3, pp. 73–87, 1986.

[18] A. C. F. Silva, J. De Backer, and G. Bolmsjö, "Temperature measurements during friction stir welding," *The International Journal of Advanced Manufacturing Technology*, vol. 88, no. 9–12, pp. 2899–2908, 2017.

[19] NASA Marshall Space Flight Center, "Controlling force and depth in friction stir welding," NASA Tech Briefs, November 2005, pp. 21–22, MFS-31310, NTRS 20110016397, 2005. [Online]. Available: https://ntrs.nasa.gov/citations/20110016397

[20] J. De Backer, G. Bolmsjö, and A.-K. Christiansson, "Temperature control of robotic friction stir welding using the thermoelectric effect," *The International Journal of Advanced Manufacturing Technology*, vol. 70, no. 1–4, pp. 375–383, 2014.

[21] NASA Marshall Space Flight Center, "Heat control via torque control in friction stir welding," NASA Tech Briefs, September 2004, p. 19, MFS-31834, NTRS 20110020531, 2004. [Online]. Available: https://ntrs.nasa.gov/citations/20110020531

[22] P. W. C. Northrop, A. Maenza, S. Tashakkor, R. Amaro, and J. Anders, "Improving friction stir welding reliability through machine learning control and physics-based modeling," Poster, 2025 Jamboree within the Marshall 65 Spring Showcase, Huntsville, AL, NTRS 20250003564, 2025. [Online]. Available: https://ntrs.nasa.gov/citations/20250003564

3.11 Exercises

3.1 Concept check · 2 min

The Temperature control selector offers five architectures: Manual, two PI loops and two predictive controllers. Which controlled variable and which manipulated variable do all five share?

(a) The axial force, held by moving the commanded plunge depth
(b) The tool temperature, held by moving the traverse speed
(c) The tool temperature, held by moving the plunge depth
(d) The tool temperature, held by moving the spindle speed

3.2 Concept check · 1 min

Which of these quantities does the simulated machine measure and pass to a controller?

(a) The tool temperature, read by a thermocouple in the rotating tool
(b) The deviation of the plate thickness from its nominal value
(c) The contact between the plate and the steel backing anvil
(d) The wear of the tool shoulder since the start of the weld

3.3 Concept check · 2 min

The predictive controller identified during the plunge carries a model gain about 1.9 times the stored traverse gain and a time constant about a sixth as long. Run over a complete weld from a cold plate, what does it do in the traverse?

(a) It oscillates with a growing amplitude until the spindle reaches a limit
(b) It holds a slowly corrected offset that runs the tool hot and lays down flash
(c) It enters the window late and leaves a long cold defect at the start
(d) It matches the traverse model, because the output bias removes model error

3.4 Theory and calculation · 6 min

The flow stress of the shear layer follows Equation (3.3) with $\sigma_0 = 1329\,\text{MPa}$, $T_\sigma = 158.8\,^\circ\text{C}$ and $\sigma_{res} = 8\,\text{MPa}$, and $\phi = 1$ below the solidus.

(a) Find the flow stress at 22, 400 and 500 °C. *Checkpoint:* 1165, 115 and 65 MPa.

(b) Find the ratio of the flow stress at 22 °C to that at 500 °C, the tool temperature of a running weld. *Checkpoint:* About 18.

(c) Raising the spindle from 450 to 900 rpm lowers the interface shear stress under the shoulder from 47.0 to 29.4 MPa. Use Equation (3.5) at a fixed contact to estimate the change in shoulder heat, and compare it with doubling the heat. *Checkpoint:* A rise of about 25 %, not 100 %.

3.5 Theory and calculation · 7 min

The stored traverse model of Table 3.3 is $K_p = 0.1356\,^\circ\text{C/rpm}$, $\tau_p = 11.78\,\text{s}$ and $\theta_p = 2.68\,\text{s}$, and the stored plunge model has $K_p = 0.2552\,^\circ\text{C/rpm}$.

(a) Find θ_p/τ_p and the ITAE setpoint tuning of Equation (3.18). *Checkpoint:* 0.2275, $K_c = 16.77\,\text{rpm}/^\circ\text{C}$, $\tau_I = 11.87\,\text{s}$.

(b) Find the ITAE disturbance tuning of Equation (3.19) and its ratios to the setpoint tuning. *Checkpoint:* $K_c = 26.91\,\text{rpm}/^\circ\text{C}$, $\tau_I = 2.90\,\text{s}$, with 1.60 times the gain and 4.1 times the reset rate.

(c) Find the loop gain $K_c K_p$ of the shipped servo tuning, 16.8 rpm/°C, on the stored traverse and plunge models. *Checkpoint:* 2.28 and 4.29.

3.6 Theory and calculation · 6 min

The quality window follows Equation (3.15) with $T_{lo,0} = 435\,^{\circ}\mathrm{C}$, $T_{hi,0} = 505\,^{\circ}\mathrm{C}$, slopes $s_{lo} = 0.20$ and $s_{hi} = 0.05\,^{\circ}\mathrm{C}\,\mathrm{min/mm}$, $v_{ref} = 120\,\mathrm{mm/min}$ and $T_{sol} = 582\,^{\circ}\mathrm{C}$.

(a) Find both edges and the width of the window at the challenge's 190 mm/min, and the margin to the cold edge of a tool at 455.6 °C. *Checkpoint:* 449 and 508.5 °C, 59.5 °C wide, and 6.6 °C of margin.

(b) At a fixed 900 rpm the steady tool temperature is 471.4 °C at 120 mm/min and 446.9 °C at 240 mm/min. Find the margin to the cold edge at each speed and estimate by linear interpolation the traverse speed at which the open-loop margin vanishes. *Checkpoint:* 36.4 °C and −12.1 °C, zero at about 210 mm/min.

(c) The design weld generates 1755 W. Find the heat input per millimeter of joint at 120 and at 190 mm/min if the heat stayed the same. *Checkpoint:* About 877 and 554 J/mm.

3.7 App challenge · 8 min

Load scenario 2, *The plunge: 20 kN and a stalled axis*, default seed, and run it to 26 s at the shipped 900 rpm. Export the CSV.

(a) Read the time at which the axis reaches depth (the first row whose `phase` column reads `DWELL`), the Peak force tile, and the tool temperature `Tt_C` in the last `DWELL` row, at the end of the dwell. *Checkpoint:* About 12.2 s, 20.0 kN, 372 °C.

(b) Reset, set Spindle speed to 1300 rpm before Run, and repeat. *Checkpoint:* About 9.0 s, 20.0 kN, 405 °C.

(c) Reload the scenario, set the halo weight w_b (Weight of the halo temperature in the load-bearing metal) to 0 in Show Model, and repeat at 900 rpm. Say what changed and why. *Checkpoint:* About 3.8 s, 18.1 kN, 319 °C.

Give the times to 0.2 s and the temperatures to 1 °C.

3.8 App challenge · 8 min

Load scenario 5, *Servo tuning through the plunge*, default seed. It starts from a cold plate with Temperature control at PID · servo already in Auto. Run to 170 s at speed 5, past WELD COMPLETE, and export the CSV.

(a) Read In window, Defect length and the tool temperature `Tt_C` in the last row whose `phase` is `DWELL`. *Checkpoint:* 97.4 %, 6 mm, 400 °C.

(b) Reset, set Temperature control to PID · regulator before Run, and repeat. *Checkpoint:* 96.9 %, 8 mm, 393 °C.

(c) The soak error at the first execution is $330 - 22 = 308\,^{\circ}\mathrm{C}$. Find the proportional kick of each tuning and explain why the servo tuning wins this weld, when Table 3.4 shows the regulator tuning reaching the window first after a bumpless entry.

Give the percentages to 0.1 % and the temperatures to 1 °C.

3.9 App challenge · 9 min

Load scenario 9, *The plunge MPC meets the traverse*, default seed. It runs the plunge-identified predictive controller in Auto from a cold plate. Run each part to 170 s at speed 5 and read In window, Defect length and Peak tool T.

(a) Run it as loaded. *Checkpoint:* 91.6 %, 21 mm, 507 °C.

(b) Reload, switch to Advanced, and set MPC move suppression to 1.0 and MPC reference time constant to 12 s before Run. *Checkpoint:* 82.5 %, 43 mm, 513 °C.

(c) Reload and set MPC move suppression to 0.02, its lowest value, with the reference time constant left at 2 s. Explain why detuning made this controller worse and less suppression made it better. *Checkpoint:* 94.7 %, 13 mm, 506 °C.

Give the percentages to 0.2 %, the lengths to 1 mm and the temperatures to 1 °C.

3.10 Application area · 3 min

A fabricator welds 6.35 mm plate with a 6.00 mm pin at 0.15 mm of shoulder plunge, which leaves 0.20 mm of root. The rolled stock carries thin patches of a tenth of a millimeter.

(a) Say what a position-controlled axial axis gives up on a 0.10 mm thin patch, and what a force-controlled axis gives up.

(b) Say how production machines combine the two, and which limit protects the anvil.

3.11 Application area · 4 min

A friction stir weld is still qualified as a recipe run open loop: a spindle speed, a traverse speed and a plunge depth.

(a) Say why that works for most of a long weld, and where the scrap of an open-loop recipe concentrates.

(b) Tool wear and a plate lifting off the backing bar are not measured on a real machine. Say how much each moves the steady tool temperature in the model, and what a closed temperature loop shows an operator when either happens.

3.12 Review the agent · 5 min

An AI assistant reviews the graded challenge and writes: “Reveal the suggested configuration (traverse-identified MPC at 130 mm/min) and then set the Z axis to Force. Force control removes the 0.09 mm thin patch as a heat disturbance, so the weld can only get sounder, and the challenge passes with a wider margin.” Find the flaw, and say how you would check the claim in the app.