INTERACTIVE CONTROL ENGINEERING

Batch, Run-to-Run, and Hybrid Control

Interactive Control Engineering: Principles and Applications

Edited by John D. Hedengren

Control technology

1 Process Dynamics, Modeling, and Identification
2 Practical PID Control
3 Feedforward, Cascade, and Ratio Control
4 Override, Split-Range, and Constraint Control
5 Delays, Slow Measurements, and State Estimation
6 Multivariable and Plantwide Control
7 Nonlinear Dynamics, Oscillations, and Stabilization
8 Batch, Run-to-Run, and Hybrid Control
9 Model Predictive Control and Economic Optimization

Industry applications

10 Control in Upstream and Midstream Oil and Gas
11 Control in Petroleum Refining
12 Control in Petrochemicals and Polymers
13 Control in Chemical Plants
14 Control in Semiconductor Manufacturing
15 Control in Mining and Mineral Processing
16 Control in Iron, Steel, and Metals Production
17 Control of Energy Conversion and Storage
18 Control of Nuclear and Fusion Systems
19 Control of Biological Systems and Human Performance
20 Motion Control, Robotics, and Autonomous Systems

INTERACTIVE CONTROL ENGINEERING
Principles and Applications

Volume 8

Batch, Run-to-Run, and Hybrid Control

Recipes, endpoints and learning from the last run

John D. Hedengren
Editor

Advanced Process Solutions, LLC

Interactive Control Engineering: Principles and Applications
Volume 8: Batch, Run-to-Run, and Hybrid Control
Edited by John D. Hedengren

First edition, 2026

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

Produced by Advanced Process Solutions, LLC.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the publisher, except for brief quotations in a review or scholarly work.

The simulators described in this book are teaching models. Their equations, parameters and results are simplified for instruction and are not validated for the design, operation or safety assessment of any real plant, vehicle or medical device. The editor and the publisher accept no liability for use of this material outside that purpose.

Names of products and companies are used for identification only and may be trademarks of their owners.

Every chapter has a companion simulator that runs in a web browser at https://apmonitor.com/apps/. This volume online, with chapter files and citation records: https://apmonitor.com/apps/books/v08-batch-hybrid/

How to cite this volume. John D. Hedengren, Ed., *Batch, Run-to-Run, and Hybrid Control*, vol. 8 of *Interactive Control Engineering: Principles and Applications*. Advanced Process Solutions, LLC, 2026.

Typeset in Times and Helvetica-family type with LaTeX and the memoir class.

Preface

This book is one of 20 volumes of *Interactive Control Engineering: Principles and Applications*. Each chapter is about one plant, machine or system and one control lesson, and each has a simulator that runs in a web browser with nothing to install. The simulators are collected at https://apmonitor.com/apps/. Nine volumes are organized by control technology and eleven by industry or application area; the last pages of this volume list them all.

How to read this book with the apps

Chapter 1 teaches the theory this volume relies on and explains, once, the simulator framework that the apps share. Each chapter after it opens with the control question it answers, its learning objectives, what to read first, and the address of its app, also given as a code to scan. The chapter then develops the model, the controllers and the measured results. Every number reported was measured on the app, so a reader can open the app, load the named scenario and reproduce it.

Every chapter ends with exercises. Each is labeled with its kind and the time it should take, and none takes more than ten minutes. Concept checks and calculations can be done on paper. App challenges name a scenario, the settings and the run length. One exercise in each chapter, labeled *Review the agent*, presents a tuning, a model or a claim proposed by an AI assistant and asks what is wrong with it. Short answers to the odd-numbered exercises are at the back of the book, and worked solutions are available to instructors.

How AI agents were used

AI agents did much of the development work behind this series. Under the editor's direction they wrote and tested the simulators, drafted the papers from which the chapters are drawn, measured the reported results on the running apps, converted the papers to book form, and drafted the introductory and closing chapters and the exercises. Other agents then checked that work: they re-measured reported numbers, audited what each app displays against the state of its model, reviewed the papers as referees, and solved the exercises without seeing the answers. The editor set the scope, chose the plants and the lessons, reviewed the results and is responsible for the content.

That way of working is also a subject of the book. The last chapter of each volume describes what the agents got wrong while this volume was made and how the errors were caught, and it sets out what the editor believes this means for engineers: more of the routine work can be delegated, and more of the engineer's value lies in checking, supervising, choosing the problem and knowing the process.

About this volume

This volume is about processes that never settle to a steady state and controllers that switch. Chapters 2 to 8 control what happens within one batch: a dosing rate that is also the safety layer, a reflux policy, a cooling trajectory inside a metastable zone, drying and curing cycles limited by the slowest and hottest part, and products whose quality is not measured during the run. Chapters 9 to 15 correct from one run or cycle to the next and handle modes, switching and actuators that take only whole values.

Volume 5 covers the delayed and inferred measurements that many batch plants depend on. Volume 14 applies run-to-run control throughout semiconductor manufacturing.

John D. Hedengren
2026

Contents

Preface v

Notation xi

1 Introduction: Batches, Repeated Runs, and Switched Plants 1
- 1.1 The Question This Volume Answers . . . 1
- 1.2 A Short History . . . 2
- 1.3 Notation and Conventions . . . 3
- 1.4 Theory on One Batch Kettle . . . 3
- 1.5 Instruments and Final Elements in This Volume . . . 12
- 1.6 The Simulator Framework . . . 14
- 1.7 Working With AI Agents While Studying . . . 17
- 1.8 Chapter Roadmap . . . 18
- 1.9 Exercises . . . 21

2 Semi-Batch Exothermic Reactor Simulator With Dosing as the Safety Layer 23
- 2.1 Introduction . . . 24
- 2.2 Accumulation, Calorimetry and the Cooling-Failure Scenario . . . 25
- 2.3 The Reactor as Modeled . . . 26
- 2.4 Where the MTSR Is Lowest . . . 31
- 2.5 Implemented Control System . . . 32
- 2.6 Measured Cases . . . 38
- 2.7 Technology in Practice . . . 47
- 2.8 Limitations . . . 49
- 2.9 Conclusion . . . 50
- 2.10 Exercises . . . 52

3 Reflux Policy on a Batch Column With No Steady State to Tune 55
- 3.1 Introduction . . . 56
- 3.2 Background . . . 57
- 3.3 Process Model . . . 58
- 3.4 Dynamic Analysis . . . 63
- 3.5 Controller Design . . . 64
- 3.6 Results . . . 67
- 3.7 Discussion . . . 74
- 3.8 Technology in Practice . . . 75
- 3.9 Limitations . . . 77
- 3.10 Conclusion . . . 78
- 3.11 Exercises . . . 80

4 Batch Crystallizer Simulator for Trajectory Control Inside a Metastable Zone 83
- 4.1 Introduction . . . 84
- 4.2 Background . . . 85
- 4.3 Crystallizer Model . . . 86
- 4.4 Cooling Profiles and the Admissible Cooling Rate . . . 92
- 4.5 Controller Design . . . 93
- 4.6 Results . . . 96
- 4.7 Discussion . . . 102

4.8 Technology in Practice . . . 103
4.9 Limitations . . . 105
4.10 Conclusion . . . 106
4.11 Exercises . . . 108

5 Hottest and Slowest Vials in an Endpoint-Aware Freeze-Drying Simulator 111
5.1 Introduction . . . 112
5.2 From Lyophile Serum to Model-Based Cycles . . . 113
5.3 The Batch as Modeled . . . 114
5.4 Controllers, Estimator and Supervisor . . . 119
5.5 Verification and Calibration . . . 123
5.6 Ten Cases, Measured . . . 124
5.7 Technology in Practice . . . 133
5.8 Limitations . . . 135
5.9 Conclusion . . . 135
5.10 Exercises . . . 137

6 Slowest Part, Hottest Core: Lead/Lag Control of a Composite Autoclave Cure 141
6.1 Introduction . . . 142
6.2 Autoclaves, Cure Models and Cure Control . . . 143
6.3 The Load as Modeled . . . 146
6.4 What Sets the Lags and the Limit . . . 150
6.5 The Control Layers as Implemented . . . 151
6.6 Code Verification and Agreement With the Source . . . 155
6.7 Nine Cases and a Challenge, Measured . . . 155
6.8 Technology in Practice . . . 163
6.9 Limitations . . . 164
6.10 Conclusion . . . 165
6.11 Exercises . . . 167

7 Injection Molding Simulator for Bumpless Switchover and Shot-to-Shot Mass Correction 170
7.1 Introduction . . . 171
7.2 Control of the Molding Cycle . . . 172
7.3 Model of One Shot . . . 174
7.4 Sequence and Control as Implemented . . . 178
7.5 Results . . . 180
7.6 Technology in Practice . . . 186
7.7 Limitations . . . 187
7.8 Conclusion . . . 187
7.9 Exercises . . . 189

8 TRISO Coating Reactor Simulator for Inferential Control of an Unmeasured Product 193
8.1 Introduction . . . 194
8.2 Coated Particle Fuel and Its Coaters . . . 195
8.3 Coater Model . . . 196
8.4 The Thickness Estimator . . . 203
8.5 Control System . . . 205
8.6 Results . . . 207
8.7 Discussion . . . 213
8.8 Technology in Practice . . . 214
8.9 Limitations . . . 215

8.10 Conclusion . . . 216
8.11 Exercises . . . 218

9 Aseptic Vial Filling Simulator With Drifting Heads and Late Checkweigher Results 221
9.1 Introduction . . . 222
9.2 Giveaway, Sampling and Feedback Adjustment . . . 223
9.3 The Filling Line as Modeled . . . 224
9.4 The Dose Records . . . 228
9.5 The Corrections as Implemented . . . 229
9.6 Measured Results . . . 230
9.7 Technology in Practice . . . 236
9.8 Limitations . . . 237
9.9 Conclusion . . . 238
9.10 Exercises . . . 239

10 Learning From the Last Layer: A Powder Bed Fusion Simulator for Iterative Learning Control 243
10.1 Introduction . . . 244
10.2 Melt Pool Monitoring and Control . . . 245
10.3 Process Model . . . 246
10.4 Implemented Control . . . 248
10.5 Results . . . 249
10.6 Technology in Practice . . . 252
10.7 Limitations . . . 254
10.8 Conclusion . . . 255
10.9 Exercises . . . 256

11 Switching a Three-Column Antibody Capture Train on Resin of Unknown Capacity 260
11.1 Introduction . . . 261
11.2 From an Immunosorbent to a Rotating Train . . . 262
11.3 The Capture Train as Modeled . . . 263
11.4 Loops, Switching Rules and the Supervisor . . . 267
11.5 Code Verification and Agreement With the Literature . . . 269
11.6 Ten Cases, Measured . . . 270
11.7 Policies Under Equal Resin and Feed . . . 273
11.8 Technology in Practice . . . 275
11.9 Limitations . . . 277
11.10 Conclusion . . . 278
11.11 Exercises . . . 279

12 Thermostats That Fall Into Step on a Simulated Supermarket Refrigeration Pack 283
12.1 Introduction . . . 284
12.2 Expansion Valves, Compressor Racks and Cases in Step . . . 285
12.3 The Pack as Modeled . . . 286
12.4 What the Line and the Shared Pressure Imply . . . 291
12.5 Thermostats, Valves, Rack and Fans as Programmed . . . 293
12.6 Measured Cases, From a Hand on the Valve to a Hot Day . . . 294
12.7 Technology in Practice . . . 301
12.8 Limitations . . . 303
12.9 Conclusion . . . 304
12.10 Exercises . . . 306

13 Milk Pasteurizer Simulator for Product History, Regeneration and Flow Diversion 309
13.1 Introduction . . . 310
13.2 Pasteurization, Holding Time and Diversion . . . 311
13.3 Parcel Model of the Pasteurizer . . . 312
13.4 Regeneration as a Recycle . . . 316
13.5 Implemented Controller . . . 317
13.6 Scenarios and Measured Results . . . 318
13.7 Technology in Practice . . . 323
13.8 Limitations . . . 325
13.9 Conclusion . . . 325
13.10 Exercises . . . 327

14 Integer Servers, Boot Delays and Retry Storms in a Cloud Autoscaling Simulator 330
14.1 Introduction . . . 331
14.2 From Admission Control to the Desired-Replica Rule . . . 332
14.3 The Service as Modeled . . . 333
14.4 Autoscalers as Implemented . . . 338
14.5 Results . . . 340
14.6 Verification, Agreement and Evaluation . . . 347
14.7 Technology in Practice . . . 347
14.8 Limitations . . . 349
14.9 Conclusion . . . 350
14.10 Exercises . . . 351

15 A Juggling Robot Arm Whose Paddle Must Be Slowing at Impact 354
15.1 Introduction . . . 355
15.2 Bouncing Balls, Return Maps and Juggling Machines . . . 356
15.3 Ball, Paddle and Arm as Simulated . . . 357
15.4 One Bounce at a Time . . . 361
15.5 The Laws as Coded . . . 364
15.6 Ten Cases, Measured . . . 366
15.7 Technology in Practice . . . 370
15.8 Limitations . . . 372
15.9 Conclusion . . . 373
15.10 Exercises . . . 374

16 Summary and Outlook 377
16.1 What the Chapters Showed . . . 377
16.2 Choosing a Method . . . 380
16.3 Where Practice Is Heading . . . 381
16.4 Engineering With AI Agents . . . 383
16.5 Gaps in Batch, Run-to-Run and Hybrid Control . . . 386
16.6 Exercises . . . 390

Answers to Selected Exercises 391

Glossary 396

Index of Control Topics 399

Index of Applications 401

About the Editor 404

Volumes in This Series **405**

Notation

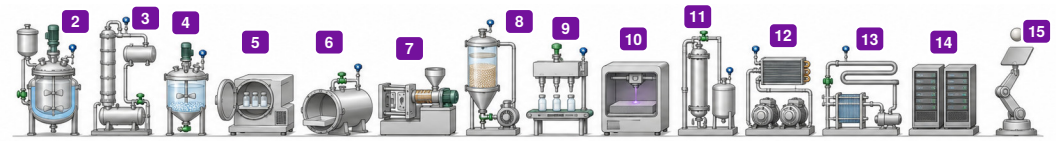
The chapters share the symbols below. A symbol used by one chapter only is defined where it first appears, and each chapter states its own units.

Symbol	Meaning
SP	setpoint, the value a controller is asked to hold
PV	process variable, the measurement a controller acts on
CV	controlled variable
MV	manipulated variable; also OP, the controller output
DV	disturbance variable
e	control error, SP $-$ PV
K_c	controller gain
τ_I, τ_D	integral (reset) time and derivative time
K_p	process gain, the steady-state change in PV per unit change in MV
τ_p	process time constant
θ_p	process dead time
τ_c	closed-loop time constant chosen in a tuning rule
Δt	controller sample time
s	Laplace variable
x, u, y, d	state, input, output and disturbance of a state-space model
A, B, C	matrices of the linear model $\dot{x} = Ax + Bu$, $y = Cx$
Q, R	weights on outputs (or states) and on input moves in an optimal controller
$\hat{x}$	an estimate of x

Abbreviations used throughout: PID, proportional-integral-derivative; PI, proportional-integral; FOPDT, first-order plus dead time; MPC, model predictive control; MHE, moving horizon estimation; IAE, integral of the absolute error; RGA, relative gain array. Times are in seconds unless a chapter says otherwise. Controller output and valve position are in percent of span. A scenario or case number refers to the numbered list in the chapter's app.

CHAPTER 1

Introduction: Batches, Repeated Runs, and Switched Plants



The chapters 2 semi-batch reactor 3 batch column 4 crystallizer 5 freeze-dryer 6 cure autoclave 7 injection molding 8 particle coater 9 vial filler 10 powder bed fusion 11 antibody capture 12 refrigeration pack 13 milk pasteurizer 14 web service 15 juggling paddle

Notice that the first five units each run one batch to an end, the next five repeat a shot, a coating, a vial, a layer or a column load, and in the last four the actuators or the dynamics switch. Illustration, not to scale.

IN PLAIN TERMS **Plants That Start, Repeat or Switch**

The main idea. Many plants never settle at one steady point. A batch is started, led through a recipe and stopped. A machine repeats the same short run hundreds of times, and a thermostat or a pool of servers switches in whole steps.

The challenge. The product of a batch is judged once, after the end, so a mistake found then cannot be undone in that batch. Each run can learn from the last. But a correction made after every result also copies the noise of one run into the next.

Why it matters. These plants make medicines, food, molded parts and fuel, and they run online services that people use every day.

THREE WAYS TO READ THIS CHAPTER

Explore	6 min	This page, Section 1.2 with its timeline and two History boxes, the Did you know? boxes in Sections 1.4.5, 1.4.8 and 1.4.10, Figure 1.11, and Key points in Section 1.8. No equations.
Study	45 min	The whole chapter in order, with each number of the kettle example worked, then Exercises 1 to 10.
Practice	5 min	This page, Section 1.4.10 with Table 1.3, Section 1.5 with Figures 1.7 and 1.9, and Section 1.8 with Figure 1.11.

1.1 The Question This Volume Answers

The loops of the earlier volumes hold a plant near one operating point for as long as it runs. The plants of this volume have no such point. A batch starts cold, follows a program and ends, and its product is judged once, after the end. A molding machine, a vial filler or a printer repeats the same run hundreds of times, and each run can learn from the last. A thermostat, a contactor or a bouncing ball switches between modes, and the switching is the control. This volume answers one question: how is a plant controlled when its operation is a sequence, a repetition or a switch, rather than a steady state?

Fourteen chapters answer it on fourteen plants, each with an app that the reader runs in a browser. Chapters 2 to 6 run one batch along a trajectory to an endpoint under a limit that holds at every instant. Chapters 7 to 11 repeat a shot, a coating run, a vial, a layer or a column cycle, end each on an event or an estimate, and correct the next one from the last. Chapters 12 to 15 are plants whose actuators or dynamics switch: thermostats on one suction header, a diversion valve under a supervisor, a fleet of servers counted in integers and a paddle that strikes a ball.

A reader who has worked through the volume can do the following on a plant that is new to the reader.

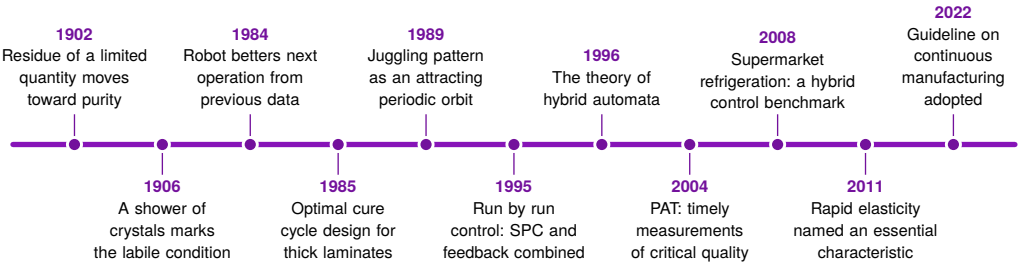


Figure 1.1 Milestones of batch, run-to-run and hybrid control in the systems of this volume. The dates are from [4] (1902), [5] (1906), [7] (1984), [8] (1985), [9] (1989), [10] (1995), [11] (1996), [12] (2004), [13] (2008), [14] (2011) and [15] (2022). Notice that the batch was described long before it was controlled, and that the last cards move the end of a phase from the clock to a measurement. Events are evenly spaced, not to a time scale.

- Write a recipe as phases with transition conditions, and choose whether a clock, a measured event or an estimate ends each phase.
- Track a trajectory with feedback and model feedforward, and say what each leaves behind.
- Find the path and endpoint constraints of a batch, the shape of the fastest feasible batch and the back-off each constraint needs.
- Design a run-to-run correction and predict its response to a step, a drift, noise, a gain error and a late measurement.
- Design an iterative learning update and test it for monotone convergence.
- Model a switched plant as a hybrid automaton, set a hysteresis band or a dwell time, handle an integer actuator, and find the cyclic steady state a repeated operation settles into.

This chapter develops each of these on one small example and names the chapter that uses it. Overviews of batch control and optimization [1, 2] and of learning in batch processes [3] give fuller treatments. PID forms and tuning are taken as known from Volume 2.

1.2 A Short History

The milestones of Figure 1.1 follow the methods of this volume through very different plants. The first two describe a batch long before anyone controlled one. In 1902 experiments on boiling mixtures showed that, as the distillation of a limited quantity proceeds, the composition of the residue moves always in one direction toward purity [4], the moving composition that Chapter 3 follows. In 1906 a cooling solution watched by its refractive index grew a few crystals and then, at a lower temperature, threw a shower of them throughout the liquid, and cooling slow enough never reached that labile condition [5], the trajectory question of Chapter 4.

> ⧖ **HISTORY** **1930s to 1980s: Batch Control Waits Its Turn**
>
> Batch processing was dominant until the 1930s, yet batch process control drew significant interest only 50 years later. When control theory emerged in the 1940s, industry was shifting from batches to continuous operation, and the dynamics and control of batch processes became a subject of study only in the 1980s. Engineers then tried to carry over what continuous plants had taught them, and the lack of a steady state made that hard. Programmable logic controllers entered the process industries in the early 1970s, following the automotive industry, and replaced relays, counters and timers that had offered much less flexibility [6].

Optimization of the batch followed: a report of 1985 treated the cure cycle of thick composite laminates as a design variable [8], the recipe that Chapter 6 reshapes.

Switching became a subject of its own. A robot that juggled by batting, presented in May 1989, was designed so that its juggling pattern was an attracting periodic orbit [9], the orbit of Chapter 15. A 1996 summary of results on hybrid automata [11] gave Section 1.4.7 its language, and by 2008 a supermarket refrigeration system had become a benchmark problem of hybrid control [13], the plant of Chapter 12. A 2011 definition of cloud computing named rapid elasticity among its essential characteristics [14].

The last milestones move the end of a phase from the clock to a measurement. A 2004 guidance on process analytical technology asked that manufacturing be designed, analyzed and controlled through timely measurements of critical quality attributes [12]. A guideline on continuous manufacturing, adopted in November 2022, noted that such technology lets disturbances be detected in real time and illustrated a therapeutic-protein process with continuous capture chromatography [15], the train of Chapter 11.

> ⧖ HISTORY **1984 to 1995: Learning From the Last Run**
>
> A paper of 1984 proposed a betterment process that improves the next operation of a robot with the data of the previous one. Each input was the last input plus an increment built from the difference between the motion achieved and the motion desired. Its convergence was assured under reasonable conditions and shown in computer simulation [7]. In chip making, a 1993 analysis gave a simple condition relating the EWMA weight and the estimated process gain that keeps such a control stable [16]. A run by run controller of 1995 combined statistical process control with feedback. A rapid mode met sudden shifts, such as those after maintenance, and a gradual mode met drifts, such as deposits building up inside a reactor [10].

1.3 Notation and Conventions

A batch has two time axes. Time within the batch is t, in minutes here. The batch number is j, which the chapters call a run, a shot, a vial, a layer, a cycle or a bounce. The framework's PID counts its executions with k inside a batch (Section 1.6.2). A recipe setting chosen before batch j is u_j, the result measured after it is y_j, its target is y^* and an unmeasured disturbance is d_j. The weight of an exponentially weighted moving average (EWMA) is λ, the gain of an iterative learning update is γ, its lead is ℓ samples and its filter is Q. A hysteresis band runs from T_{lo} to T_{hi}, an integer actuator has the count n, and a cycle-to-cycle map has the eigenvalue μ.

The chapters keep their own symbols. Chapter 9 writes the EWMA weight of each filling head as λ_h, and Chapter 11 uses λ for its capacity filter. Chapter 10 writes the learning gain γ as here but uses λ for the lead, a length in millimeters. Chapter 7 writes the loop gain of its shot-to-shot correction as k, which is $\lambda\xi$ in Section 1.4.5. Chapter 14 uses δ for a deadband in servers. The framework's h is its integration step and a its controller action, so neither appears here with another meaning.

1.4 Theory on One Batch Kettle

1.4.1 The Running Example

The running example is a kettle that heats a charge of liquid food with three identical electric elements, as shown in Figure 1.2. Its temperature obeys Equation (1.1), with the time constant $\tau = 20\,\text{min}$, the room at $T_a = 20\,^\circ\text{C}$ and each element worth $K = 40\,^\circ\text{C}$ of steady rise.

$$\tau \frac{\mathrm{d}T}{\mathrm{d}t} = T_a - T + K u, \qquad 0 \le u \le 3 \tag{1.1}$$

The input u counts elements. Under a thyristor it varies continuously, and under contactors it is an integer, so one, two or three elements hold the kettle at 60, 100 or 140 °C. The product is judged on a thermal dose accumulated over its whole history, Equation (1.2).

$$F = \int_0^{t_f} 10^{(T-80)/10}\,\mathrm{d}t \tag{1.2}$$

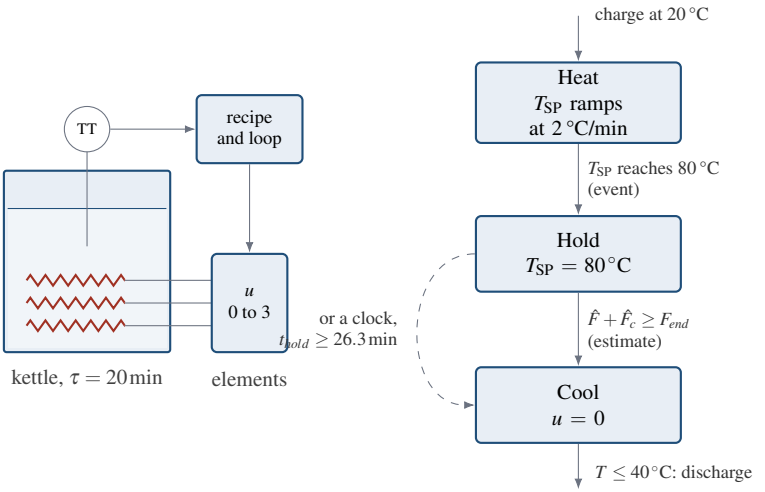


Figure 1.2 The Running Example. Left, a Kettle Heated by Three Elements Under a Thyristor (u Continuous) or Contactors (u an Integer). Right, Its Recipe as Three Modes With the Guard That Ends Each: an Event, an Estimate of the Dose, or a Clock (Dashed).

A minute at 80 °C counts as one minute, and each 10 °C hotter counts ten times as much, the form used for the lethality of a thermal treatment. The specification is $F \geq 30$ min, an endpoint constraint, and the product scorches above 90 °C, a path constraint.

The kettle shows at once what separates a batch from a continuous plant. It has no steady state to hold, since it is charged cold, heated, held, cooled and emptied. Its result is a functional of the whole path, so the same final temperature can come with any dose. And its sensitivity moves along the batch: near 80 °C a 1 °C error changes the dose rate by 25.9 %, while at 50 °C the dose rate is a thousandth of its value at 80 °C and the same error costs almost nothing. Every chapter of the volume has these three properties. The batch column of Chapter 3 grows harder to separate with no disturbance applied, and its process gain decays toward zero over the run. The crystallizer of Chapter 4 is graded only after the batch, on crystals that carry its whole cooling history.

1.4.2 Recipes and Phases

A recipe divides the batch into phases. Each phase has a setpoint program, an active controller, and an end condition, the guard on the way out of the phase in Figure 1.2. A guard can be a clock, a measured event or an estimate, and the choice matters as much as any tuning. The kettle's heat phase ends on an event, the setpoint reaching 80 °C. Its hold ends on an estimate: the dose $\hat{F}$ integrated from the thermocouple, plus the dose the model predicts for free cooling from the present temperature, $\hat{F}_c(T)$, must reach the endpoint setting F_{end}. From 80 °C the cooling still adds 1.57 min of dose, and from 88 °C it adds 8.66 min.

Table 1.1 compares that guard with a clock set to the nominal hold of 26.3 min, on four batches with $F_{end} = 30$ min. With a well-tracked heat-up both deliver 30.09 min. A batch heated by feedback alone lags the ramp, as Section 1.4.3 shows, and the clock then delivers 21.0 min while the estimate extends the hold to 35.4 min and delivers 30.1 min. Weaker elements cost the clock 1.8 min of dose and the estimate nothing. A thermocouple that reads 1 °C high fools both alike, to 23.9 min, because the estimate integrates the same wrong reading. An estimate removes what the model can see and keeps what the instrument cannot.

Every phase change also hands the actuator from one controller or program to the next. The framework's PID tracks the applied output while it is not in charge, so the controller that takes over starts from the value its predecessor left (Section 1.6.2), and Volume 2 treats bumpless transfer in

Table 1.1 Ending the Hold on a Clock or on the Dose Estimate, With $F_{end} = 30\,\text{min}$

Batch	Dose, clock (min)	Dose, estimate (min)	Hold, estimate (min)
PI with feedforward, nominal	30.09	30.09	26.3
PI alone, lagging the ramp	21.03	30.11	35.4
Elements 10 % weak	28.30	30.10	28.1
Thermocouple 1 °C high	23.93	23.93	26.3

full [17]. The chapters use all three kinds of guard. Chapter 8 ends each coating layer on a timer, on an integrated thickness estimate or on that estimate corrected by a load cell, and abandons a phase with a latched fault when it cannot finish. Chapter 7 switches from fill to pack on screw position, cavity pressure or time, with a time limit behind all three, and hands the valve from the speed loop to the pressure loop by tracking. Chapter 5 ends primary drying on a recipe time, on its probes, on a gauge ratio held for 15 min or on its estimator. Chapter 11 ends each loading on a time, an integrated load, a measured breakthrough or a learned capacity. Chapter 13 logs each parcel's history apart from the instruments, the dose idea applied to every parcel of a continuous stream.

1.4.3 Trajectory Tracking

The heat phase asks the kettle to follow a ramp of $\rho = 2\,^\circ\text{C/min}$ from 20 to 80 °C. A PI loop with $\tau_I = \tau = 20\,\text{min}$ and $K_c = 0.1\,\text{element}/^\circ\text{C}$ cancels the plant's lag and leaves a first-order closed loop with the time constant $\tau_c = \tau/(K_c K) = 5\,\text{min}$. On a ramp such a loop lags by Equation (1.3).

$$e_{ss} = \rho\,\tau_c = 2\,^\circ\text{C/min} \times 5\,\text{min} = 10\,^\circ\text{C} \tag{1.3}$$

The lag is no tuning fault. The output must keep rising, and an integral term rises at a constant rate only on a constant error. The simulated kettle reads 70.01 °C when the setpoint reaches 80, collects 0.22 min of dose on the ramp against 2.17 min for a ramp tracked exactly, and reaches 79 °C at 41.5 min, which is $30 + \tau_c \ln 10$. The supersaturation controller of Chapter 4 meets the same fact from the other side: its integral term generates the cooling ramp, and a proportional term alone could not.

The cure is to compute the input the trajectory needs and give it to the loop as its bias, Equation (1.4), so that feedback corrects only what the model misses.

$$u_{ff}(t) = \frac{T_{\text{SP}}(t) - T_a + \tau\,\dot{T}_{\text{SP}}(t)}{\hat{K}} \tag{1.4}$$

On the ramp u_{ff} runs from 1 to 2.5 elements, and in the hold it is 1.5. With the model exact the ramp error has an RMS of 0.012 °C and the batch takes 78.3 min instead of 87.4 min. With elements 10 % weaker or stronger than the model, the RMS is 0.96 or 0.79 °C, and the stronger set overshoots the hold by 0.90 °C. That residue repeats in every batch, which is the opening for Section 1.4.6. Chapter 3 enters its reflux trajectory as the controller bias in the same way, with a proportional trim around it, and finds that a composition loop holding its setpoint closely collects less product than the open-loop trajectory. In Chapter 4 three cooling curves between the same endpoints give mean sizes that differ by a factor of 4.3, and the best-tracked batch gives the second-worst product. Tracking well matters less than tracking the right trajectory.

1.4.4 Path and Endpoint Constraints

A path constraint holds at every instant, here $T(t) \le 90\,^\circ\text{C}$. An endpoint constraint holds once, at the end, here $F(t_f) \ge 30\,\text{min}$. The fastest batch that meets both is made of arcs on which an input or a constraint is at a bound [18]. For the kettle it heats at full power until the temperature meets the path constraint, holds it there with the holding input, and cools with the elements off. The dose then accumulates fastest at every instant.

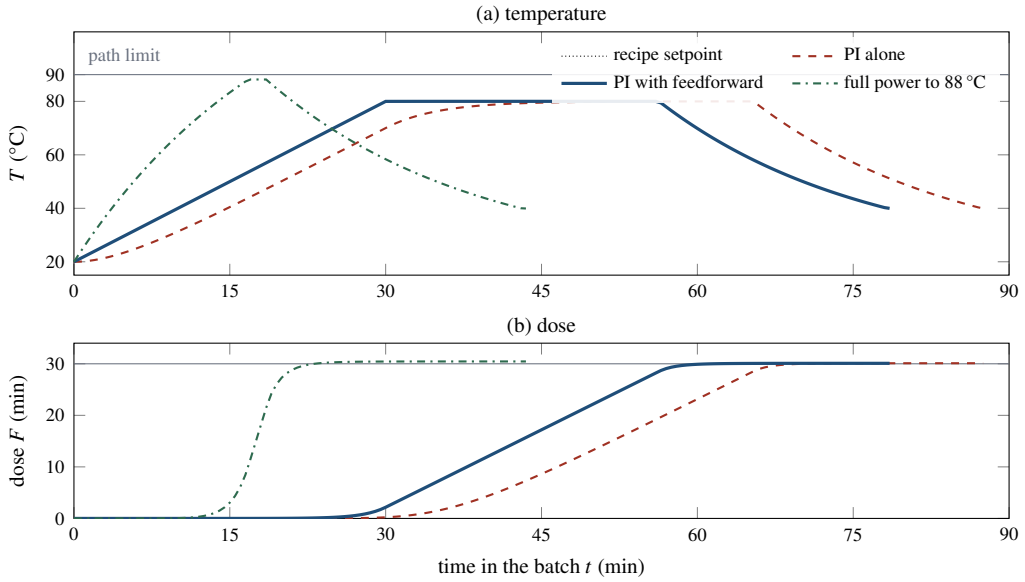


Figure 1.3 One Batch Under Three Policies, Each Ending Its Hold on Equation (1.5). (a) Temperature: PI Alone on the Recipe (Dashed), PI With the Feedforward of Equation (1.4) (Thick Solid), and Full Power to a 2 °C Back-Off From the Path Limit (Dash-Dotted), With the Recipe Setpoint Dotted. (b) The Dose, With the Endpoint Constraint at 30 min.

The constraint cannot be ridden at its value, so the policy switches at 88 °C, a back-off of 2 °C. Full power reaches it in $\tau \ln(120/52) = 16.7$ min, the feedforward of Equation (1.4) then holds it with 1.7 elements and the PI trims, and the hold ends on the guard of Section 1.4.2, written as Equation (1.5).

$$\hat{F}(t) + \hat{F}_c\big(T(t)\big) \geq F_{end} \tag{1.5}$$

Figure 1.3 compares this policy with the recipe. The dose is 11.6 min when the hold starts and 21.5 min after a hold of only 1.7 min, and the cooling supplies the last 9.0 min, so the batch takes 43.1 min against 78.3 min for the recipe, 45 % shorter. Ending the hold on the dose so far, without the predicted remainder, would overshoot the endpoint by the whole cooling dose.

The back-off pays for what the model does not know. With elements 10 % weaker, nominal or stronger, the fast policy peaks at 88.07, 88.19 and 88.57 °C, so 2 °C covers that uncertainty with 1.4 °C to spare. A noisier measurement or a slower loop would need more. The same structure, a fast arc to a limit and a hold on it with a margin, recurs through the batch chapters, each with its own limit. In Chapter 2 the limit is on the temperature the batch would reach if cooling failed, held below the decomposition threshold by the dosing controller rather than by the temperature loop. Chapter 4 states its wall limit as a lower output limit of the jacket loop that moves with the measurement. In Chapter 5 the hottest vials limit the shelf while the slowest decide the endpoint, and in Chapter 6 the slowest thermocouple sets the ramp while the hottest one and a core no thermocouple reaches set the limit, through a selector (Volume 4). Several of these chapters also offer a shrinking-horizon predictive controller that plans to the endpoint (Section 1.6.3, and Volume 9).

1.4.5 Run-to-Run Control

Some errors no estimate inside the batch can see. The thermocouple bias of Table 1.1 is one, and only a measurement of the product, here a laboratory test of the dose after each batch, reveals it. Run-to-run control uses that result to adjust the next batch's recipe. Near its operating point the result

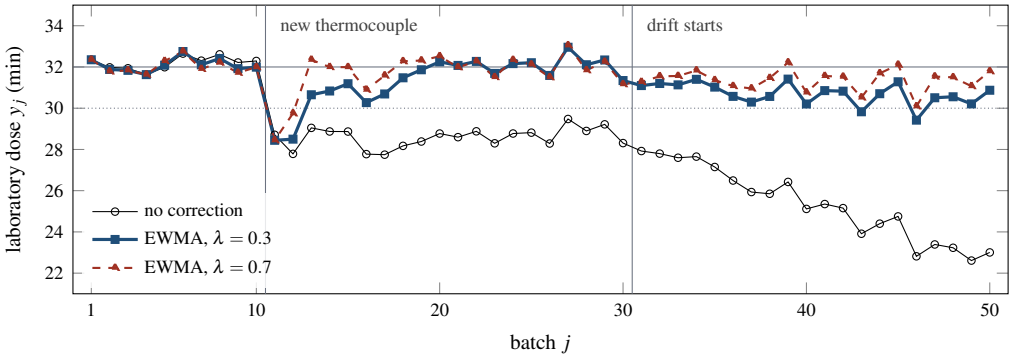


Figure 1.4 Run-to-Run Control of the Kettle's Dose Over 50 Batches With the Same Laboratory Noise: No Correction (Thin Solid, Circles), EWMA With $\lambda = 0.3$ (Thick Solid, Squares) and $\lambda = 0.7$ (Dashed, Triangles). Thin Horizontal Lines Mark the Target of 32 min and the Specification of 30 min (Dotted). A Thermocouple Reading 0.5 °C High Arrives at Batch 11 and Drifts From Batch 31.

of batch j is modeled by Equation (1.6), with the true gain β, a slowly moving disturbance d_j and a noise ε_j new in every batch.

$$y_j = \beta\, u_j + d_j + \varepsilon_j \tag{1.6}$$

The EWMA controller [10, 19] estimates the intercept from each result less the model's prediction from the setting that produced it, and sets the next batch to meet the target with the estimate, Equation (1.7), where b is the model's gain.

$$\hat{d}_j = \lambda\,(y_j - b\,u_j) + (1-\lambda)\,\hat{d}_{j-1}, \qquad u_{j+1} = \frac{y^* - \hat{d}_j}{b} \tag{1.7}$$

With $\xi = \beta/b$, substituting Equation (1.7) into Equation (1.6) gives the four properties collected in Equation (1.8). A step in d shrinks by the factor $1-\lambda\xi$ each batch, so the loop is stable for $0 < \lambda\xi < 2$ [16]. A drift of δ per batch leaves a steady offset δ/λ, and white noise of variance σ^2 leaves an output variance raised by the factor $2/(2-\lambda)$.

$$\mu = 1 - \lambda\xi, \qquad e_{drift} = \frac{\delta}{\lambda}, \qquad \sigma_y^2 = \frac{2\,\sigma^2}{2-\lambda} \tag{1.8}$$

A simulation of the linear model checks each: a drift of 0.1 per batch leaves 0.333 at $\lambda = 0.3$ and 0.143 at 0.7, and 200,000 batches of unit noise give standard deviations of 1.0844 and 1.2406 against 1.0847 and 1.2403. The weight trades the two: $\lambda = 0.3$ brings a step within 10 % in seven batches and widens the noise by 8.5 %, $\lambda = 0.7$ takes two batches and widens it by 24 %. Correcting after every result can only widen pure noise, the lesson of the funnel experiment, and pays only against disturbances that persist [20].

On the kettle the setting is F_{end} and the result is the laboratory dose, with a target of 32 min, two minutes above the specification, and a laboratory noise of 0.5 min. The model gain $b = 1$ is right, since the simulated gain is 1.0013. Figure 1.4 runs 50 batches. From batch 11 a new thermocouple reads 0.5 °C high, which cuts the dose by 3.48 min and the gain to 0.89, and from batch 31 it drifts up by 0.05 °C a batch, about 0.33 min of dose per batch. Without correction 40 of the 50 batches fail the specification. With $\lambda = 0.3$ four fail and with $\lambda = 0.7$ two. Over batches 41 to 50 the two weights fall short of the target by 1.5 and 0.7 min, against δ/λ of 1.2 and 0.5 min with the drift of batch 45, a drift that is itself speeding up.

A result that arrives late changes the picture. Written as an integral action on the error, $u_{j+1} = u_j + (\lambda/b)(y^* - y_{j-D})$ with a delay of D batches, the loop has the characteristic polynomial of Equation (1.9).

$$z^{D+1} - z^D + \lambda\xi = 0 \tag{1.9}$$

It is stable for $\lambda\xi$ below 2, 1, 0.618 and 0.445 at delays of 0, 1, 2 and 3 batches, and with $D = 1$ it is the equation that governs a laboratory-only loop in Volume 5. The intercept form of Equation (1.7) instead compares each late result with the setting that produced it, so with $b = \beta$ it is unaffected by any delay, and a simulation confirms it at $\lambda = 0.7$ with delays of one and two batches. A sufficient condition for stability under any metrology delay with a gain error is known [21].

Three chapters use this section. Chapter 7 corrects pack pressure from each weighed part by the integral form, with $\lambda\xi = 0.55$, and derives the stability bound and the variance factor above. Chapter 9 meets the delay: about six results per head are in flight, the integral law counts each error several times and oscillates at long latencies, and the record-based law, which is Equation (1.7), is unaffected. Chapter 11 learns the resin capacity from earlier switches with an EWMA of weight 0.5. Volume 14 applies run-to-run control to wafers, where it was developed.

> **+ DID YOU KNOW? The Funnel Experiment, Answered in 1990**
>
> Deming's funnel experiment has been used to show that actively controlling a process can increase the variance of its output. A note of 1990 answered that such situations seldom occur in the continuous process industries, where active control to a target had served with excellent results for decades [20].

1.4.6 Iterative Learning Control

Run-to-run control corrects one number per batch. When the error is a whole trajectory that repeats, the correction can be a whole trajectory too. Iterative learning control (ILC) stores an input profile $v_j(t_i)$ on a grid of samples and updates it after each batch from the error of that batch, shifted ahead by a lead of ℓ samples to cover the plant's lag, as shown in Equation (1.10) [7, 22].

$$v_{j+1}(t_i) = Q\left[v_j(t_i) + \frac{\gamma}{\hat{K}}\, e_j(t_{i+\ell})\right] \tag{1.10}$$

For a plant whose sampled transfer function is G, an error component of frequency ω is multiplied by the factor of Equation (1.11) each batch, and the error shrinks monotonically if the factor is below one at every frequency [23].

$$\mu(\omega) = \left|Q(\omega)\left[1 - \frac{\gamma}{\hat{K}}\, G(\mathrm{e}^{\mathrm{j}\omega})\, \mathrm{e}^{\mathrm{j}\omega\ell}\right]\right| < 1 \tag{1.11}$$

On the kettle the heat phase and the first ten minutes of the hold run open loop on a one-minute grid, starting from the feedforward of Equation (1.4) with elements 10 % weaker than the model, so the first batch ends the ramp 6.6 °C low with an RMS error of 4.78 °C. The thermocouple has 0.1 °C of noise. With $\ell = 1$ and no filter the factor is $1 - 0.9\gamma$ at zero frequency and $1 - 0.9\gamma(1-p)/(1+p)$ at the shortest period, with $p = \mathrm{e}^{-1/20}$, so with $\gamma = 1$ a slow error falls tenfold each batch and an error that alternates every minute by only 2 %. The RMS falls to 2.59, 1.43 and 0.84 °C in the next three batches and reaches the noise, about 0.1 °C, by batch 40, as Figure 1.5 shows. Gains up to 2 converge monotonically on this plant, and $\gamma = 0.5$ reaches only 0.55 °C by batch 10.

Learning fails where the lead is wrong. Put the thermocouple in a thermowell that lags 2 min and the one-sample lead falls short: the factor rises to 1.036 at a period of about 20 min. The RMS falls to 1.30 °C at batch 13, then grows back to 2.9 °C by batch 60, because each batch adds a little of the error it cannot remove. A zero-phase low-pass Q, run forward and backward with a pole of 0.3, pulls the factor down to 1.003 and the error settles between 0.71 and 0.80 °C, the price being the short-period detail it no longer learns. A lead of two samples fits the lag better, holds the factor at 1.0005 and keeps falling, to 0.40 °C at batch 100. Learning also copies the noise of each batch into

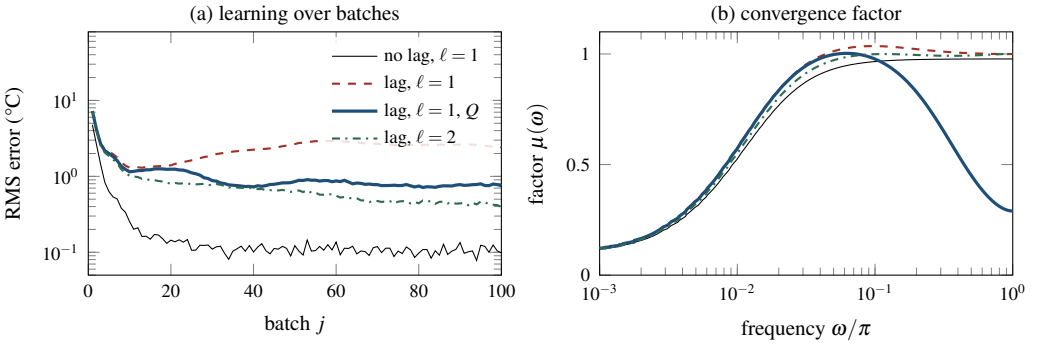


Figure 1.5 Iterative Learning on the Kettle's Heat Phase With $\gamma = 1$ and Elements 10 % Weaker Than the Model. (a) RMS Error per Batch. (b) The Factor of Equation (1.11). Thin Solid: No Thermowell Lag, Lead of One Sample. With a 2 min Thermowell Lag: Lead of One Sample (Dashed), the Same With a Q-Filter (Thick Solid), and a Lead of Two Samples (Dash-Dotted).

the next and has nothing for a batch that does not repeat. Chapter 10 learns a laser power profile along each layer's outline with this update, $\gamma = 0.5$, a lead of 0.6 mm and a Q-filter, adds the feedback effort into the profile, and measures each failure named here: growth without the filter, a corner blurred by too wide a filter, and a layer that breaks the repetition [24].

1.4.7 Hybrid Automata, Switching and Hysteresis

A hybrid automaton [11, 25] is a set of modes, each with its own differential equation, joined by guards that say when the mode changes and resets that say what jumps when it does. The recipe of Figure 1.2 is one. So is a thermostat. Let the kettle hold 80 °C on contactors, with one element always on and the second switched on below $T_{lo} = 79\,^{\circ}\mathrm{C}$ and off above $T_{hi} = 81\,^{\circ}\mathrm{C}$. The kettle then heats toward 100 °C and cools toward 60 °C, and the on and off times follow from Equation (1.1) as Equation (1.12).

$$t_{on} = \tau \ln \frac{100 - T_{lo}}{100 - T_{hi}}, \qquad t_{off} = \tau \ln \frac{T_{hi} - 60}{T_{lo} - 60} \tag{1.12}$$

Both are $20 \ln(21/19) = 2.00$ min, a period of 4.0 min. Scanned every 0.1 min the switch acts one scan late, so the kettle swings from 78.95 to 81.05 °C with a period of 4.2 min, 14.3 starts of the second element an hour. Halving the band halves the swing and doubles the starts, to 27.3 an hour.

The band is what keeps the switching finite. With no band the element switches at every scan, 300 times an hour at a 0.1 min scan and 3000 at 0.01 min, so the rate is set by the scan and not by the plant. In continuous time it would switch infinitely often in a finite time, a Zeno solution. A measurement lag widens the swing beyond the band: a thermowell of 0.5 min gives 78.45 to 81.55 °C at 9.7 starts an hour. A minimum on and off time of 3 min, the protection a compressor needs, gives 78.49 to 81.50 °C at 10 starts an hour. Switching can also destabilize a plant whose every mode is stable, which is one reason for dwell times [26]. Relay oscillation in general is treated in Volume 7.

Safe switching in this volume rests on a few devices. Hysteresis and dwell bound the switching rate. A latch makes a protective switch stay until a condition is cleared and someone resets it. Tracking makes the controller that takes over start from the output in use. Persistence asks a condition to hold for a stated time before it counts. Chapter 12 runs a hysteresis thermostat on each case, stages compressors only after the suction has been outside a 2 K dead band for 15 s, enforces minimum run and rest times, and shows four thermostats that never exchange a signal falling into step through the suction pressure they share. Chapter 13 diverts at once on any of four conditions and returns to forward flow only after two tube volumes have flushed. Chapter 2 cuts the feed 2 K above its setpoint,

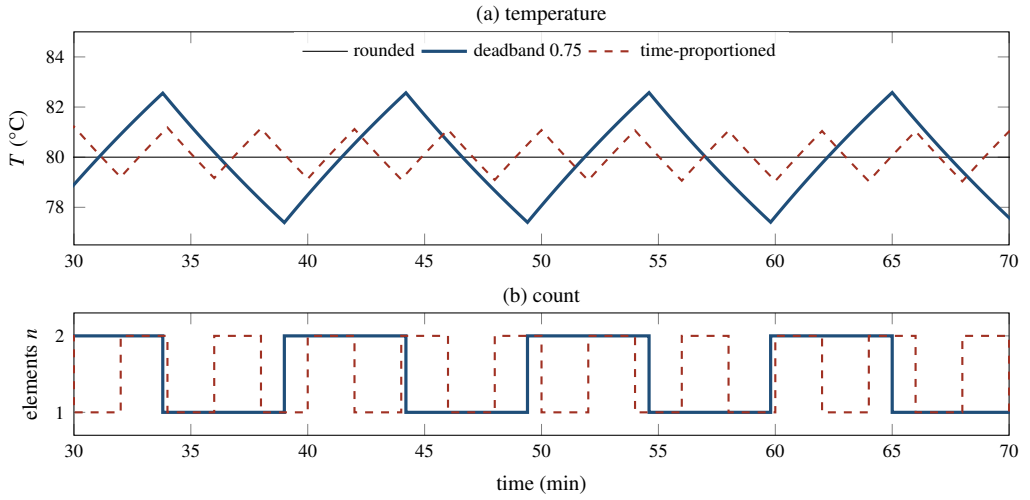


Figure 1.6 The Kettle Held at 80 °C by a PI Loop on Three Contactor Elements. (a) Temperature With the Demand Rounded (Thin Solid), Moved Only Past a Deadband of 0.75 Element (Thick Solid), or Time-Proportioned Over 4 min (Dashed). (b) The Element Count of the Last Two.

resumes it 0.3 K below that threshold, and latches its high-high interlock. Chapter 15 is a hybrid system by nature, smooth flight broken by impacts at which the ball's velocity jumps.

1.4.8 Integer Actuators

An actuator that comes in whole units, an element, a compressor, a server or a column, cannot take the value a PI loop asks for. Let the kettle's PI, with its 0.1 min period, compute a continuous demand of 1.5 elements at 80 °C and let contactors apply an integer count. Figure 1.6 shows three ways to do it. Rounding the demand changes the count at almost every scan, 300 times an hour, and holds 80 °C within 0.1 °C: it is the thermostat with no band. A deadband that moves the count only when the demand leaves the present count by more than 0.75 element changes it 11.5 times an hour and lets the kettle swing from 77.4 to 82.6 °C, and a deadband of 1.0 gives 5.9 changes and 74.9 to 85.0 °C. Time-proportioning switches the second element on for the demand's fractional part of each 4 min period, 30 changes an hour, and holds 79.0 to 81.2 °C. In each case the integral term finds the mean: with elements 10 % weak the deadband and time-proportioning hold 1.654 and 1.659 elements against the 1.667 required.

Every integer actuator trades swing against switching, and the right point depends on what a switch costs: wear on a contactor, a start on a compressor, money on a server. Chapter 14 orders servers only when its PI's demand leaves a deadband around the present count, adds a cooldown that stops flapping, and back-calculates the integral against the count actually applied, which also has a boot dead time. Chapter 12 stages two fixed compressors around a variable one, so that a continuous actuator absorbs what the integer ones cannot, and compares a scheduler that plans the starts. Volume 9 treats scheduling of integer inputs by optimization.

Table 1.2 Maps From One Cycle to the Next and Their Eigenvalues on the Kettle

Cycle and map	Eigenvalue μ	On the kettle
Batch, EWMA correction	$1-\lambda\xi$	0.7 or 0.3 for $\lambda = 0.3$ or 0.7, $\xi = 1$
Batch, learning update	$\mu(\omega)$ of Eq. (1.11)	0.1 to 0.98 with $\gamma = 1$, up to 1.036 with a short lead
Period, fixed duty	$e^{-t_p/\tau}$	0.8187 for $t_p = 4$ min
Period, minimum dwell	$e^{-2t_d/\tau}$	0.741 for $t_d = 3$ min
Period, thermostat	0	Back on the cycle after one switch
Bounce, impact map	Modulus of the restitution inside the window	Chapter 15

+ DID YOU KNOW? A Small Coil Switches a Large Motor

A contactor is a large relay whose contacts carry a motor's inrush current over many start and stop cycles. One rated at 75 horsepower on 480-volt three-phase power has a coil rated for 120 volts, so a small switch or a programmable controller can start the motor [27]. The kettle's contactors of Section 1.4.8 switch its elements the same way.

1.4.9 Cyclic Steady State

A repeated operation settles, if it settles at all, into a cycle rather than a point. Sampled once per cycle, the state follows a map $x_{c+1} = \Phi(x_c)$, and the cyclic steady state is its fixed point x^*. Near it a deviation is multiplied each cycle by the Jacobian of Φ, so the cycle is stable when every eigenvalue μ lies inside the unit circle, and $|\mu|$ says how fast it is reached [28]. For the time-proportioned element of Section 1.4.8 at a fixed 50 percent duty and a period $t_p = 4$ min, Equation (1.1) gives the map of the temperature at the start of each period, Equation (1.13).

$$T_{c+1} = e^{-t_p/\tau}\,T_c + 14.32 = 0.8187\,T_c + 14.32 \tag{1.13}$$

Its fixed point is a cycle from 79.0 to 81.0 °C, and a deviation falls to 10 % only after 12 periods, 48 min. The thermostat of Section 1.4.7 has the same cycle but a map with $\mu = 0$, because each switch happens at a fixed temperature and forgets how the kettle got there. After a 5 °C drop it needs one long on time, 5.5 min from 75 to 81 °C, and is back on its cycle. The difference goes further. With elements 10 % weak the fixed duty settles at 73.1 to 74.9 °C, while the thermostat keeps 79 to 81 °C at a duty of 0.67. A switch triggered by the state is feedback. A switch triggered by a clock is open loop.

Table 1.2 lists the cycle-to-cycle maps of this section, of Sections 1.4.5 and 1.4.6, and of the chapters that compute their own. The impact map of Chapter 15 is stable only while the paddle is slowing at impact, and over most of that window its eigenvalues have a modulus equal to the ball's restitution. In Chapter 10 heat accumulates from layer to layer, so the layers repeat only as far as the part's temperature does. Chapter 11 rotates three columns through a cycle whose feasibility it checks at each switch, and in Chapter 12 two cases on one suction header close a gap of 0.39 of a cycle to 0.10 by their fifth cut-in, a phase map that attracts.

1.4.10 Strategies and Safe Switching in Each Chapter

The sections above give each chapter two questions. How is the product brought to its end: by a trajectory, by an endpoint rule, by a correction from one run to the next, or by learning? And where the plant switches, what keeps the switching safe? Table 1.3 lists the answers each chapter tests, as groundwork for the comparison in Chapter 16. The measured results are left to the chapters.

Table 1.3 What Each Chapter Judges at the End, the Strategies It Compares and What Keeps Its Switching Safe

Ch.	Judged at the end	Strategies compared	Switching kept safe by
2	The whole dose, with the temperature after a cooling failure below a limit throughout	Fixed, temperature-limited and accumulation-limited dosing, one-step optimizer, MPC	Feed cut with hysteresis, latched interlock
3	Amount and purity in the receiver	Constant reflux, constant composition, optimized ramp, MPC	Policy change with tracking
4	Size and fines of the filtered crystals	Three cooling curves, cascade, supersaturation feedback, MPC	Moving output limit, transfer on a change of structure
5	Every vial dry, none collapsed	Fixed recipe, probe-limited, endpoint-aware with a learned resistance, MPC	Endpoint rules held for a set time
6	Cure of the slowest part, exotherm of the hottest core	Air loop, lead/lag selector, observer feedforward, optimized recipe, MPC	Selector with external reset
7	Mass of every part	Phase sequencer, shot-to-shot correction	Switchover by tracking, time limit behind each trigger
8	Four layer thicknesses	Timer, integrated estimate, estimate corrected by a load cell	Ramped gas changes, timeout fault
9	Fill of every vial	Pooled, per-head and record-based corrections	Validity window, step limit
10	Melt pool width along every layer	Feedback, learning, both	Q-filter, profile kept or reset at a new outline
11	Antibody pooled per liter of resin and hour, antibody lost	Time, load, breakthrough and learned-capacity triggers	Switch only when the next column is ready, pressure override
12	Goods temperature, energy, starts	Thermostats, staging, floating pressures, hybrid MPC	Band, minimum run and rest, desynchronization
13	Hold of every parcel	Cascade with feedforward under a supervisor	Diversion at once, return after a flush
14	Latency percentile and cost	Platform rule, PI, forecast floor, MPC	Deadband, cooldown, admission control
15	The period-1 bounce	Open-loop sinusoid, once-per-bounce PI, mirror law	Paddle slowing at impact, inside the window

DID YOU KNOW? A Quarter of Batch Control on Personal Computers

Batch operations run on stand-alone controllers, programmable logic controllers, distributed control systems and personal computers. A quarter of batch process control has been reported to involve personal computers, often in combination with the other platforms [6].

1.5 Instruments and Final Elements in This Volume

Figure 1.7 lists what the controllers of each chapter read and what they move. Temperature and pressure are each read in seven chapters, and a modulating valve or a flow loop is a final element in nine. The quality that decides the product is seldom among the readings, and only three chapters read a composition. Chapter 8 infers four layer thicknesses from a thermocouple and a load cell, and Chapter 2 infers the reagent that has piled up from two temperatures and a feed flow.

Weighing recurs in this volume. A load cell is one or more strain gauges bonded to a metal structure of precisely known elastic properties, so that the force on it becomes a change of electrical resistance [27], as Figure 1.8 shows. A scale closes the loop between shots of Chapter 7, tare cells and a checkweigher the loop between vials of Chapter 9, and a load cell under the coater the bed-mass correction of Chapter 8.

Many final elements here have only two states. A discrete control element has a limited number of

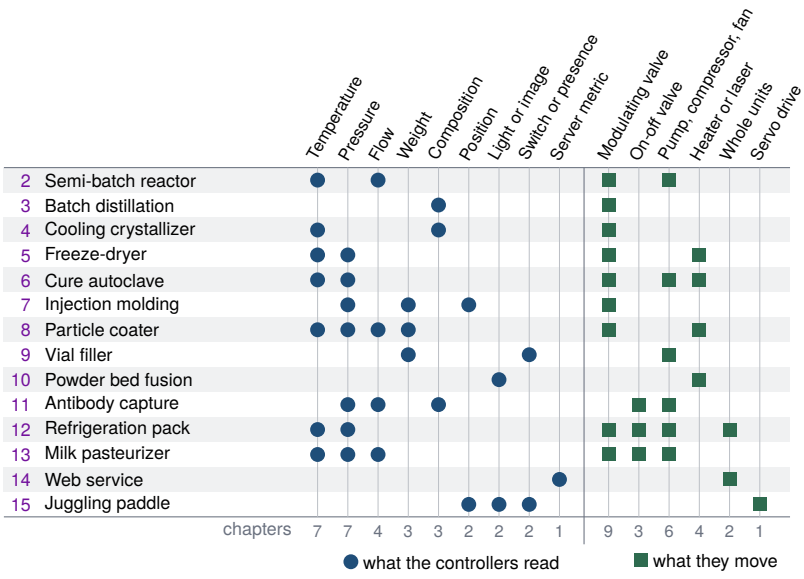


Figure 1.7 What the controllers of each chapter read (round marks) and move (square marks), from each chapter's variables table and technology section. Pressure includes differential pressure, a flow loop counts as a modulating valve, a heater includes the freeze-dryer's thermal unit, and a computed value, such as an inferred thickness, is not marked. Notice the scales, switches and whole units of the repeated and switched plants.

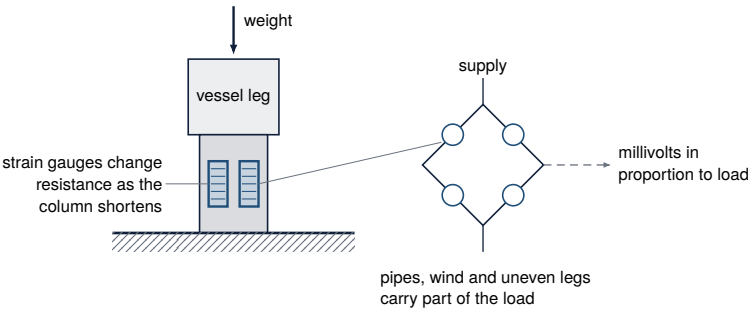


Figure 1.8 Principle of a strain-gauge load cell under a vessel leg. Notice that pipes, wind and uneven legs can carry part of the load past the cell, which then reads less than the true weight.

states, usually on and off, and the solenoid valve, a coil that pulls a ferrous armature against a return spring, is a very common form of on-off valve [27]. The case valves of Chapter 12, the diversion device of Chapter 13 and the valve block of Chapter 11 switch between positions in this way, and the staged compressors and the servers of Chapter 14 come in whole units, the integer actuators of Section 1.4.8.

The controllers sit in the layers that the batch control standards name. Their functional model relates recipe management, scheduling, sequential control, regulatory control and safety interlock systems [29]. Batch operations are controlled from stand-alone controllers, programmable logic controllers, distributed control systems and personal computers [6], and supervisory and lower-level loops run with cycle times from milliseconds to minutes [30]. Figure 1.9 places the loops of this volume on their platforms.

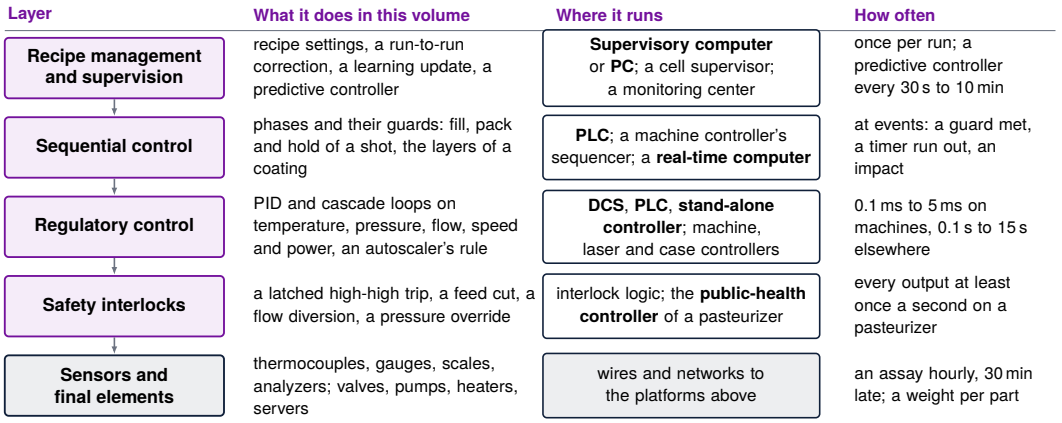


Figure 1.9 The control layers of this volume, named after the functional model of the batch control standards, the platform that hosts each and how often each acts, from [6, 29, 30] and the chapters' technology sections. Notice that a correction between runs acts once per run, however fast the loops beneath it.

Table 1.4 What Every App Provides. Standalone Apps Are Described in Their Own Chapters.

Element	What it does
Run, Pause, Reset	Start, hold and restart the run. Reset keeps the reader's settings.
Speed	Simulated time per second of real time. It changes the pace and never the result.
Manual and Auto	Hand each actuator to the reader or to the controller, without a bump.
Scenario list	Numbered cases, each with a briefing, a completion rule and a reload button.
Control panel	Setpoints, manual outputs, tuning, options and disturbances, at a Beginner or an Advanced level.
Trends	SP and PV above the MV for each loop, with limits, event marks and a cursor.
Scorecard	IAE, overshoot, settling time, MV travel and time at a limit for each loop.
Learn card	Predict, Try It, two hints and Explain, opened by the reader in that order.
Show Model	Equations, live variable values, editable parameters and assumptions.
Export CSV	Every logged sample of the run as a text file with a header row.
Instructions	A separate page with background, an operating guide and the model's limitations.

1.6 The Simulator Framework

Every chapter of this book has an app that runs in a web browser, collected with the others at https://apmonitor.com/apps/. About 240 apps exist across the series. All but a few are built on one shared framework, so the engine, the PID algorithm, the trends and the data export behave the same way in every chapter. The exceptions are standalone studios, such as the cascade studio and the two Temperature Control Lab apps, which follow the same ideas in their own code. This section describes the framework once and states where the standalone apps differ. Table 1.4 lists what a reader meets.

1.6.1 The Simulation Engine

Each app carries a plant model written as ordinary differential equations $\frac{dx}{dt} = f(x,u,d,p)$ in the states x, the inputs u, the disturbances d and the parameters p. The engine integrates the model with the classical fourth-order Runge–Kutta method on a fixed step h set for the fastest dynamics of the app. Some apps supply their own step, such as explicit Euler for a discontinuous right-hand side or an exact update for a model driven by events.

Three clocks run at separate rates. The plant advances every step h. The controller executes every control period Δt, equal to h unless the app declares a slower controller, and holds its outputs between executions. The log takes one sample at a fixed interval of simulated time for the trends, the scorecard

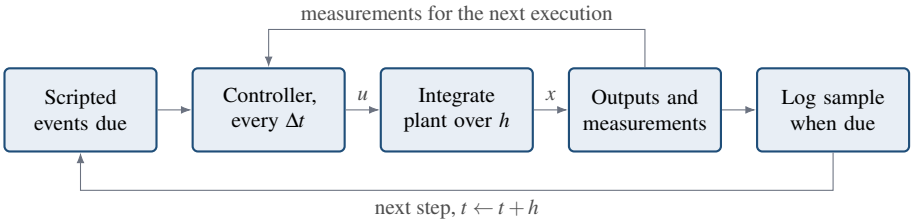


Figure 1.10 Work of the Engine in One Integration Step

and the exported file. The order of work in one step is shown in Figure 1.10. The controller also executes once when a scenario is loaded, so every run starts from an initialized controller.

Drawing the screen is kept apart from these clocks. A speed of 5 means five units of the app's time base per second of real time. Speed, frame rate and pauses change how fast a run is shown and never what it computes, so a phone and a workstation produce the same trajectory and the same file.

Noise comes from a deterministic pseudorandom generator. Its seed is declared by the app, or by a scenario that needs its own noise record, and it is restored whenever a scenario is loaded or Reset is pressed. The framework gives the reader no seed setting, so *default seed* in this book means the seed the app ships with. A run repeats bit for bit when the same scenario is loaded and the same settings are made before Run. A change made by hand during a run lands on a different step each time, which is one reason numeric answers carry a tolerance. The standalone simulators draw noise from the browser's own generator, so a noisy run there repeats in its statistics and not sample for sample.

Run and Pause start and hold the clock. Reset restarts the run at $t = 0$, paused, and keeps the reader's settings: every setpoint, tuning value, option, model parameter, disturbance slider and mode the reader changed is put back before the plant and the controller are initialized. Scripted events replay, and a one-shot disturbance button pressed during the run is cleared. The reload button beside the scenario list restores the scenario's own settings instead.

1.6.2 The Velocity-Form PID Algorithm

One PID implementation serves the PID loops of every framework app. It is the velocity (incremental) form: each execution adds the change in the proportional and derivative terms to a running sum, and the integral is a separate term, as shown in Equations (1.14) to (1.17) for execution k and control period Δt.

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k, \qquad D_k = -a\,K_c\,\tau_D\,\frac{\mathrm{PV}^f_k - \mathrm{PV}^f_{k-1}}{\Delta t} \tag{1.14}$$

$$S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \tag{1.15}$$

$$v_k = u_b + S_k + I_{k-1}, \qquad u_k = \mathrm{sat}(v_k) \tag{1.16}$$

$$I_k = I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{T_t}\right]\Delta t \tag{1.17}$$

where $a = \pm 1$ is the controller action, PV^f is the measurement after an optional first-order filter, u_b is the output bias, S is the running sum of proportional and derivative increments, v is the output before limiting, u is the output sent to the actuator, I is the integral term and T_t is the tracking time constant, equal to τ_I unless a loop sets another value. The function sat clamps v_k to the actuator range and then applies a rate limit if the loop declares one.

The limit acts on the emitted output u_k only. The sum S_k is never clamped, so it equals $P_k + D_k$ and the output before limiting is the position form $u_b + K_c e_k + D_k + I_{k-1}$. Anti-windup touches the integral alone. While the output is limited, the last term of Equation (1.17) drives I toward the value that reproduces the output applied, which is back-calculation [31]. With no integral action,

Equation (1.17) is skipped. Clamping the running sum would instead discard each increment that crosses a limit and keep each one that returns, so symmetric noise would push the output off the limit even in a proportional-only loop.

The derivative acts on the measurement, so a setpoint step gives a proportional kick and no derivative kick. A change of K_c acts on the whole error at the next execution, as in the position form. The action $a = +1$ suits a process whose PV rises with the MV, a reverse-acting controller in the ISA convention, and $a = -1$ gives the direct-acting one.

In Manual the controller keeps executing and sets $I_k = u_{\text{man}} - u_b - S_k$, so the output it would send equals the manual value u_{man}. The first automatic output differs from it by one period's change in P and D. In Auto the engine copies each controller output to the manual station, so a switch to Manual is also bumpless. The standalone studios with PID loops use velocity-form code of their own, with the limit on the output and anti-windup on the integral alone, in which a gain change applies only to later increments.

1.6.3 The Predictive Control and Estimator Options

More than 100 apps offer model predictive control (MPC) as an option beside the structure they open with, which stays the default. Volume 9 develops the theory. In most of them the controller plans with a small linear model in discrete time as shown in Equation (1.18).

$$x_{k+1} = A x_k + B u_k + E w_k + c, \qquad y_k = C x_k \tag{1.18}$$

where w is a measured or scheduled disturbance and c is a constant left by linearization. The model comes from a step test at the design point, from the app's balances linearized there or again at every decision, or from the model the reader fits. It never reads the true states or parameters of the run, so a parameter the reader changes in Show Model is model mismatch, as on a real plant.

At each decision the controller minimizes the objective of Equation (1.19) over a horizon of N steps.

$$J = \sum_{k=1}^{N} (y_k - r_k)^{\top} Q (y_k - r_k) + \sum_{j} \Delta u_j^{\top} R \Delta u_j + \sum_{i} \rho_i s_i^2 \tag{1.19}$$

where r is the setpoint, Q weights the tracking error, R is the move suppression on each input change Δu_j and ρ_i is the price of a slack $s_i \geq 0$ that lets an output band be broken when it cannot be held. Actuator limits are hard constraints. The quadratic program is solved in the browser by a dual active-set method [32]. The first move is applied, and the plan is made again at the next decision.

The plan starts from an estimate of the model state. About half of these apps use a steady-state Kalman filter on the model with an added disturbance state, which absorbs what the model did not predict and removes steady offset without an integral term [33]. Others correct the model output with a filtered bias to the same end. About a dozen apps also offer moving horizon estimation (MHE), which solves the estimation problem over a window of recent samples and can hold bounds that a Kalman filter cannot, such as a concentration that is never negative [34].

With the MPC selected, the tuning fields of the loops it replaces leave the Control panel. In most of these apps the Advanced level shows up to three settings, each a multiple of the shipped value: the prediction horizon, the move suppression and the band price. The trend draws the accepted plan ahead of the present time, and a list of constraint prices shows which limits bind. A plan that is not optimal, not finite or outside its limits is rejected, and a fallback, in most apps the shipped controller, holds the actuator until a plan succeeds. The Model Fitting and Tuning Studio builds the same kind of controller from the reader's own data in its tab, MPC from the fit.

1.6.4 Scenarios, Progress and Data

An app teaches through scenarios, the numbered cases at the top of the Control panel, nine to thirteen in most apps. Loading one sets the model parameters, the controller settings, the disturbances, the mode, the initial state and any scripted events. The information button beside the list opens a briefing

in four parts: what the case is, why it matters, what to do and one detail for the practicing engineer. The Learn card holds the lesson back, with Predict and Try It before the run, hints on request and Explain afterward.

Each scenario has a completion rule that the app judges from the run on the simulated clock: a run length, the scripted events passed, a setting made by the reader, a state held for a stated time or a target met. Choosing a case earns nothing. A completed case is marked in the list, and a badge row shows the count and a level, Explorer after one case, Practitioner at half and Completed at all. Progress is kept in the reader's browser with no account. A certificate of progress or of completion can be issued from it, as a record of self-paced work and not an assessed credential.

One case, usually the last, is the graded challenge. It states a target, such as an error band to hold through a sequence of upsets. Changing the model parameters or the disturbances voids the attempt in most apps, and a record made with Reveal suggested tuning says so.

Show Model opens the equations, live values of the variables, the editable parameters and the assumptions. The trends plot SP and PV above the MV for each loop, with actuator limits and marks at setpoint moves, disturbances, mode changes and scripted events. The scorecard reports for each loop the integral of the absolute error (IAE), overshoot, settling time, MV travel and share of time at a limit, with the app's own measures, and it does not combine them into one score. IAE and travel accumulate from Reset, so two runs are compared at the same clock reading.

Export CSV saves the log as a text file with a header row of column names that carry their units and one row per logged sample. In most apps the time column is in seconds, minutes or hours according to the run length. The log holds the most recent 60 000 samples unless the app sets another size. Apps with the MPC option begin the file with comment lines that start with #, which a script must skip.

An app exercise in this book names the scenario by its number and name as the app shows them, every setting that differs from the scenario's defaults, the run length in simulated time, the seed (the default seed unless stated), what to read and where (a trend, a scorecard tile or a column of the exported file) and a tolerance on a numeric answer. Settings are made before Run unless the exercise gives a time.

1.7 Working With AI Agents While Studying

A reader of this book is likely to have an AI assistant within reach, and the apps make its work easy to test. This section gives the editor's view of how to use one while working through a chapter. In short, the assistant may do the algebra and the scripting, and the reader keeps the checking.

What to delegate. Tasks with a clear input and a result that can be tested are good candidates. An assistant can derive the linearization of a chapter's balance at an operating point and the transfer function that follows from it. It can write a short script that reads the exported CSV file and fits a first-order-plus-dead-time model to a step test. It can propose a first tuning from that fit with a named rule, convert units, explain a term in a model or outline a report. The assistant knows nothing about the app until it is told. Give it the equations and parameter values from Show Model, the column names of the exported file and the scenario that produced the data, as described in Section 1.6.4. Ask it to state its assumptions and the units of every result.

What to check, and how. An assistant returns a fluent answer whether or not the answer is right, so each result needs a test that it could fail. Four tests cover most of the errors a chapter will produce.

- *Units.* Every term of an equation carries the same units, and a process gain is in PV units per MV unit. A gain computed with the valve as a fraction is 100 times the gain of a loop whose valve is in percent.
- *Signs.* Decide from the physics which way the PV moves when the MV rises. Compare that with the sign of the fitted gain and with the controller action of Section 1.6.2.

Table 1.5 Roadmap of the Volume

Ch.	Plant	The lesson	Level	Time
2	Semi-batch exothermic reactor	Accumulated reagent as a path constraint. Dosing as the safety layer.	intermediate	3 h
3	Batch distillation column	No operating point. An open-loop reflux trajectory against a composition loop.	intermediate	3.5 h
4	Seeded cooling crystallizer	The shape of the trajectory decides the product, not the tuning.	intermediate	3 h
5	Pharmaceutical freeze-dryer	Hottest vials set the limit, slowest the endpoint. An estimator that learns.	advanced	2.5 h
6	Composite cure autoclave	Lead/lag selection among twelve thermocouples, an exotherm the selector cannot stop.	advanced	2.5 h
7	Injection molding machine	Bumpless switchover inside each shot, correction between shots.	introductory	2 h
8	Fuel particle coater	A recipe as a state machine. Phases ended by an estimate and its failures.	intermediate	3 h
9	Four-head vial filler	Run-to-run correction per head with results in flight.	introductory	2 h
10	Laser powder bed fusion	Learning a power profile layer by layer, and why the Q-filter matters.	introductory	2 h
11	Three-column antibody capture	When to switch a column whose capacity is unmeasured. A cyclic schedule.	advanced	2 h
12	Supermarket refrigeration pack	Hysteresis thermostats that synchronize, staged compressors.	intermediate	2.5 h
13	Milk pasteurizer	Product history, regeneration and a diversion supervisor.	introductory	2 h
14	Autoscaled web service	Integer servers with a boot dead time, flapping and retry storms.	intermediate	2.5 h
15	Juggling robot paddle	A hybrid orbit stable through its impact map, with or without a sensor.	advanced	2.5 h

- *Steady state.* Set the derivatives to zero. The model must return the operating point that Show Model reports, and a linearization must give zero deviation there.
- *A run on the app.* Load the named scenario at the default seed, enter the proposed numbers and run. The trend and the scorecard decide. A tuning that reads well on paper and holds the valve at a limit for half the run has failed.

A fitted model gets one more test: compare it with data it was not fitted to, such as a second step in the opposite direction. When a check fails, report the failure to the assistant with the evidence and ask for the cause, then check the correction the same way.

How the exercises train the habit. One exercise in each chapter is marked *Review the agent*. It presents a tuning, a model or a claim as an AI assistant might propose it, with a flaw, and asks the reader to find the flaw in about five minutes. The flaws are ordinary ones: a unit, a sign, a steady state that does not close, a result measured on the wrong scenario or at a different run length, or a limit that was ignored. Each is found with one of the tests above. The purpose is not distrust. An engineer who can say why an answer is right can hand an assistant much larger tasks and still stand behind the result. That larger picture, in which agents do most of the routine work and the engineer supervises and checks, is the subject of Chapter 16.

1.8 Chapter Roadmap

Table 1.5 lists the chapters with the lesson each adds to this one. The level says how much of Section 1.4 a chapter assumes. The time is an estimate for reading the chapter once and running each case of its app once, at about eight minutes a case. Figure 1.11 places the same chapters by the kind of operation that Section 1.1 sorts them into.

One batch along a trajectory	Repeated runs corrected from the last	Switched actuators or dynamics
Sections 1.4.2 to 1.4.4	Sections 1.4.2, 1.4.5, 1.4.6	Sections 1.4.7 to 1.4.9
(2) unreacted reagent as a path limit	(7) switchover inside a shot, correction between shots	(12) thermostats that fall into step
(3) an open-loop reflux trajectory against a loop	(8) a recipe whose phases end on an estimate	(13) a diversion supervisor on product history
(4) the shape of the cooling curve decides the product	(9) a correction per head with results in flight	(14) integer servers with a boot dead time
(5) hottest vials set the limit, slowest the end	(10) a power profile learned layer by layer	(15) a hybrid orbit stable through its impacts
(6) a selector among twelve thermocouples	(11) when to switch a column of unmeasured capacity	

Figure 1.11 Method map of the volume: each chapter by the kind of operation of Section 1.1, with its lesson from Table 1.5 and the subsections it rests on. Notice that the recipe sections serve both the single batch and the repeated run.

Chapters 7, 9, 10 and 13 form a first course: a sequencer with a switchover, a run-to-run correction, a learning update and a supervisor, each on a plant that is quick to run. They rest on Sections 1.4.2, 1.4.5, 1.4.6 and 1.4.7. Chapters 2 to 6 are best read after Sections 1.4.3 and 1.4.4, in any order, and Chapter 8 after them. Chapters 12, 14 and 15 rest on Sections 1.4.7 to 1.4.9, and Chapter 11 draws on every section. A reader who knows batch operation can take a shorter path through Chapters 2, 3, 9, 10 and 15.

KEY POINTS

- A batch has no steady state, and its product is judged once, after the end.
- Each phase ends on a clock, a measured event or an estimate.
- Correcting the next run from the last pays only against disturbances that persist.
- A switch triggered by the state is feedback, and one triggered by a clock is open loop.

Where else this applies. Wafers corrected run to run in Volume 14, Chapter 6, and a tap changer behind a deadband in Volume 17, Chapter 8.

References

[1] D. Bonvin, "Optimal operation of batch reactors—a personal view," *Journal of Process Control*, vol. 8, no. 5–6, pp. 355–368, 1998.

[2] D. Bonvin, B. Srinivasan, and D. Hunkeler, "Control and optimization of batch processes," *IEEE Control Systems Magazine*, vol. 26, no. 6, pp. 34–45, 2006.

[3] J. H. Lee and K. S. Lee, "Iterative learning control applied to batch processes: An overview," *Control Engineering Practice*, vol. 15, no. 10, pp. 1306–1318, 2007.

[4] L. Rayleigh, "On the distillation of binary mixtures," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 4, no. 23, pp. 521–537, 1902.

[5] H. A. Miers and F. Isaac, "The refractive indices of crystallising solutions, with especial reference to the passage from the metastable to the labile condition," *Journal of the Chemical Society, Transactions*, vol. 89, pp. 413–454, 1906.

[6] D. Bonvin and G. François, "Control and optimization of batch chemical processes," in *Coulson and Richardson's Chemical Engineering*. Elsevier, 2017, pp. 441–503.

[7] S. Arimoto, S. Kawamura, and F. Miyazaki, "Bettering operation of robots by learning," *Journal of Robotic Systems*, vol. 1, no. 2, pp. 123–140, 1984.

[8] J. W. Hou, "Optimal cure cycle design for autoclave processing of thick composites laminates: A feasibility study," Old Dominion University, for NASA Langley Research Center, Norfolk, VA, Progress Report, NASA Contractor Report NASA-CR-181261, Nov. 1985, NASA Technical Reports Server 19870018301.

[9] M. Buhler, D. E. Koditschek, and P. J. Kindlmann, "A family of robot control strategies for intermittent dynamical environments," *IEEE Control Systems Magazine*, vol. 10, no. 2, pp. 16–22, 1990.

[10] E. Sachs, A. Hu, and A. Ingolfsson, "Run by run process control: Combining SPC and feedback control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 8, no. 1, pp. 26–43, 1995.

[11] T. A. Henzinger, "The theory of hybrid automata," in *Proceedings 11th Annual IEEE Symposium on Logic in Computer Science*, 1996, pp. 278–292.

[12] U.S. Food and Drug Administration, "Guidance for industry. PAT: A framework for innovative pharmaceutical development, manufacturing, and quality assurance," U.S. Department of Health and Human Services, Food and Drug Administration, Rockville, MD, Tech. Rep., Sep. 2004.

[13] C. Sonntag, A. Devanathan, and S. Engell, "Hybrid NMPC of a supermarket refrigeration system using sequential optimization," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 13 901–13 906, 2008.

[14] P. Mell and T. Grance, "The NIST definition of cloud computing," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-145, 2011.

[15] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, "ICH harmonised guideline Q13: Continuous manufacturing of drug substances and drug products," Adopted 16 November 2022, 2022. [Online]. Available: https://database.ich.org/sites/default/files/ICH_Q13_Step4_Guideline_2022_1116.pdf

[16] A. Ingolfsson and E. Sachs, "Stability and sensitivity of an EWMA controller," *Journal of Quality Technology*, vol. 25, no. 4, pp. 271–287, 1993.

[17] R. Hanus, M. Kinnaert, and J.-L. Henrotte, "Conditioning technique, a general anti-windup and bumpless transfer method," *Automatica*, vol. 23, no. 6, pp. 729–739, 1987.

[18] B. Srinivasan, S. Palanki, and D. Bonvin, "Dynamic optimization of batch processes: I. characterization of the nominal solution," *Computers & Chemical Engineering*, vol. 27, no. 1, pp. 1–26, 2003.

[19] E. Del Castillo and A. M. Hurwitz, "Run-to-run process control: Literature review and extensions," *Journal of Quality Technology*, vol. 29, no. 2, pp. 184–196, 1997.

[20] J. F. MacGregor, "A different view of the funnel experiment," *Journal of Quality Technology*, vol. 22, no. 4, pp. 255–259, 1990.

[21] R. P. Good and S. J. Qin, "On the stability of MIMO EWMA run-to-run controllers with metrology delay," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 1, pp. 78–86, 2006.

[22] D. A. Bristow, M. Tharayil, and A. G. Alleyne, "A survey of iterative learning control," *IEEE Control Systems Magazine*, vol. 26, no. 3, pp. 96–114, 2006.

[23] M. Norrlöf and S. Gunnarsson, "Time and frequency domain convergence properties in iterative learning control," *International Journal of Control*, vol. 75, no. 14, pp. 1114–1126, 2002.

[24] R. W. Longman, "Iterative learning control and repetitive control for engineering practice," *International Journal of Control*, vol. 73, no. 10, pp. 930–954, 2000.

[25] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Systems Magazine*, vol. 29, no. 2, pp. 28–93, 2009.

[26] D. Liberzon and A. S. Morse, "Basic problems in stability and design of switched systems," *IEEE Control Systems Magazine*, vol. 19, no. 5, pp. 59–70, 1999.

[27] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[28] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, ser. Applied Mathematical Sciences. Springer, 1983.

[29] International Society of Automation, "ISA88: Batch control, committee scope and purpose," www.isa.org, standards committees, 2026, accessed 22 November 2026.

[30] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[31] K. J. Åström and L. Rundqwist, "Integrator windup and how to avoid it," in *Proc. 1989 American Control Conference*, 1989, pp. 1693–1698.

[32] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic programs," *Mathematical Programming*, vol. 27, no. 1, pp. 1–33, 1983.

[33] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[34] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003.

1.9 Exercises

1.1 Theory and calculation · 4 min

The kettle's dose is Equation (1.2), in which one minute at 80 °C counts as one minute and each 10 °C hotter counts ten times as much.

(a) Find the dose of a hold of 6 min at 85 °C. *Checkpoint:* 18.97 min.

(b) Find the dose of a heat-up tracked exactly from 20 to 80 °C at 4 °C/min. *Checkpoint:* 1.09 min.

(c) Find how long a hold at 75 °C must last to deliver 30 min of dose. *Checkpoint:* 94.9 min.

1.2 Concept check · 2 min

The kettle's thermocouple reads 1 °C high. The hold can end on a clock set to the nominal 26.3 min or on the dose estimate with the predicted cooling dose, as in Section 1.4.2. Which statement is right?

(a) ending on the dose estimate removes the error, because the estimate integrates the measured temperature

(b) neither guard removes it, because both rely on the same reading, so only a test of the product shows it

(c) ending on the clock removes the error, because a clock does not depend on the thermocouple at all

(d) either guard removes it, because the PI loop's integral action cancels a constant measurement bias

1.3 Theory and calculation · 5 min

The kettle of Equation (1.1) is heated on a ramp by a PI loop with $\tau_I = \tau = 20\,\text{min}$ and $K_c K = 5$.

(a) Find the closed-loop time constant and the lag on a ramp of 1.5 °C/min. *Checkpoint:* 4 min and 6 °C.

(b) Find the feedforward of Equation (1.4), with $\hat{K} = K = 40\,°\text{C}$ per element, when that ramp's setpoint passes 50 °C. *Checkpoint:* 1.5 elements.

1.4 Theory and calculation · 5 min

The product's limit is lowered to 87 °C, and the fast policy of Section 1.4.4 keeps the same 2 °C back-off.

(a) Find the time to heat from 20 to the switching temperature with all three elements on. *Checkpoint:* 15.60 min.

(b) Find the holding input at the switching temperature. *Checkpoint:* 1.625 elements.

(c) Find the time of free cooling from the switching temperature to 40 °C. *Checkpoint:* 23.57 min.

1.5 Theory and calculation · 6 min

An EWMA controller, Equation (1.7), runs with $\lambda = 0.4$ and a model gain $b = 1$.

(a) With $\beta = 1$, a step of -2 in d arrives. Find the error left after three corrections. *Checkpoint:* -0.432.

(b) Find the steady offset under a drift of -0.2 per batch. *Checkpoint:* -0.5.

(c) Find the standard deviation of the result when the only disturbance is white noise with a standard deviation of 0.5. *Checkpoint:* 0.559.

(d) The true gain is $\beta = 1.5$. Find the factor by which a step shrinks each batch, and the largest λ that keeps the loop stable. *Checkpoint:* 0.4 and 1.33.

1.6 Concept check · 2 min

A filler's results arrive one batch late, and its correction is the integral form $u_{j+1} = u_j + (\lambda/b)(y^* - y_{j-1})$ with $\lambda\xi = 0.8$. Which statement is right?

(a) it is unstable, because 0.8 exceeds the limit of 0.618 for a delay
(b) it is stable and monotone, because 0.8 is below the limit of 2
(c) it is unstable, because any delay destabilizes an integral correction
(d) it is stable, with an error that oscillates and shrinks by 0.89 per batch

1.7 Theory and calculation · 4 min

A learning update, Equation (1.10), with $\ell = 1$ and no filter acts on the kettle sampled every minute, with the model gain exact, so that the factor of Equation (1.11) is $1 - \gamma$ at zero frequency and $1 - \gamma(1-p)/(1+p)$ at the shortest period, with $p = \mathrm{e}^{-1/20}$.

(a) Find both factors for $\gamma = 0.5$. *Checkpoint:* 0.5 and 0.9875.

(b) Find the factor at zero frequency for $\gamma = 1.5$ and say what it means for a constant error. *Checkpoint:* -0.5.

1.8 Theory and calculation · 4 min

The kettle holds 80 °C with one element always on and the second switched by a thermostat with $T_{lo} = 78\,^{\circ}\mathrm{C}$ and $T_{hi} = 82\,^{\circ}\mathrm{C}$. Ignore the scan and any measurement lag.

(a) Find the on and off times of the second element. *Checkpoint:* 4.01 min each.

(b) Find the period and the number of starts of the second element per hour. *Checkpoint:* 8.03 min, 7.47.

1.9 Concept check · 2 min

The kettle's elements have aged and deliver 10 % less heat. Its hold runs either on the thermostat of Section 1.4.7, with a band from 79 to 81 °C, or with the second element on for a fixed half of every 4 min period. Which statement is right?

(a) the thermostat still holds 79 to 81 °C, because each switch happens at a measured temperature
(b) the fixed duty still holds 79 to 81 °C, because a clock does not depend on the elements
(c) both still hold 79 to 81 °C, because both had that cycle with nominal elements
(d) neither holds the band, because only integral action can compensate a change of gain

1.10 Theory and calculation · 5 min

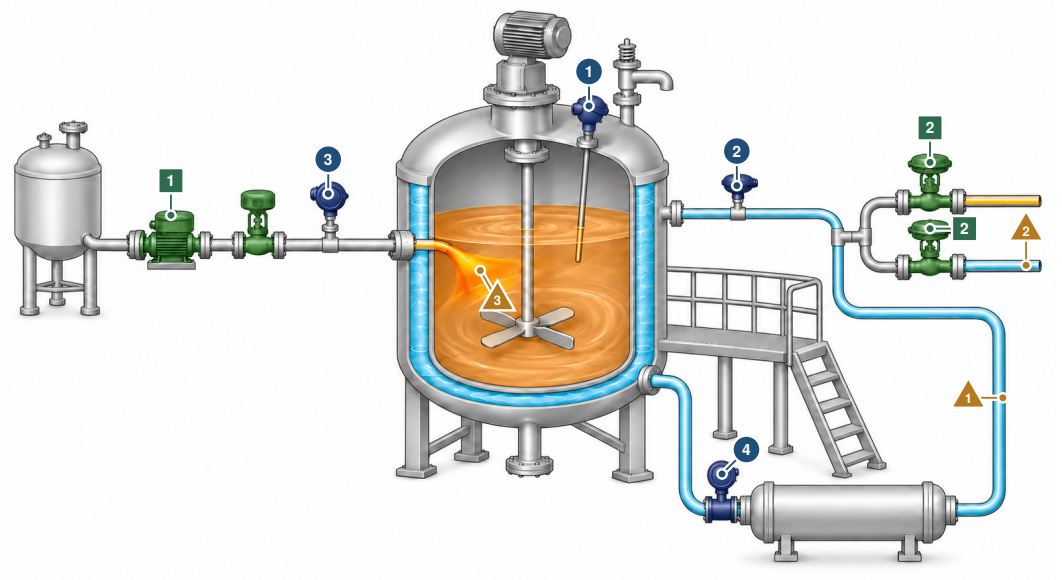
The second element is time-proportioned at a fixed 50 % duty with a period of 6 min.

(a) Find the eigenvalue of the map from one period's starting temperature to the next, and the number of periods that shrink a deviation to 10 %. *Checkpoint:* 0.741 and 8.

(b) Find the cyclic steady state, the lowest and highest temperature of the cycle. *Checkpoint:* 78.50 °C and 81.50 °C.

CHAPTER 2

Semi-Batch Exothermic Reactor Simulator With Dosing as the Safety Layer



The controller moves 1 pump that doses the second reagent 2 cooling-water and hot-water valves on the jacket loop

The controller reads 1 batch temperature 2 jacket temperature 3 feed flow meter 4 low-flow switch on the jacket loop

What disturbs it 1 jacket circulation lost 2 warmer cooling water 3 weaker catalyst, so fed reagent reacts later

Notice that the feed enters the stirred liquid while every utility stays outside the wall, so only the dosing pump decides how much unreacted reagent the batch holds. Illustration, not to scale.

IN PLAIN TERMS **The Danger Is the Feed Not Yet Used**

The main idea. This reactor is filled with one liquid and then fed a second one slowly, because the two give off heat as they react. Feeding slowly lets a cooling jacket around the tank keep up with that heat.

The challenge. If the tank is kept too cool, the fed liquid piles up without reacting while the thermometer looks perfect. Should the cooling then fail, the stored liquid reacts all at once and the tank can overheat until a safety disc bursts.

Why it matters. The feed controller, not the temperature controller, decides how much hidden heat a batch carries at any moment.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on a stirrer left off (Section 2.1), the history at the end of Section 2.2, the two try-its and the fact in Section 2.6, and the key points
Study	65 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the measured cases (Section 2.6, through the challenge), Technology in Practice (Section 2.7), Limitations (Section 2.8) and the open challenge

The control question A batch dosed cold can hold its setpoint within a fraction of a kelvin while the reagent fed into it piles up unreacted. What protects a semi-batch reactor against a cooling failure, the temperature loop or the dosing controller, and how far can that protection trust an estimate?

Learning objectives

- Compute the adiabatic temperature rise of the unreacted feed, the MTSR and the criticality class from the balances.
- Explain why the highest MTSR of a dose has a minimum in the process temperature, and why a colder batch can be the more dangerous one.
- Compare fixed, temperature-limited and accumulation-limited dosing through a cooling failure, a weak catalyst and warm cooling water.
- Predict the drift of a heat-flow calorimeter from the error in its assumed heat-transfer coefficient.
- Judge what a predictive dosing controller gains over a one-step optimizer, and what relinearizing along its plan does not fix.

Prerequisites Sections 1.4.2 (recipes and guards) and 1.4.4 (path constraints and back-off), 1.6.2 (the PI law) and 1.6.1 (engine, speed and seed). Section 1.6.3 is needed only for the predictive controller.

The app https://apmonitor.com/apps/semi-batch-reactor/

See also Volume 4, Chapter 2 for a protective override that rests at its clamp until a limit nears, Volume 4, Chapter 6 for a safety limit that sets the line speed, Volume 7, Chapter 2 for the ignition boundary of a cooled exothermic reactor, Volume 13, Chapter 4 for catalyst beds run close to an ignition floor, and Volume 13, Chapter 2 for an inventory that accumulates unseen until a purge removes it.

2.1 Introduction

A semi-batch reactor is charged with one reagent and fed the other, and when the reaction is strongly exothermic the feed rate is fixed before anything else, because the dose caps the rate at which heat can be released [1–3]. The recipe of this chapter doses 1200 kg of a reagent B, 8000 mol of it, into 3000 kg of a reagent A in solvent held at 70 °C. At 200 kJ/mol the dose carries 1.6 GJ, enough to heat the finished batch by 211.6 K if it were released at once. Spread over four hours it releases about 111 kW, against a jacket that could remove 194 kW with its cooling water at 8 °C.

The misconception this reactor corrects is that a batch on its temperature setpoint is a safe batch. What endangers it is not the heat being made but the heat not yet made, the B that has been fed and has not reacted [4]. If the cooling fails, that B reacts without heat removal and the batch climbs toward the maximum temperature of the synthesis reaction (MTSR), the quantity on which the thermal assessment of a cooling failure rests [5, 6]. Measured on the model, the recipe at 70 °C never holds more than 234 mol of unreacted B, and its MTSR never passes 76.2 °C. The same four-hour dose into a batch held at 30 °C keeps the true temperature between 30.0 and 30.2 °C. Yet at 2 h 30 min it holds 4398 mol of B unreacted and an MTSR of 160.4 °C, 30 K above the temperature at which the product's decomposition would reach its maximum rate within a day. The colder batch is the more dangerous one, and its temperature trend does not show it.

DID YOU KNOW? A Stirrer Left Off

In a semi-batch plant described by the British safety regulator, an operator filled the reactor with the first reagent and began adding the second without starting the stirrer. When the agitator was started, the reaction was so violent that the reactor burst from overpressure, and 11 people were killed [7].

The problem is one of a general class. A controller that has to limit an inventory it cannot measure, whose release after a failure is the hazard, appears wherever a reactor is fed rather than charged. Semi-batch emulsion polymerization has the same shape, where calorimetry estimates the conversion and feedback on it has kept monomer from accumulating after a deliberate inhibition [8], and so do

fed-batch systems in which online optimization is joined to runaway prediction and prevention [9]. In each case the estimate comes from a balance, and calorimetric state estimators for reactor diagnosis and control form a family of methods of their own [10]. Feed profiles for a second-order reaction under safety constraints have been optimized [11] and computed online from a heat balance [12]. A transfer function from feed to temperature carries none of this, because in it the feed is a load and the accumulation is not a state.

The thermal safety literature presents MTSR and the criticality classes as quantities of a design assessment [6], and the optimization literature presents the feed profile as the solution of a problem [12]. This chapter works instead with a reactor that holds its setpoint while its criticality class moves from 2 to 5 and then loses its cooling. The app computes the accumulation, MTSR and Stoessel's classes live from the true state, logs the calorimeter's estimate beside the truth, and ends a cooling failure at the rupture disc when the accumulation says it will. The kinetics, the criticality classes, heat-flow calorimetry, cascade and split-range control and accumulation-limited dosing are all standard methods. The model has 18 states, and its energy and mole ledgers close to round-off. Its decomposition is anchored on a declared threshold through the adiabatic time to maximum rate. Its calorimeter sees the accumulation only through the heat-transfer coefficient it assumes, and its four shipped dosing strategies act on one constraint. An optional predictive controller on the dose is measured against those strategies on the same runs, and a second setting of it, which relinearizes its model along its own plan, is measured against the single linearization. A sweep of the process temperature finds a minimum of the highest MTSR near 66 °C. The ten cases and the challenge separate the heat being made from the heat stored, the temperature controller from the safety layer, and the estimate from the truth.

The chapter is laid out as follows. Section 2.2 gives the history of accumulation, calorimetry and the cooling-failure scenario. Section 2.3 sets out the model, from the kinetics and the balances to the criticality classes and the events. Section 2.4 measures the highest MTSR against the process temperature. Section 2.5 describes the cascade, the calorimeter, the dosing strategies and the interlocks, with the predictive option in Sections 2.5.8 and 2.5.9. Section 2.6 runs the cases and the challenge and compares the predictive controller with the shipped strategies (Sections 2.6.9 and 2.6.10). Section 2.7 describes the plant's instruments, dosing pump and control system, Section 2.8 states the limitations, Section 2.9 concludes, and the exercises of Section 2.10 run the cases.

2.2 Accumulation, Calorimetry and the Cooling-Failure Scenario

Feeding one reagent into the other became the standard way to run a strongly exothermic liquid-phase reaction because it hands the release of heat to a pump. In 1986 the limits of safe operation of a semi-batch reactor were compared with those of a continuous stirred tank [1], and the conditions for thermally safe operation of a semibatch reactor were mapped for slow liquid-liquid reactions in 1990 [2]. Safety and runaway prevention in batch and semibatch reactors were reviewed together in 2006 [3].

The assessment rests on measurements of the decomposition that a runaway can set off. An accelerating rate calorimeter for thermal hazard evaluation was described in 1980 [13], and the adiabatic time to maximum rate that such an instrument measures defines T_{D24}, the temperature at which that time is 24 h. The same time can be estimated from dynamic differential scanning calorimetry (DSC) [14]. Chemical reaction engineering was turned toward safety in 1988 [5], and the hazards linked to accumulation in semi-batch reactors were assessed in 1996 [4]. Stoessel's textbook sets out the cooling-failure scenario, which orders the process temperature, MTSR, T_{D24} and the maximum temperature for technical reasons (MTT) into five criticality classes, together with its use in process design [6].

The estimate that a dosing controller needs came from calorimetry. Heat-flow calorimetry was developed as a tool for process optimization and process safety at once [15], and calorimetric state estimators, which infer the reaction rate from a reactor's heat balance, were reviewed for diagnosis

and control in 1992 [10]. In 1997 on-line calorimetry estimated the conversion of a semibatch emulsion copolymerization, and feedback on that estimate avoided monomer accumulation, and with it a potential runaway, when an inhibitor was added on purpose [8]. Optimal feed profiles for a second-order reaction in a semi-batch reactor under safety constraints followed in 1999 [11]. By 2001 the feed that maximizes conversion was being computed on-line from a heat balance, bounded by the heat release in normal operation and by the temperature the batch could reach after a cooling failure [12]. Online model-based optimization has since been combined with runaway prediction and prevention in fed-batch systems [9].

The fourth case of the app follows an accident. On December 19, 2007, T2 Laboratories in Jacksonville, Florida, was making its 175th batch of a gasoline additive when a runaway reaction burst the reactor, and the explosion and fire killed four employees and injured 32 people [16]. The U.S. Chemical Safety and Hazard Investigation Board found that a loss of sufficient cooling likely caused the runaway, that the cooling system was susceptible to single-point failures, that the relief system could not relieve a runaway, and that the company had not recognized the hazard. Its recommendations went to AIChE and ABET, to add reactive hazard awareness to the requirements for baccalaureate chemical engineering programs [16]. The Board had found that only 11 % of the universities it surveyed required process safety education in the core of the degree [16].

> ⧖ HISTORY **2008: A Feed Started Before the Vessel Was Hot**
>
> In August 2008 a 4500-gallon pressure vessel exploded in West Virginia during the restart of a pesticide unit, after an outage to upgrade the unit's control system, and two employees died. The vessel was meant to be filled with solvent first and held near 135 °C, the temperature at which the incoming material decomposed fast enough not to accumulate, and interlocks kept its feed valve shut until it was hot and its circulation was running. During the restart the fill and the heat-up were skipped and the interlocks were bypassed. Hours after the feed began the vessel was still at 60 to 65 °C, the material piled up, and the runaway overwhelmed the relief system [17].

2.3 The Reactor as Modeled

Reactor R-101 holds 6 m^3. It is charged with 3000 kg of A in solvent, 12 000 mol of A, and dosed with 1200 kg of feed carrying 6.667 mol/kg of B, so that A stands in 50 % excess over the 8000 mol of B. The feed pump P-103 delivers up to 600 kg/h. The jacket loop holds 150 kg of water, circulated at 2 kg/s by P-102, and a split-range pair of valves replaces part of its flow with cooling water at 8 °C or steam-heated water at 140 °C. The arrangement, with every variable that the equations use, appears in Figure 2.1, and Table 2.1 sorts the variables by role.

The state vector has 18 entries. Six carry the process, the moles of A, of unreacted B and of product, the batch and jacket temperatures, and the split-range signal behind its actuator. Twelve are ledgers integrated by the same step, and two of them, the feed mass m_{fed} and the product decomposed n_X, also enter the volume, the heat capacity and the heat-transfer law. The other ten book every joule released by the synthesis and the decomposition, removed into the jacket, lost to the room, brought in by the feed and added by the utilities (with the cooling-water and hot-water shares apart), and every mole of B fed and reacted. Units are SI inside every rate, and the panels show kg/h, kW, kW/K and MJ/K.

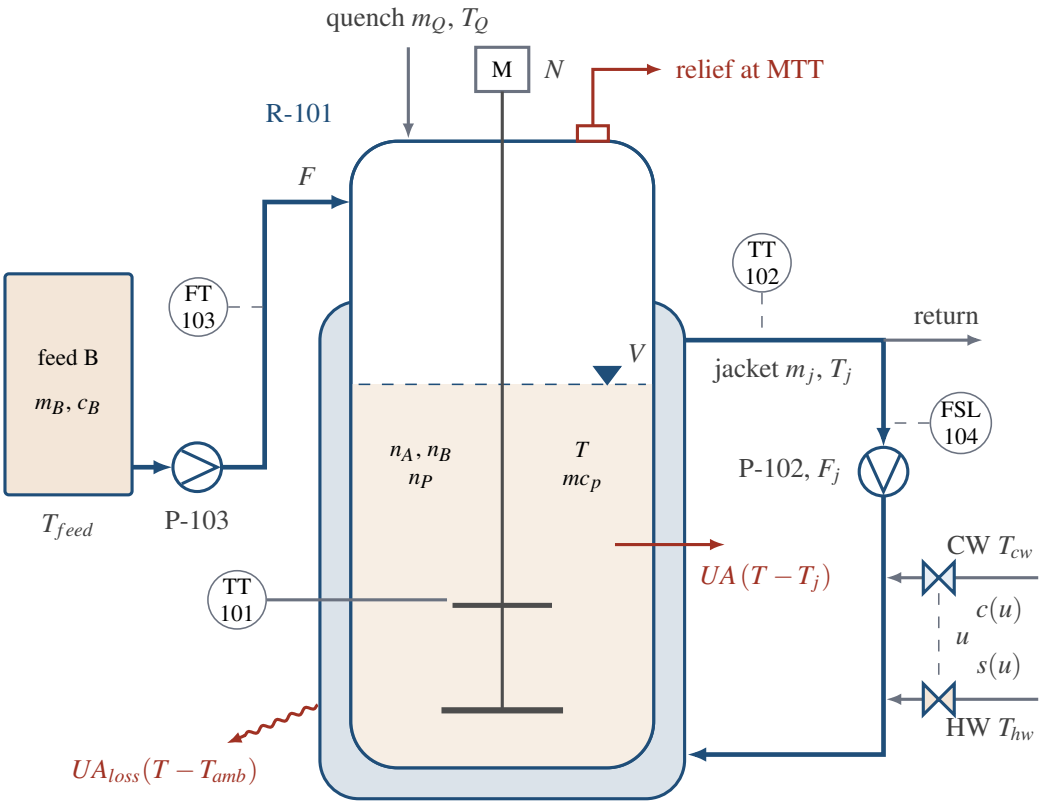


Figure 2.1 Semi-batch reactor R-101 as modeled. Feed of strength c_B at T_{feed} is dosed at F by P-103 through FT-103 into a batch of volume V holding n_A, n_B and n_P at temperature T, stirred at speed N. The jacket, holding m_j of water at T_j, takes $UA\,(T - T_j)$ from the batch and circulates at F_j through P-102, with fractions $c(u)$ and $s(u)$ of its flow replaced by cooling water at T_{cw} and hot water at T_{hw} on the split-range signal u. TT-101 and TT-102 are the batch and jacket thermometers and FSL-104 the low-flow switch of the loop. The rupture disc relieves the batch at MTT, the quench adds m_Q of solvent at T_Q, and the batch loses $UA_{loss}(T - T_{amb})$ to the room.

Table 2.1 Variable Roles, Reset Values and Ranges in Case 2

Variable	Role	Reset value	Range
T	CV, TT-101	70.0 °C SP	30–90
T_j	CV, TT-102	70.2 °C	SP 5–135
MTSR_{est}	CV, AIC-104	SP 80 °C	0–90 K below
$T_{j,SP}$	MV of TIC-101	70.2 °C	5–135
u	MV, split range	50.09 %	0–100
F	MV, feed	0, then 300 kg/h	0–600
T_{cw}	DV, cooling water	8 °C	5–40
a, activity	DV, unmeasured	100 %	25–150
f, fouling	DV, unmeasured	0 %	0–60
feed strength	DV, unmeasured	100 %	80–120
N/N_0	DV, agitator	100 %	20–100
F_j	DV, circulation	2 kg/s	0–100 %
n_A, n_B, n_P	states	12000, 0, 0 mol	—
T, T_j, u	states	70.0, 70.2 °C, 50.09 %	—
m_{fed}, n_X	states, ledgers	0 kg, 0 mol	—
ten ledgers	states	0	—

2.3.1 Kinetics

The synthesis A + B → P is second order, and the product decomposes in a first-order reaction, as shown in Equation (2.1).

$$\begin{aligned} r &= k(T)\,\frac{n_A\,n_B}{V}, \qquad r_d = k_d(T)\,n_P \\ k(T) &= k_{70}\,\frac{a}{100}\,\exp\Big[-\frac{E_r}{R}\Big(\frac{1}{T_K}-\frac{1}{343.15}\Big)\Big] \\ k_d(T) &= k_{d,24}\,\exp\Big[-\frac{E_d}{R}\Big(\frac{1}{T_K}-\frac{1}{T_{D24,K}}\Big)\Big] \end{aligned} \tag{2.1}$$

where r and r_d are in mol/s, T_K is the absolute temperature, $k_{70} = 2.4 \times 10^{-6}\,\mathrm{m^3/(mol\,s)}$, $E_r = 120\,\mathrm{kJ/mol}$, a is the catalyst activity in percent of nominal, $E_d = 300\,\mathrm{kJ/mol}$ and $R = 8.314\,\mathrm{J/(mol\,K)}$. The activation energy of the synthesis makes it 13.5 times slower at 50 °C than at 70 °C and 257 times slower at 30 °C, and 2840 times faster at 150 °C. The decomposition constant is anchored on the declared $T_{D24} = 130\,^\circ\mathrm{C}$ through the zero-order estimate of the adiabatic time to maximum rate, as shown in Equation (2.2).

$$\mathrm{TMR}_{ad}(T) = \frac{mc_p\,R\,T_K^2}{(-\Delta H_d)\,k_d(T)\,n_P\,E_d} \tag{2.2}$$

The constant $k_{d,24}$ is chosen so that $\mathrm{TMR}_{ad}(T_{D24}) = 24\,\mathrm{h}$ for the finished batch, with $mc_p = 7.56\,\mathrm{MJ/K}$ and n_P equal to the 8000 mol of the nominal dose, which gives $k_{d,24} = 1.97 \times 10^{-7}\,\mathrm{/s}$ with $(-\Delta H_d) = 250\,\mathrm{kJ/mol}$. Near T_{D24} the decomposition doubles its rate about every 3.1 K. The finished batch would reach its maximum rate in 222 h at 120 °C, 2.9 h at 140 °C, 23 min at 150 °C and 3.4 min at 160 °C, and at 70 °C the decomposition runs 6.3 million times slower than at T_{D24}. A heat release that grows exponentially with temperature against a removal that grows only linearly is also what sets the ignition boundary of the continuous reactor in Volume 7, Chapter 2.

2.3.2 Mole Balances and the Accumulation

With the feed mass flow F in kg/s and c_B the moles of B it carries per kilogram, the mole balances and the batch's volume and heat capacity follow Equation (2.3).

$$\begin{aligned} \frac{\mathrm{d}n_A}{\mathrm{d}t} &= -r, \qquad \frac{\mathrm{d}n_B}{\mathrm{d}t} = F\,c_B - r, \qquad \frac{\mathrm{d}n_P}{\mathrm{d}t} = r - r_d \\ \frac{\mathrm{d}n_X}{\mathrm{d}t} &= r_d, \qquad \frac{\mathrm{d}m_{fed}}{\mathrm{d}t} = F \\ V &= (m_{A0} + m_Q)/\rho_A + m_{fed}/\rho_B \\ mc_p &= (m_{A0} + m_Q)\,c_p + m_{fed}\,c_{pB} \end{aligned} \tag{2.3}$$

Here c_B is the label value of 6.667 mol/kg, a reagent of 150 g/mol, times the feed strength in percent over 100, $\rho_A = 900\,\mathrm{kg/m^3}$, $\rho_B = 1000\,\mathrm{kg/m^3}$, $c_p = c_{pB} = 1800\,\mathrm{J/(kg\,K)}$, and m_Q is the quench mass, zero until the quench fires. The unreacted B, n_B, is the accumulation, what has been fed less what has reacted. The heat capacity grows from 5.40 MJ/K for the charge to 7.56 MJ/K for the finished batch, and the volume from 3.33 to 4.53 $\mathrm{m^3}$, 56 to 76 % of the vessel.

2.3.3 Energy Balances and the Split Range

The batch and the jacket water exchange heat across the wall, as shown in Equation (2.4).

$$\begin{aligned} m c_p \frac{\mathrm{d}T}{\mathrm{d}t} &= (-\Delta H_r)\, r + (-\Delta H_d)\, r_d - UA\,(T - T_j) \\ &\quad + F\, c_{pB}\,(T_{feed} - T) - UA_{loss}\,(T - T_{amb}) \\ m_j c_{pj} \frac{\mathrm{d}T_j}{\mathrm{d}t} &= F_j c_{pj}\big[c(u)\,(T_{cw} - T_j) + s(u)\,(T_{hw} - T_j)\big] \\ &\quad + UA\,(T - T_j) \end{aligned} \tag{2.4}$$

with $(-\Delta H_r) = 200\,\mathrm{kJ/mol}$, $T_{feed} = 25\,^\circ\mathrm{C}$, $UA_{loss} = 20\,\mathrm{W/K}$ to a room at $T_{amb} = 20\,^\circ\mathrm{C}$, $m_j = 150\,\mathrm{kg}$, $c_{pj} = 4180\,\mathrm{J/(kg\,K)}$ and F_j equal to 2 kg/s times the circulation disturbance, and zero while the circulation has failed. The jacket inlet is the circulating water with a fraction $c(u)$ of it replaced by cooling water or $s(u)$ by hot water, so that at the crossover, with both valves shut, the loop only recirculates. The openings follow the signal as shown in Equation (2.5).

$$\begin{aligned} c(u) &= \mathrm{clamp}\Big(\frac{u_{lo} - u}{u_{lo}}, 0, 1\Big), \qquad u < u_{lo} \\ s(u) &= \mathrm{clamp}\Big(\frac{u - u_{hi}}{100 - u_{hi}}, 0, 1\Big), \qquad u > u_{hi} \\ u_{lo} &= X - \tfrac{1}{2}\Delta u, \qquad u_{hi} = X + \tfrac{1}{2}\Delta u \\ \tau_v \frac{\mathrm{d}u}{\mathrm{d}t} &= u_{cmd} - u \end{aligned} \tag{2.5}$$

where c and s are zero outside the stated ranges, the crossover is $X = 50\,\%$ and the dead band $\Delta u = 0$ as shipped (both adjustable), and $\tau_v = 10\,\mathrm{s}$ is the lag of the valve actuator behind the command u_{cmd}. The jacket inlet temperature $T_j + c\,(T_{cw} - T_j) + s\,(T_{hw} - T_j)$ rises continuously and monotonically with u, but its slope differs on the two sides. With the jacket at 50 °C it is 0.84 K per percent on the cooling side and 1.8 K per percent on the hot side, which one slave tuning has to live with. The pressurizer of Volume 18, Chapter 3 shares one controller output between two unequal actuators in the same way.

The batch enthalpy $m c_p T$, referenced to 0 °C with one heat capacity per kilogram and heats of reaction independent of temperature, closes against the ledgers exactly. Over case 2 to its completion the synthesis releases 1592 MJ, the jacket takes 1500 MJ and the room 16 MJ, the cooling water carries away 1505 MJ while the hot water brings 1.45 MJ, and the energy balance closes to 3.2×10^{-16} relative.

2.3.4 Heat Transfer

The conductance of the wall follows the wetted area and a viscosity factor, as shown in Equation (2.6).

$$UA = U_0\,(1 - \varphi_v x_P)\,(a_0 + a_1 V)\,(N/N_0)^{0.6}\,(1 - f) \tag{2.6}$$

where $U_0 = 500\,\mathrm{W/(m^2\,K)}$, $\varphi_v = 0.30$, $x_P = (n_P + n_X)/n_{B,nom}$ is the product made as a fraction of the 8000 mol dose, $a_0 = 2.6\,\mathrm{m^2}$ is the bottom head, $a_1 = 2.22\,\mathrm{m^2}$ per $\mathrm{m^3}$ of batch, N/N_0 is the agitator speed as a fraction of design and f the fouling fraction. In case 2, UA is 5.00 kW/K at the start of the dose, 4.84 half-way and 4.49 at its end, the area gaining 27 % while U loses 29 %. The law, the agitator exponent and the fouling factor are illustrative. In series with its circulation the jacket's effective conductance, $UA\,F_j c_{pj}/(UA + F_j c_{pj})$, is 3.13 kW/K at the start of the dose, so full cooling could remove 194 kW from a 70 °C batch with water at 8 °C and 125 kW with water at 30 °C.

2.3.5 The Hazard and the Criticality Classes

The accumulation turns into temperature through Equation (2.7).

$$\Delta T_{ad,acc} = \frac{(-\Delta H_r)\, n_B}{m c_p}, \qquad \mathrm{MTSR} = T + \Delta T_{ad,acc} \tag{2.7}$$

MTSR is where the batch would go if the cooling stopped now and all the unreacted B reacted with no heat removed. It rises 26.5 K for every 1000 mol of accumulation in the finished batch, and 37.0 K in the charge, where mc_p is smallest. The class follows from the order of the four temperatures of the cooling-failure scenario [6], as the application computes it. Class 1 has $T_p < \mathrm{MTSR} < \mathrm{MTT} < T_{D24}$, class 2 $T_p < \mathrm{MTSR} < T_{D24} < \mathrm{MTT}$, class 3 $T_p < \mathrm{MTT} < \mathrm{MTSR} < T_{D24}$, class 4 $T_p < \mathrm{MTT} < T_{D24} < \mathrm{MTSR}$, and class 5 has T_{D24} below both MTSR and MTT, where T_p is the process temperature. Here MTT is 170 °C, the boiling point of the solvent at the set pressure of the rupture disc. Because it lies above T_{D24} only classes 2 and 5 can occur, and the batch moves from 2 to 5 the moment MTSR passes 130 °C. T_{D24}, MTT, the MTSR margins and the class thresholds are teaching values. Real ones come from adiabatic calorimetry and DSC [13, 14], assessed by the method of Stoessel [6].

2.3.6 Events, Initialization and Numerics

Discrete events act after each integration step. A cooling failure sets $F_j = 0$ for 10 min or until the circulation is restored, and the water standing in the jacket then warms toward the batch. The quench adds $m_Q = 1000$ kg of solvent at $T_Q = 20$ °C at once, as shown in Equation (2.8).

$$T^{+} = \frac{m c_p T + m_Q c_p T_Q}{m c_p + m_Q c_p} \tag{2.8}$$

It cools by mixing and dilutes, and carries no inhibitor. When T reaches MTT the rupture disc bursts, the batch is lost and the state is frozen, because the model does not describe the venting. The decomposition is flagged as having taken over once $T > T_{D24}$ and its heat $q_d = (-\Delta H_d)\, r_d$ exceeds both the synthesis heat $q_r = (-\Delta H_r)\, r$ and the heat the batch is losing, $q_j + q_{loss}$. The dose is complete when the measured feed total reaches 1200 kg, and the batch is complete once the dose is done, the circulation is running and n_B has fallen below 0.5 % of the nominal dose.

The reset state is the charge at its starting temperature T_0, the setpoint or 30 °C in the challenge, with nothing fed and the jacket holding the charge against its heat loss. That gives $T_j = T_0 + UA_{loss}(T_0 - T_{amb})/UA$, 70.2 °C for a 70 °C charge, with the hot-water valve just cracked at $u = 50.09$ %. A setpoint chosen before Run, directly or carried through Reset, also moves the starting charge, while one changed during a run leaves the charge where it was and the batch is heated or cooled to the new value. The engine of Section 1.6.1 keeps time in minutes for this app and multiplies every rate by 60. The plant advances by fixed fourth-order Runge-Kutta steps of 1 s. In a runaway the reaction time scale falls to seconds, so each step is divided into $n = \lceil \max(\lambda h, |\dot{T}|h/1\,\mathrm{K}, 1) \rceil$ equal substeps, at most 400, with $\lambda = k(n_A + n_B)/V + k_d$ and $h = 1$ s, and n is chosen from the state at the start of the step so that a run replays exactly. The controllers act every 5 s, the historian logs every 2 s, and all of the noise is seeded at 20260926. The speed selector runs 1, 5, 20, 60 or 300 simulated minutes per second, 5 by default, at which the four-hour dose plays out in 48 s. Halving the step moves the cold batch's MTSR at 2 h 30 min by less than 1×10^{-8} K and the relief of case 4 by 0.5 s, and tightening the substep bound from 1 to 0.2 leaves the synthesis peak of case 4 at 157.1398 °C. The energy, jacket and mole ledgers close below 1×10^{-9} relative through a normal batch, a runaway to the relief and a quench, one of the 61 checks the application passes.

Table 2.2 collects the parameters. The disturbances of Table 2.1 are sliders and events of their own, and the scenarios set them at reset or fire them at stated times.

2.3.7 Assumptions

The batch is one well-mixed liquid with one heat capacity per kilogram for every component, heats of reaction that do not change with temperature and a reaction that does not change the volume. Real

Table 2.2 Model Parameters. The Show Model Panel Offers the Visible Ones as Sliders, While the Hidden Ones Stay Out of It

Symbol	Meaning	Value	Class
m_{A0}, n_{A0}	charge of A in solvent	3000 kg, 12 000 mol	visible
m_B	feed to dose	1200 kg	visible
$c_{B,lab}$	B per kilogram of feed	6.667 mol/kg	hidden
c_p	heat capacity of the batch	1800 J/(kg K)	visible
c_{pB}	heat capacity of the feed	1800 J/(kg K)	hidden
ρ_A, ρ_B	densities of charge and feed	900, 1000 kg/m^3	hidden
T_{feed}	feed temperature	25 °C	visible
k_{70}	synthesis rate constant, 70 °C	2.4×10^{-6} m^3/(mol s)	visible
E_r	synthesis activation energy	120 kJ/mol	visible
$-\Delta H_r$	heat of the synthesis	200 kJ/mol	visible
T_{D24}	decomposition threshold	130 °C, teaching value	visible
E_d, $-\Delta H_d$	decomposition	300, 250 kJ/mol	visible
MTT	boiling point at the disc	170 °C, teaching value	visible
V_r	reactor volume	6 m^3	hidden
U_0	clean U at the start	500 W/(m^2 K)	visible
φ_v	fall in U at full conversion	0.30	visible
a_0, a_1	wetted area law	2.6 m^2, 2.22 m^2/m^3	hidden
m_j, F_j	jacket water, circulation	150 kg, 2.0 kg/s	visible
c_{pj}	jacket water heat capacity	4180 J/(kg K)	hidden
T_{hw}	steam-heated water	140 °C	visible
τ_v	split-range actuator lag	10 s	hidden
UA_{loss}, T_{amb}	heat loss, room	20 W/K, 20 °C	hidden
F_{max}	feed pump capacity	600 kg/h	visible
m_Q	quench mass	1000 kg	visible
T_Q	quench temperature	20 °C	hidden

batches change heat capacity and density by a few percent as they convert, and a poorly mixed feed point can hold unreacted B that no well-mixed model shows.

The synthesis is an elementary second-order reaction with an Arrhenius constant chosen for teaching rather than fitted to a chemistry, and the catalyst activity simply scales it. A reaction with an induction period or an autocatalytic step accumulates in ways this model does not reproduce.

The decomposition is first order in the product and anchored through the zero-order estimate for the finished batch. A batch holding less product takes longer to reach its maximum rate at the same temperature, 30 min rather than 21 min at 150.6 °C for the batch of case 4, and the estimate says nothing about autocatalytic decompositions.

The rupture disc ends the model at MTT. Two-phase venting, disc sizing, the pressure history and boiling below MTT are not modeled, because the reactor is treated as closed and rated up to the disc.

The jacket is one well-mixed loop with a constant circulation, and its standing water keeps absorbing heat during a lost circulation without being allowed to boil. The heat-transfer law, the viscosity factor, the agitator exponent and the fouling factor are illustrative, and the sizes, rates and limits were picked to teach one recipe, not to design a plant.

2.4 Where the MTSR Is Lowest

A cold batch holds more B and a hot batch starts higher, and between the two the highest MTSR of a dose has a minimum in the process temperature. Figure 2.2 measures it. Each point is a run of the cold-start configuration with the charge initialized at the setpoint, the fixed 300 kg/h profile from $t = 0$ and the high-temperature feed stop switched off, both set before Run, so that only the temperature changes. The highest MTSR falls from 206.7 °C at 30 °C to a minimum of 75.6 °C at 66 and 67 °C, and rises again to 93.5 °C at 90 °C.

Below the minimum the accumulation dominates, 6679 mol at the end of the dose at 30 °C against 234 mol at 70 °C. Above it the batch temperature dominates, and from 75 °C upward the highest MTSR comes in about the first 8 min of the dose, during the overshoot as the feed starts, rather than

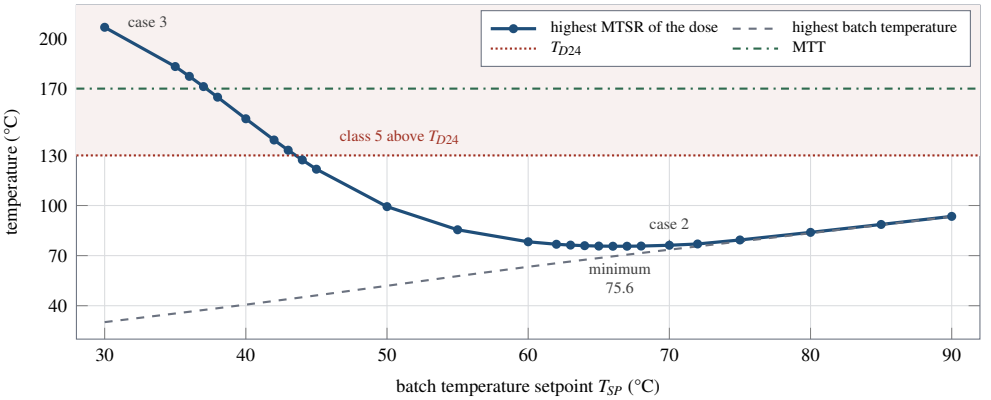


Figure 2.2 Highest MTSR of the fixed four-hour dose against the batch temperature setpoint, with the highest batch temperature of each run. Every point is an engine run of case 3 with the setpoint changed and the high-temperature feed stop off before Run, so that the charge starts at the setpoint. The highest MTSR is the solid line with dots, the highest batch temperature the dashed line, T_{D24} the dotted line and MTT the dash-dotted line. The shaded band lies above T_{D24}, where the batch is in criticality class 5.

at its end. The curve crosses MTT near 37 °C and T_{D24} near 43.5 °C, so every setpoint of 43 °C or less ends its dose in class 5, while the shipped 70 °C sits 3 to 4 K above the minimum with a highest MTSR 0.6 K above it.

A quasi-steady argument places the minimum. When the reaction keeps up with the feed, the unreacted B is roughly the molar feed divided by the pseudo-first-order rate constant, as shown in Equation (2.9).

$$n_B \approx \frac{F c_B V}{k(T)\, n_A}, \qquad \mathrm{MTSR} \approx T + \frac{(-\Delta H_r)\, F c_B V}{m c_p\, k(T)\, n_A} \tag{2.9}$$

With the Arrhenius law the derivative of this MTSR with temperature vanishes where $\Delta T_{ad,acc} = RT_K^2/E_r$, which is 8.0 K near 67 °C. At the end of the dose, where the highest MTSR occurs, the estimate of n_B is within 6 to 13 % of the engine's between 62 and 70 °C, 247 mol against 234 at 70 °C. The engine's own $\Delta T_{ad,acc}$ passes 8.0 K between 67 °C, where it is 8.6 K, and 68 °C, where it is 7.7 K. The measured curve is flat within 0.11 K from 65 to 68 °C, which is why the minimum is best read as a band. Running below it trades a few kelvin of process temperature for a large accumulation.

2.5 Implemented Control System

Three PI loops, a heat-flow calorimeter, a dosing controller with four shipped strategies and an optional fifth, and a layer of interlocks run every $\Delta t_c = 5$ s, as shown in Figure 2.3. Within a physics step they act in the order of Figure 1.10, and the plant's own events (the failure, the quench, the relief) follow the integration.

2.5.1 Transmitters

No controller reads the plant's amounts. TT-101 lags the true batch temperature by a 15 s first-order lag and adds first-order colored noise with a standard deviation of 0.05 K and a 30 s correlation time. TT-102 lags the jacket by 6 s with 0.05 K of noise correlated over 18 s. The Coriolis meter FT-103 lags the feed by 3 s with 0.3 % of multiplicative noise correlated over 12 s, and its integral is the measured total that ends the dose. FSL-104 trips when the true circulation falls below 20 % of design.

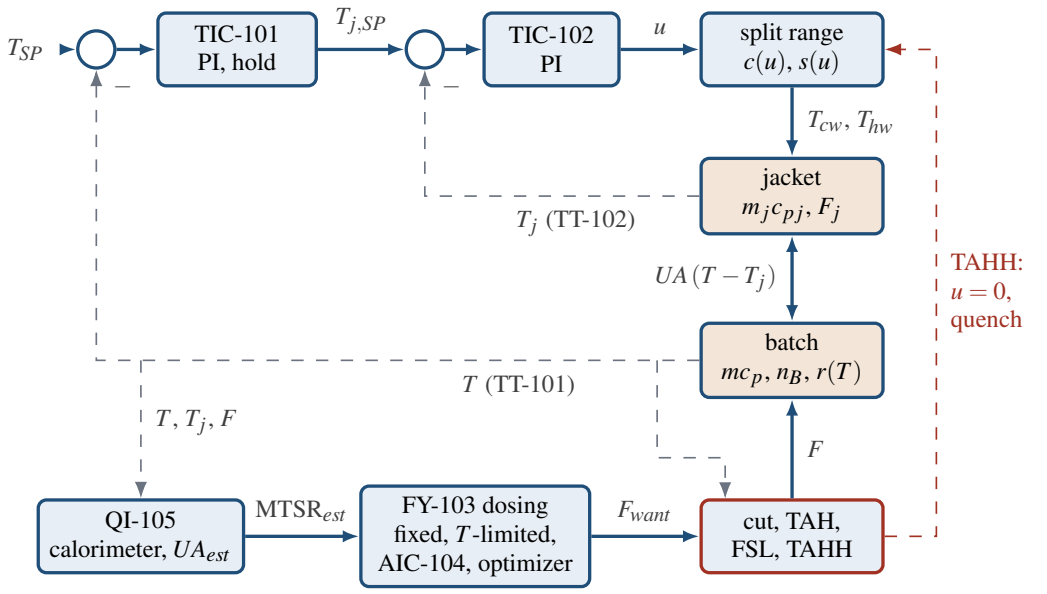


Figure 2.3 Control structure as implemented. TIC-101 sets the jacket setpoint and TIC-102 the split-range signal. QI-105 estimates MTSR from TT-101, TT-102, FT-103 and an assumed UA_{est}, FY-103 turns it into the wanted feed under the chosen strategy, and the temperature cut and the interlocks limit or stop the feed, the high-high trip also commanding full cooling and the quench.

Table 2.3 Loop Pairings, Tunings and Limits as Shipped

Tag	CV	MV	γ	K_c	τ_I, output
TIC-101	T (TT-101)	$T_{j,SP}$	+1	4	20 min, 5–135 °C
TIC-102	T_j (TT-102)	u	+1	3	2 min, 0–100 %
AIC-104	MTSR_{est}	F	+1	40	15 min, 0–F_{ceil}

Limit $L = 80\,°\mathrm{C}$, $F_{ceil} = 600\,\mathrm{kg/h}$, step $h = 5\,\mathrm{min}$
Profile 300 kg/h, cut 2 K with 0.3 K of hysteresis
TAH 80 °C (clears at 78), FSL 20 %, TAHH 115 °C

Gains in K per K, % per K and kg/h per K, in the order of the rows.

2.5.2 The Shared PI Law

All three loops run the framework's velocity-form PI law, Equations (1.14) to (1.17) of Section 1.6.2, with $\Delta t = \Delta t_c$. The action, written γ in Table 2.3, is +1 for all three loops and the bias is zero, with the integral seeded at reset to the actuator's reset value. The derivative term is unused in all three loops, which leaves the running sum equal to the proportional term. Windup is handled on the integral alone, by back-calculation whose tracking time is τ_I, the clamp limits only what is emitted, and none of the loops is rate limited. A loop in MANUAL tracks the operator's value and returns to automatic without a bump. The pairings and shipped settings are listed in Table 2.3.

2.5.3 The Temperature Cascade

TIC-101 turns the batch temperature error into the jacket setpoint, limited to 5 to 135 °C, and TIC-102 moves the split-range signal between 0 and 100 %. Both act directly, so a colder batch raises the jacket setpoint and a colder jacket raises the signal toward the hot water. With the integral hold on, TIC-101 runs with $\tau_I = \infty$ while the signal sits at 0 or 100 % or FSL-104 reports lost flow, so its integral neither

grows nor unwinds while the jacket cannot answer, a form of conditional integration [18]. In Local the jacket loop holds its own setpoint and TIC-101 tracks.

The application identifies no control model, and a step from the case 2 reset with the jacket loop closed in Local and the dose start moved to 120 min measures the two plants the cascade sees. Stepped from its reset value of 70.2 to 65 °C, the jacket reaches 63 % of the change 24 s after the new setpoint reaches TIC-102, with a 0.83 K undershoot. The batch follows the jacket with a gain of 0.995 and a 63 % time of 17.9 min, which is $mc_p/UA = 18.0$ min at the start of the dose and grows to 28.1 min at its end. The inner loop is about 45 times faster than the plant the outer loop sees, which is what makes the cascade worth having [18]. Volume 3, Chapter 3 measures how fast an inner loop must be on a fermenter whose lags lie within a factor of two of one another.

2.5.4 The Heat-Flow Calorimeter

QI-105 infers the reaction heat from the batch's heat balance using only TT-101, TT-102, FT-103 and a conductance it has to assume, as shown in Equation (2.10).

$$\begin{aligned} q_{est} &= UA_{est}(T_m - T_{j,m}) + (mc_p)_{est}\,\dot{T}_f \\ &\quad - F_m c_{pB}(T_{feed} - T_m) + UA_{loss}(T_m - T_{amb}) \\ n_{B,est} &= c_{B,lab}\,m_{fed,m} - \textstyle\sum q_{est}\,\Delta t_c/(-\Delta H_r) \\ \mathrm{MTSR}_{est} &= T_m + (-\Delta H_r)\max(0, n_{B,est})/(mc_p)_{est} \end{aligned} \tag{2.10}$$

where the subscript m marks a measurement, $\dot{T}_f$ is the slope of TT-101 through a first-order filter of 1 min, $m_{fed,m}$ is the measured feed total and $(mc_p)_{est}$ is built from it. The assumed conductance UA_{est} is the clean-jacket law of Equation (2.6) without fouling or agitator terms, evaluated at the measured volume and the calorimeter's own conversion and multiplied by a correction that is 100 % as shipped. A True UA option, which no real calorimeter has, hands it the plant's conductance for comparison. Every term but UA_{est} is measured, every constant is the declared value of Table 2.2, and a wrong conductance drifts the estimate by $n_{B,est} - n_B \approx -\int (UA_{est} - UA)(T - T_j)\,dt/(-\Delta H_r)$. A jacket that has fouled since the calibration removes less heat than the calorimeter believes, so it credits the reaction with heat it did not make and reports an accumulation that is too low, the failure in the safe-looking direction. In case 2, with nothing fouled, the estimate stays within 21.1 mol of the truth and its MTSR within 0.63 K.

2.5.5 Dosing Strategies

FY-103 sets the feed. The fixed profile spreads the dose evenly, $F_{prof} = m_B/t_{dose}$, which is 300 kg/h for four hours, from the dose start until the measured total reaches 1200 kg. Temperature-limited dosing runs the same profile but cuts it while TT-101 is more than $\delta_T = 2$ K above the setpoint, resuming 0.3 K below that. Accumulation-limited dosing gives the feed to AIC-104, a PI loop on MTSR_{est} with its setpoint the limit $L = T_{D24} - \Delta_{MTSR}$, 80 °C for the shipped margin of 50 K, and it tracks the applied feed whenever it is not in charge. The limit on the estimate is a path constraint in the sense of Section 1.4.4, held below T_{D24} with a back-off, and like the explosive-limit override of Volume 4, Chapter 6 it lets a safety limit rather than the recipe set the throughput. Here the limited quantity is not measured at all and comes from an estimate. The one-step optimizer takes the largest feed whose predicted MTSR one step h ahead meets the same limit, with the reaction rate held at its calorimetric value $r_{est} = \max(0, q_{est})/(-\Delta H_r)$, as shown in Equation (2.11).

$$\begin{aligned} F^* = \max F \quad &\text{subject to} \quad 0 \le F \le F_{ceil} \text{ and} \\ T_m + \frac{(-\Delta H_r)\left[n_{B,est} + (F c_{B,lab} - r_{est})\,h\right]}{(mc_p)_{est} + F c_{pB} h} &\le L \end{aligned} \tag{2.11}$$

The constraint is linear in F, so the optimum is closed form. With $\ell = L - T_m$, $\alpha = h\,[(-\Delta H_r)c_{B,lab} - \ell\, c_{pB}]$ and $\beta = \ell\,(mc_p)_{est} - (-\Delta H_r)(\max(0, n_{B,est}) - r_{est}h)$, the feed is $F^* = \mathrm{clamp}(\beta/\alpha, 0, F_{ceil})$

when $\alpha > 0$, and otherwise F_{ceil} if $\beta \geq 0$ and zero if not. Holding the rate fixed is conservative for a second-order reaction, which speeds up as B accumulates. The optimizer is a one-step constrained maximization and not a model predictive controller. Both limited strategies keep the temperature cut behind them, every strategy delivers the last of the dose exactly, and the application's own test confirms that the optimizer's feed meets the limit with equality and that 1 % more breaks it.

2.5.6 Interlocks

Above the loops, and in MANUAL as well as in automatic, TAH stops the feed while TT-101 reads above 80 °C and releases it 2 K below. FSL-104 stops the feed while the circulation is lost. TAHH at 115 °C is latched until the interlocks are reset, stopping the feed for good, driving the signal to full cooling and the jacket setpoint to its 5 °C floor, and firing the quench when it is armed. The interlocks stop the hazard from growing. None of them removes the B already in the batch.

2.5.7 Simulation Model, Loop Plants and Controller Models

The simulation model is the one of Section 2.3. The plants the loops see are the measured lags above, the 24 s inner loop and the 18 min to 28 min batch, and no shipped controller uses an identified model. The shipped controllers' internal models are the clean-jacket conductance and the heat balance of Equation (2.10), the adiabatic MTSR of Equation (2.7), which ignores the jacket water and the room and is conservative for that reason, and the one-step prediction of Equation (2.11). None of them contains the kinetics, which the predictive option of Section 2.5.8 carries in a model of its own.

2.5.8 Model Predictive Control

A fifth dosing strategy, model predictive control (MPC), is an option and never the default of any case, and only scenarios 9 and 10 load it. Every $T_s = 30\,\mathrm{s}$, six scans of Δt_c, it plans the feed over $N = 10$ steps (5 min) in seven blocks whose lengths are 1, 1, 1, 1, 2, 2 and 2 steps. As in Section 1.6.3, only the first block's feed is applied, here held for 30 s, and the initial plan is computed on the priming call at $t = 0$ [19, 20]. Its request replaces the wanted feed of FY-103, so the temperature cut, TAH, FSL, TAHH and the quench act on it as on the other strategies. Nonlinear predictive control has been applied to a semi-batch reactor to avoid thermal runaway [21], while this controller solves a quadratic program on a model relinearized at every decision.

The model is written from the balances of Section 2.3 with the declared values of Table 2.2 and the design disturbances (water at 8 °C, full activity and circulation), never the run's values, so a weak catalyst, a fouled jacket or warm water is mismatch it is not told about. Its six states are the unreacted B and A, the batch and jacket temperatures, the integral I of TIC-101 and the feed total, as shown in Equation (2.12).

$$\begin{aligned}
\frac{\mathrm{d}n_B}{\mathrm{d}t} &= F c_{B,lab} - \hat{r}, \quad \frac{\mathrm{d}n_A}{\mathrm{d}t} = -\hat{r}, \quad \frac{\mathrm{d}m_{fed}}{\mathrm{d}t} = F \\
mc_p \frac{\mathrm{d}T}{\mathrm{d}t} &= (-\Delta H_r)\,\hat{r} - UA_{est}(T - T_j) + F c_{pB}(T_{feed} - T) \\
&\quad - UA_{loss}(T - T_{amb}) + mc_p\,\hat{d}_T \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{K_{c,T}}{\tau_{I,T}}(T_{SP} - T) \\
\frac{\mathrm{d}T_j}{\mathrm{d}t} &= \mathrm{clamp}\big((T_{j,SP} - T_j)/\tau_s,\; g_c,\; g_h\big)
\end{aligned} \tag{2.12}$$

Here $\hat{r} = \hat{\theta}\,k(T)\,n_A n_B / V$ at full activity, UA_{est} is the calorimeter's clean-jacket law, $T_{j,SP} = K_{c,T}(T_{SP} - T) + I$ clamped to 5 to 135 °C with the tuning of TIC-101 ($K_{c,T} = 4$, $\tau_{I,T} = 20\,\mathrm{min}$, I held at a jacket limit when the hold is on), and g_c and g_h are the rates of T_j at full cooling and full hot water, $[UA_{est}(T - T_j) + F_j c_{pj}(T_x - T_j)]/(m_j c_{pj})$ with T_x equal to $\hat{T}_{cw}$ or T_{hw}. The lag $\tau_s = 22\,\mathrm{s}$ stands for the closed jacket loop. Steps of 10 K up and down in its setpoint, on case 2 in Local with no feed, fit a first-order-plus-dead-time model with a gain of 1.02, a dead time of 0.16 min and a lag of 0.17 to

0.20 min, with residuals of 2.8 to 3.3 % of the step, and τ_s lumps the two. With water at 30 °C and the jacket near 45 °C, as in scenario 9, the true loop takes 91 s to reach 63 % of a step, because the cooling valve's authority falls with $T_j - T_{cw}$, an error that matters only while the valve is in its range.

With its estimates settled and the feed stepped to a new constant rate, the model's batch temperature stays within 0.54 K of the truth over 60 min in six open-loop checks. On a jacket fouled by 15 % it stays within 0.15 K while the predicted MTSR runs up to 6.94 K low, because the calorimeter's error passes into the estimates. Replaying the feed the plant received in scenario 9 is harder. From 2 min, with $\hat{T}_{cw}$ still at 8 °C, the error within 5 min reaches 0.84 K on the MPC's run and 1.29 K on the optimizer's, from 10 to 180 min it is at most 0.23 and 0.30 K, and over the 30 min from 200 min, across the end of the dose, it reaches 1.47 K.

At each decision the nonlinear model is integrated with no feed over the horizon by fourth-order Runge-Kutta, its Jacobian at the measured point (equal to a hand derivation within 3×10^{-9}) gives the feed's effect, the pair is discretized exactly over T_s, and an affine term per step makes the linear free response equal the nonlinear one. Over the 436 decisions of scenario 9 the planned temperature differs from the nonlinear model under the same plan by 0.42 K on average and 1.45 K at most. In the 210 decisions at full cooling the plan over-predicts the peak of its own feed by 0.50 K on average and 1.22 K at most, the safe direction, while the first plan, with the valve at the crossover, under-predicts it by 1.45 K.

Four estimates make the prediction offset free through an augmented disturbance model [22], all from instruments the calorimeter already reads, and all updated on every scan under every strategy. The multiplier $\hat{\theta}$ is the calorimeter's reaction rate over the model's, each filtered over 10 min, so the weak catalyst of case 7 appears as $\hat{\theta} = 0.25$. The water temperature $\hat{T}_{cw}$ comes from the jacket's heat balance while the commanded cooling valve is more than 15 % open, filtered over 5 min. At each decision $\hat{d}_T$ moves by 0.3 of the 30 s prediction residual of T, within ±7.2 K/h, and the bias b_M, the calorimeter's MTSR less the model's, makes the limit mean what it means to AIC-104. No Kalman filter is used. Each plan solves Equation (2.13) for the block feeds F_j in units of 100 kg/h.

$$
\begin{aligned}
\min \; & \sum_{k=0}^{N-1} -w_k F_k + S \sum_{j=0}^{6} (F_j - F_{j-1})^2 \\
& + \textstyle\sum_{i=M,T} \left(\rho\, \varepsilon_i^2 + \sigma \varepsilon_i \right) \\
\text{s.t. } & 0 \le F_k \le \min(F_{ceil}, F_{max}) \\
& T_s \textstyle\sum_k F_k \le m_B - m_{fed,m} \\
& \widehat{\mathrm{MTSR}}_k \le \max\left(L,\, \widehat{\mathrm{MTSR}}_k^0\right) + \varepsilon_M, \quad \varepsilon_M \ge 0 \\
& \hat{T}_k \le \max\left(T_{SP} + \delta_T - \beta,\, \hat{T}_k^0\right) + \varepsilon_T, \quad \varepsilon_T \ge 0
\end{aligned}
\tag{2.13}
$$

where F_k is the feed of the block holding step k (in kg/s in the dose row), F_{-1} the feed applied now, $w_k = 0.01\,(1 - k/2N)$ a reward that favors the near steps, $S = 0.001$ the move weight, $\hat{T}_k$ and $\widehat{\mathrm{MTSR}}_k$ the predictions over steps 1 to N (the latter linearized at the measured point, plus b_M), and the superscript 0 the free response with no feed. The ceiling sits $\beta = 0.25$ K below the cut, at 71.75 °C, and $L = 80$ °C. The feed bounds and the rest of the dose are hard, and the two bands are soft at $\rho = 10^4$ per K^2 and $\sigma = 10^3$ per K, with one slack per band for the whole horizon. The larger of the limit and the free response makes each band a funnel, so a slack counts only what the feed adds, and both were zero at every decision of every case. A flat reward spread the rest of the dose over the horizon and never finished it, hence the fall of w_k to half. Four plans of scenario 9 re-solved by an independent ADMM method agree to 3.2×10^{-6}, and a plan of nine variables takes 0.10 to 0.15 ms on average in Node.js.

The settings were chosen once on scenario 9, over 60 combinations of horizons from 5 to 60 min, back-offs from 0 to 1 K and move weights from 0.001 to 0.1, by a rule fixed beforehand: no cut, no TAH, true $T \le 72$ °C and true MTSR ≤ 81 °C, then the earliest end of the dose, ties within 1 % going

to the smaller feed travel. No other case is retuned. The end of the batch, 200 to 430 min away, lies beyond the solver's 120 steps, and within that limit the rule chose 5 min because every 30 and 60 min candidate was cut or ended its dose later, so the end of the dose enters through the hard row on the rest of the dose rather than a shrinking horizon.

A plan is accepted only when its solver status is optimal, all of its outputs are finite, its slacks are not negative and its moves lie within bounds. Otherwise the run is marked and AIC-104, whose integral has tracked the applied feed, doses until the next valid plan. Forced at five instants with the noise off, the hand-over moves the feed by at most 3.02 kg/h, the size of AIC-104's own next move (3.16 kg/h). The return is not bumpless (4 to 116 kg/h) but lies within the plan's ordinary changes between decisions, 83 kg/h on average and 412 kg/h at most. The controller reads only TT-101, TT-102, FT-103 and its total, the calorimeter, its own commands and the declared parameters, and the application's tests catch each of 16 doctored copies that read the plant or skip a protection under the MPC alone.

2.5.9 Linearization Along the Plan

The MPC has a second setting, its linearization, single in every case but case 10. Linearizing once per decision at the current state is the successive linearization of nonlinear MPC [23], and it holds the feed's effect at the measured point for the whole 5 min, although the plan's own feed moves the batch away from that point. The synthesis speeds up by 13.0 % per kelvin at 70 °C, and the jacket stops following its setpoint once the cooling valve is fully open. The iterated setting keeps the single plan as iteration 1 and relinearizes the same model along the path that plan predicts. With Equation (2.12) written as $\dot{x} = f(x,F)$, $x = (n_B, T, T_j, I, m_{fed}, n_A)$, and $\bar{F}_k$ the previous iteration's plan, each step is linearized at its own predicted state and planned feed, as shown in Equation (2.14).

$$\begin{aligned} \bar{x}_0 &= x_m, \qquad \bar{x}_{k+1} = \Phi(\bar{x}_k, \bar{F}_k) \\ A_k &= e^{J_k T_s}, \qquad B_k = \textstyle\int_0^{T_s} e^{J_k \tau}\, d\tau\; g_k \\ c_k &= \bar{x}_{k+1} - A_k \bar{x}_k - B_k \bar{F}_k \\ \hat{x}_{k+1} &= A_k \hat{x}_k + B_k F_k + c_k, \qquad \hat{x}_0 = x_m \\ \widehat{\mathrm{MTSR}}_k &= \nabla \mathrm{MTSR}(\bar{x}_k)\,(\hat{x}_k - \bar{x}_k) + \mathrm{MTSR}(\bar{x}_k) + b_M \end{aligned} \tag{2.14}$$

Here x_m is the state the decision starts from, Φ the model over T_s at constant feed (two Runge-Kutta substeps), J_k and g_k the derivatives of f in x and F at $(\bar{x}_k, \bar{F}_k)$ by the central differences used at the measured point, and $\hat{T}_k$ the second entry of $\hat{x}_k$, for $k = 1$ to N. The affine term c_k makes the model reproduce the plan's own path, so a plan that stops moving is a fixed point and a converged plan predicts what the nonlinear model gives for it. The weights, bounds, dose row, soft prices and funnels of Equation (2.13) are iteration 1's.

With A_k and B_k changing along the horizon, the condensed program carries products $A_{k-1} \cdots A_{j+1} B_j$ where the shared solver, written for one model, carries matrix powers. The application states it on the same dense active-set QP solver, and given one model at every step it returns the shared solver's plan bit for bit at all 351 decisions of case 10, while an independent log-barrier solve agrees within 0.0036 kg/h. The iteration stops when no planned feed moves by 1 kg/h, below the noise of FT-103, or after three relinearizations, both values declared before any run. A program that fails the acceptance test, or a plan still moving after the third relinearization, falls back to iteration 1's plan, with a mark in the historian. Each relinearization is a Gauss-Newton step on the plan, the step that real-time iteration takes once per sample [24], repeated here until the plan settles, between linear and fully converged nonlinear MPC [25]. A plan of case 10 perturbed by up to 100 kg/h per step is predicted within 3.2×10^{-4} K of the nonlinear model at the median decision, and within 3.2×10^{-5} K at 10 kg/h, while the single linearization misses by 0.51 K at every size because it does not reproduce the plan's own path. The converged plans of case 10's dose match the nonlinear model within 6.2×10^{-5} K, where the single plans at the same 324 decisions err by up to 1.66 K.

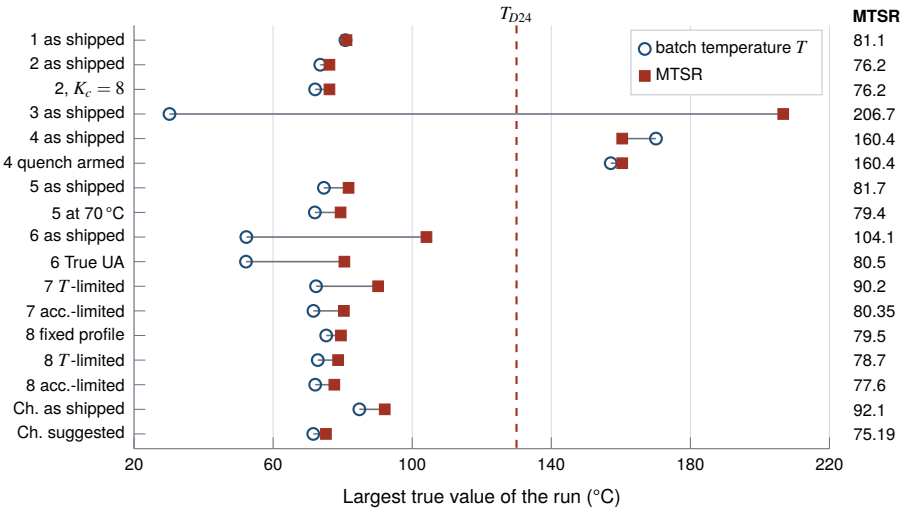


Figure 2.4 The largest true batch temperature (open circles) and the largest true MTSR (filled squares) of every run of Table 2.5, with T_{D24} dashed and the MTSR values at the right. The line between the two marks is the heat the batch stores as unreacted B. Notice case 3, whose batch never passes 30.2 °C while its MTSR reaches 206.7 °C, and case 4, whose temperature passes its MTSR only because the run continues to the relief (footnote a of the table).

While case 10 doses, 324 of its 331 plans converge, in 3.58 programs on average and at most four. Before the pump starts 12 of 19 do not, moving by about 295, then 9 to 17, then 2 to 3.4 kg/h. A decision takes 0.57 ms in Node.js against 0.12 ms for the single plan. In Chrome on the page's frame loop, under the load of other jobs on the host, it takes 1.3 to 1.5 ms at 1× speed, 0.3 % of a core, and at 300 simulated minutes per second, about ten decisions to a frame, the MPC takes 36 to 37 % of a core while the page holds 60 frames a second. A switch between the settings, or a fallback, moves the feed only at a decision and is not ramped. With the noise off, a fallback or a switch to the single plan steps the feed by up to 248 kg/h and the return by up to 255 kg/h, against ordinary moves of up to 279 kg/h, and over the next 10 min the batch is at most 0.10 K warmer than without it, with no cut.

2.6 Measured Cases

The scenario sequence of Table 2.4 runs through ten cases to a graded challenge. Every case starts from the shipped settings of Tables 2.3 and 2.2 and departs from them only as its row states. A departure is made at $t = 0$ before Run, and the application gives the same run when it is carried through Reset. The engine itself, run at its default seed, supplied each value quoted here, and times are clock times from the start of the batch. Table 2.5 gathers the largest true values of each run, and Figure 2.4 sets the largest MTSR of each run beside its largest batch temperature.

2.6.1 A Batch Whose Heat Follows the Pump

Case 1 is run in MANUAL, the pump at 300 kg/h and the split-range signal left at 50 % with both utility valves shut. The reaction makes 122 kW and no utility removes it, so the batch reaches 73.7 °C at 5 min and 79.5 °C at 10 min, and the TAH stops the feed at 10.7 min, when TT-101 reads 80.1 °C. The heat follows the pump because at 70 °C, with A in excess, the reaction takes B almost as fast as it arrives. After 10 min, 331 mol have been fed and 24 mol are unreacted, and stopping the pump then drops the reaction heat from 122 to 4.9 kW by 12 min. The *Hold the temperature* helper, pressed at 5 min, sets the signal to 33.9 %, the cooling valve 32 % open, and the batch peaks at 75.1 °C and is back at 69.2 °C by 60 min.

Table 2.4 Scenario Sequence of the Application

#	Scenario and configuration	Teaching point
1	MANUAL, 300 kg/h, valve at 50 %	the heat follows the pump
2	Cascade, fixed profile from 10 min	a feed-controlled batch
3	Charged and dosed at 30 °C	accumulation, no symptom
4	Case 3, circulation lost 2:30 to 5:00	an honest runaway
5	Case 4, accumulation-limited	safe, and slow
6	50 °C, AIC-104, U fouled 15 %	a fooled calorimeter
7	Temperature-limited, activity 25 %	no temperature symptom
8	Water at 30 °C, fixed 3 h profile	the jacket sets the pace
9	Case 8 dosed by the MPC	the cut planned ahead
10	Case 2 dosed by the MPC, iterated	a plan exact to its model
Ch.	Cold charge, two utility upsets	graded challenge

Table 2.5 Largest True Accumulation, MTSR and Temperature of Each Run, With the End of the Dose and the Outcome. Runs Stop at 8 h, at Completion or at the Relief, and Last 1 h in Case 1 and 7 h in the Challenge

Case and configuration	n_B (mol)	MTSR (°C)	T (°C)	Dose ends	Outcome
1 as shipped	49	81.1	80.8	—	TAH, 10.7 min
2 as shipped	234	76.2	73.56	4:10	done 4:26
2 with $K_c = 8$	234	76.2	72.08	4:10	done 4:25
3 as shipped	6679	206.7	30.2	4:00	class 5 at 1:44
4 as shipped	4398	160.4[a]	170.1	—	lost 3:55
4 quench armed	4398	160.4	157.1	—	quenched 3:10
5 as shipped	1641	81.7	74.6	—	563 kg by 8 h
5 at 70 °C[b]	324	79.4	72.0	6:47	done 7:08
6 as shipped	2056	104.1	52.3	3:42	limit +24.1 K
6 True UA	1145	80.5	52.2	5:54	limit +0.5 K
7 T-limited	765	90.2	72.35	4:09	done 5:47
7 acc.-limited	391	80.35	71.6	6:00	done 7:13
8 fixed profile	181	79.5	75.27	3:00	done 3:08
8 T-limited	275	78.7	72.84	3:34	done 3:50
8 acc.-limited	235	77.6	72.1	4:45	done 5:00
Ch. as shipped	1196	92.1	84.8	4:15	worst 22.76 K
Ch. suggested[c]	198	75.19	71.55	3:43	worst 1.61 K

n_B, MTSR and T are the largest true values of the run, and times are h:min. [a] At the failure. Once no B is left the running maximum is T itself, 170.1. [b] Setpoint raised before Run, so the charge starts at 70 °C. [c] Optimizer, 55 K margin, 450 kg/h ceiling, dose from 20 min, batch loop $K_c = 6$ and $\tau_I = 15$ min, integral hold on.

In case 2 the cascade holds 70 °C and the pump starts at 10 min. Before the feed the jacket sits 0.2 K above the batch and the noise keeps the signal between 49.3 and 51.0 % around the crossover. The signal crosses 50 % for good at 10.3 min and settles near 35 %, between 33 and 37 % with the noise, the cooling valve 26 to 33 % open. The feed is a load step of about 110 kW that the PI cascade can correct only after it has warmed the batch, so the batch peaks at 73.56 °C at 19.2 min and returns within 0.5 K of its setpoint for the rest of the dose at 64 min. With the batch loop gain doubled to 8 the peak is 72.08 °C at 16.4 min and the return comes at 47.6 min. From 70 min to the end of the dose the jacket runs 20.5 to 22.0 K below the batch and removes 96.5 to 105 kW. The unreacted B never exceeds 234 mol, 2.9 % of the dose, reached as the dose ends at 4 h 10 min, MTSR never passes

76.2 °C, and the batch is complete at 4 h 26 min. At 70 °C this batch is feed-controlled.

2.6.2 Accumulation Without a Symptom

Case 3 charges the batch at 30 °C and doses it on the same profile without heating it. During the dose the true temperature stays between 30.0 and 30.2 °C, TT-101 reads between 29.9 and 30.3 °C with its noise, the signal stays between 38.7 and 50.6 %, and the cooling valve never opens more than 22.5 %, because the reaction never makes more than 29.5 kW. At 1 h 44 min MTSR passes T_{D24} with 91 % of the B fed so far unreacted, and the class moves from 2 to 5 at the same moment (the calorimeter's class 13 s later). At 2 h 30 min, 4398 mol are unreacted, 55 % of the dose, with MTSR at 160.4 °C, and at 4 h 6677 mol, 83 %, with MTSR at 206.7 °C, above MTT. Nothing on the temperature trend changes, which is the point of the case and the left end of Figure 2.2.

> **TRY IT IN THE APP** **3 · Cold Start**
>
> Load the scenario and press Run. Watch the trends: the batch temperature stays flat near 30 °C while the true MTSR climbs. At 1 h 44 min it passes the 130 °C decomposition threshold and the criticality class steps from 2 to 5, with no change on the temperature trend. By 4 h the MTSR stands at 206.7 °C.

2.6.3 An Honest Runaway

In case 4 the jacket circulation fails at 2 h 30 min on the cold batch of case 3 and returns at 5 h, with the quench not armed. FSL-104 stops the feed at the scan that coincides with the failure, with 4398 mol unreacted, the most of the run, and MTSR at 160.4 °C. That stops the hazard from growing and does nothing about the B already there. For 35 min the batch creeps from 30.1 to 45.4 °C, warming the water standing in the jacket with it, until 3 h 05 min. The synthesis then accelerates, to 57.8 °C at 3 h 09 min and 75.0 °C at 3 h 10 min, and in the next 13 s it runs to 157.1 °C, gaining 37.6 K in the last one-second step and releasing 250 MW at the start of that step. TT-101, behind its 15 s lag, reads 78.6 °C at the true peak and trips the TAH at 100.9 °C and the TAHH at 116.7 °C, 2 s and 7 s after it, commanding full cooling that cannot flow. The synthesis peak lies 3.2 K below the MTSR at the failure because the standing water and the room took some heat during the creep, which is why MTSR is the conservative number. The jacket water, then at 59.9 °C, absorbs heat from the batch, which falls to 150.56 °C at 3 h 17 min, 20.6 K above T_{D24}, where the decomposition's 18.8 kW exceeds the 18.7 kW the batch loses and its time to maximum rate for this batch is about 30 min. The batch then heats itself, to 154.8 °C at 3 h 40 min and 159.9 °C at 3 h 50 min, and reaches MTT at 3 h 55 min with the decomposition making 699 kW (between 3 h 54 min and 3 h 55 min on eight other seeds). The rupture disc bursts and the batch is lost 65 min before the cooling would have returned. The run at 230 min is shown in Figure 2.5. Armed, the quench fires at the same high-high trip and brings the batch from 156.7 °C to 127.9 °C, below T_{D24}, and the decomposition never takes over. The reactor survives and the batch is scrapped.

The screenshot also shows the calorimeter's limits. During the jump its filtered slope and its integral lag the truth, so its MTSR spikes to 214.4 °C at 190.6 min while no B is left. Afterwards it credits the decomposition's heat to the synthesis, and its estimate of the unreacted B has fallen to –387 mol by 230 min, which the badges show as zero.

Pause. Try Exercise 2.1 before reading on.

2.6.4 The Same Failure, Accumulation-Limited

Case 5 repeats case 4 with AIC-104 holding the calorimeter's MTSR at 80 °C. At 30 °C the limit lets the pump run only as fast as the slow reaction takes the B, so by 2 h 30 min it has fed 276 kg against 750 kg on the fixed profile, with 1476 mol unreacted, MTSR at 80.0 °C and the pump at 28 kg/h. After the failure the stored B reacts slowly and the batch peaks at 74.6 °C at 4 h 25 min, class 2 throughout and 55 K below T_{D24}. When the cooling returns at 5 h the dose resumes, and by 8 h the pump runs at 23 kg/h with 563 kg fed, the true MTSR having peaked at 81.7 °C as AIC-104 brought

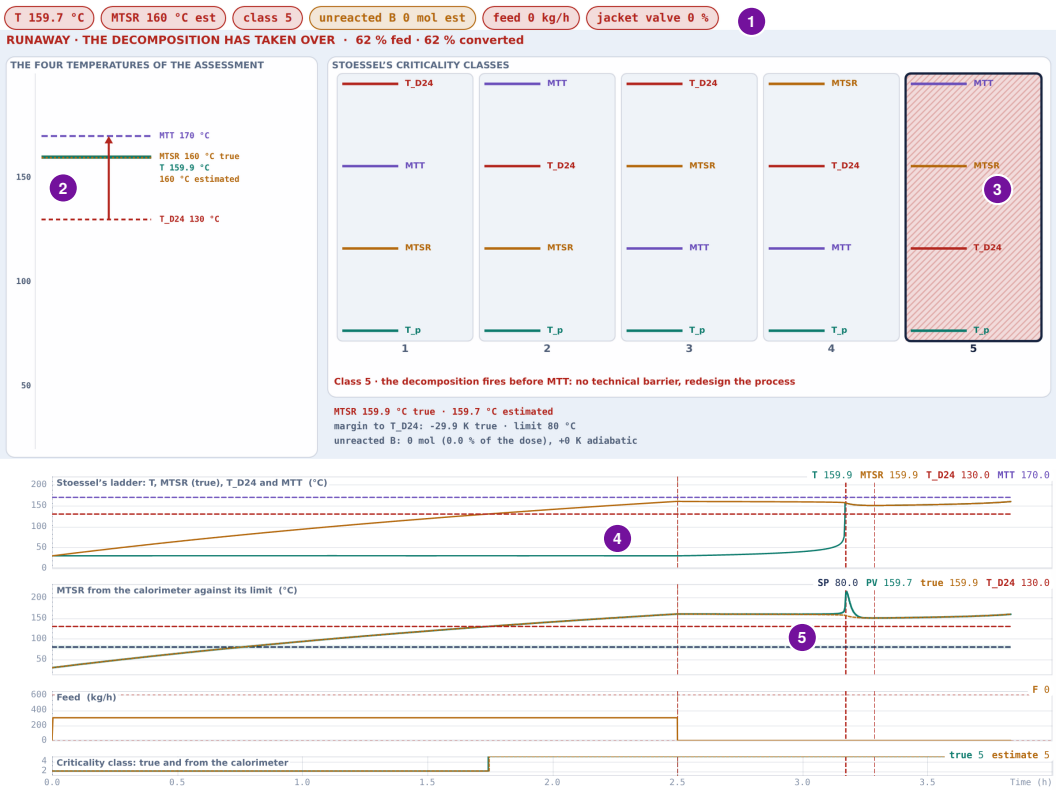


Figure 2.5 The semi-batch reactor in case 4, cooling failure at the wrong moment, at $t = 3.83$ h (230 min) with no control changed, a composite of the Safety view (top) and the trends (bottom), with the header, the controls and the Learn panel cut. The status line reads runaway, the decomposition has taken over, with 62 % of the feed in and 62 % converted, and the badges (1) give TT-101 at 159.7 °C, the calorimeter's MTSR at 160 °C, class 5, no unreacted B estimated, no feed and the split-range valve at 0 %. The ladder of the four temperatures (2) places the true batch at 159.9 °C and the true MTSR at 160 °C (160 °C estimated) between T_{D24} at 130 °C and MTT at 170 °C, with the decomposition's arrow drawn from T_{D24} up to MTT, and class 5 is picked out and hatched among Stoessel's five (3), over a line reading that the decomposition fires before MTT with no technical barrier. The margin line reads −29.9 K to T_{D24} against the 80 °C limit. The trend strips, over a 4.0 h window, show the true MTSR climbing while the batch holds 30 °C (4), to its value at the failure, 160.4 °C in Table 2.5, and the feed of the fixed profile stopping at 2.5 h when the circulation is lost. The batch creeps and then jumps, the calorimeter's MTSR spikes during the jump against its 80 °C limit (5), and the class steps from 2 to 5 long before the failure in the truth and the estimate. Dashed markers stand at the failure, at the crossing of T_{D24} with the TAH and TAHH trips, and at the decomposition taking over.

the feed back. It is safe, and very slow. Raising the setpoint to 70 °C before Run, which starts the charge at 70 °C, shows where the time goes. The same limit leaves 53 mol unreacted at the failure, the batch peaks at 72.0 °C at 57 min, and the dose finishes at 6 h 47 min, two and a half hours of it spent waiting for the cooling.

> **TRY IT IN THE APP** **5 · The Same Failure, Accumulation-Limited**
>
> Load the scenario and press Run, then watch the feed. The accumulation limit lets the pump run only as fast as the cold batch reacts, 28 kg/h at 2 h 30 min against 300 kg/h on the fixed profile. When the circulation fails, the batch peaks at 74.6 °C and stays in class 2. Compare the feed trend with 4 · Cooling failure at the wrong moment.

2.6.5 A Calorimeter Fooled by Fouling

Case 6 runs at 50 °C, where the reaction is 13.5 times slower than at 70 °C, dosed by AIC-104 at the same 80 °C limit, with a jacket that has lost 15 % of its U since the calorimeter was calibrated. The calorimeter overestimates the heat removed, counts too much B as reacted, and AIC-104 feeds into an accumulation it cannot see. The dose finishes at 3 h 42 min instead of the 5 h 54 min it takes with the true conductance. At that moment the calorimeter assumes 4.72 kW/K where the jacket has 4.18 kW/K, reports 1192 mol unreacted where 2054 mol are, and puts MTSR at 81.3 °C while the true value is 104.0 °C. The departure of −861.9 mol matches the −863.8 mol predicted from the integral of the conductance error to 0.2 %. With the True UA option or a correction of 85 % the dose takes 5 h 54 min and the true MTSR never exceeds 80.5 °C. At 70 °C the same fouling hardly matters, because the reaction keeps up with the feed, and the true MTSR peaks at 81.0 °C with the setpoint raised before Run and the charge started at 70 °C.

Pause. Try Exercise 2.2 before reading on.

2.6.6 A Weak Catalyst

Case 7 runs at 70 °C with a catalyst at a quarter of its activity and temperature-limited dosing. The cut acts only twice, at 8.6 and 23.9 min, and the batch never goes above 72.35 °C, less than the 73.56 °C of case 2. Yet the unreacted B reaches 765 mol against 234 mol, MTSR reaches 90.19 °C, and the dose finishes at 4 h 09 min as if nothing were different. A slower reaction releases its heat later rather than hotter. Accumulation-limited dosing reads the slower reaction in the calorimeter and slows the feed, so MTSR peaks at 80.35 °C, the unreacted B at 391 mol, and the dose takes 6 h. On identical seeds the temperature-limited dose overshoots the 80 °C limit by 10.2 K and the accumulation-limited dose by 0.35 K. Anything that slows the reaction makes the two strategies disagree in the same way, an inhibitor in the feed or a missed initiation as well as a weak catalyst.

Pause. Try Exercise 2.3 before reading on.

2.6.7 Warm Cooling Water

Case 8 runs the fixed profile in three hours, 400 kg/h, with cooling water at 30 °C. After the first 30 min the jacket removes 130 to 138 kW, all that full cooling can, while the reaction makes 142 to 146 kW until the last minute of the dose. The split-range signal sits at 0 % for 93 % of the dose, and the batch runs at 73.9 to 74.8 °C with a peak of 75.27 °C. Temperature-limited dosing cuts the feed 14 times, holds the batch below 72.84 °C, and stretches the dose from 3 h 00 min to 3 h 34 min. Accumulation-limited dosing, with the same cut behind it, finishes at 4 h 45 min. The jacket, not the recipe, sets the pace, and a saturated temperature loop has no handle on it but the feed.

2.6.8 The Challenge

The challenge charges the batch at 30 °C, to be heated to 70 °C and dosed, warms the cooling water from 8 to 30 °C at 2 h, and loses the circulation for 10 min at 2 h 30 min. The target is the dose finished by 4 h 00 min, the true MTSR never above 130 °C, the true batch temperature within 2 K of 70 °C from 45 min until the dose is complete, and no relief or quench, judged when the batch is complete. Settings are changed before Run, and a changed model parameter or disturbance voids the attempt (Section 1.6.4). As shipped, the dose starts at once on the four-hour profile with a sluggish batch loop ($K_c = 1.5$, $\tau_I = 60$ min, no hold). At 45 min the batch is still at 47.2 °C, and the B fed

into it cold ignites at about 1 h and takes it to 84.8 °C, tripping the TAH at 1 h 00 min. The worst deviation after 45 min is 22.76 K, 20.76 K outside the band, MTSR reaches 92.1 °C and the dose finishes at 4 h 15 min, 15 min late. The suggested configuration of Table 2.5 starts the dose at 20 min, once the batch is hot, and runs the one-step optimizer against a 75 °C limit under a 450 kg/h ceiling, with the batch loop at 6 and 15 min and its integral held while the jacket cannot answer. The dose then finishes at 3 h 43 min (223.0 min), 17.0 min early, with a worst deviation of 1.61 K, 0.39 K inside the band, never more than 198 mol unreacted and MTSR never above 75.19 °C, 54.8 K under its bound. AIC-104 in place of the optimizer, at the same limit and ceiling, stretches the dose to 5 h 04 min and misses the band by 0.68 K. A fixed profile can pass as well, because the failure comes at a known time. With the suggested batch loop, a fixed 3.5 h profile started at 10 min finishes at 3 h 50 min within 1.85 K, while the same profile started at 20 min misses the deadline by 6 s. A second ten-minute failure, fired by hand outside the graded attempt, tests a configuration further. At 3 h 20 min it leaves the optimizer's MTSR at 75.2 °C and its dose done at 3 h 53 min, although the batch strays 2.08 K, while the fixed profile reaches an MTSR of 77.1 °C, strays 2.70 K and ends its dose after 4 h 00 min. Over second failures every 5 min from 25 min to 3 h 45 min the optimizer's MTSR stays below 75.5 °C, but the band is missed at 26 of the 41 times, by up to 1.39 K, so the accumulation limit protects the MTSR wherever the failure falls but not the band.

> ✦ **DID YOU KNOW? Faster and Safer at Once**
>
> From the simulator. In the challenge the shipped settings end the dose at 4 h 15 min with up to 1196 mol of B unreacted and an MTSR of 92.1 °C, the suggested ones at 3 h 43 min with at most 198 mol and 75.19 °C (Table 2.5). The faster batch stores less heat.

2.6.9 Predictive Dosing Against the One-Step Optimizer

Scenario 9 is the warm-water batch of case 8 dosed by the MPC, with the 80 °C MTSR limit, a 600 kg/h ceiling and the cut at 72 °C. Table 2.6 runs it, the dosing cases and the challenge under the MPC and the shipped strategies with the same seed, instruments, noise and events, to the completion of the batch or at most 600 min (420 min in the challenge). Neither structure sees anything in advance, the shipped strategies keep their shipped tuning (the suggested one in the challenge), and the MPC is never retuned.

On the warm-water batch the MPC is never cut and the batch never passes 71.87 °C, while the one-step optimizer is cut 21 times, the first at 2.9 min, spends 74.4 min of the dose above 72 °C and peaks at 73.20 °C at 4.3 min, as shown in Figure 2.6. From 45 min to the end of the dose the MPC holds the batch 0.73 K from its setpoint on average against 1.48 K. It is slower. Its dose ends at 218.0 min against 207.9 min, its batch is complete at 231.6 min against 225.8 min, 5.8 min later, and it moves the pump more, 36 199 kg/h of summed feed changes against 29 060 kg/h. The optimizer gains its time by letting the batch run above the cut, and a hotter batch sheds more heat into a jacket that is already at full cooling. On five other noise seeds the MPC is never cut and completes the batch at 231.4 to 233.2 min, while the optimizer is cut 21 times on each and completes at 225.0 to 225.8 min, and halving the integration step leaves both counts and both completion times unchanged.

Neither part of the controller produces the result alone. Without the temperature ceiling the MPC is cut 21 times at every look-ahead from one to ten steps, the optimizer's behavior. With the ceiling kept and a single 30 s step of look-ahead it is cut 19 times, because a change of feed reaches the batch temperature over minutes rather than seconds, and two steps cut it twice and four steps once. One move held over the full 5 min is never cut, but its dose ends at 290.6 min, 73 min later, so the seven moves are what keep it fast. Longer horizons do not help on this batch. The 10 to 240 min sweep is scripted, because the interactive horizon control stops at 60 min. Each horizon is cut once, between 2.9 and 3.9 min. The cause is not isolated, but each cut falls while $\hat{T}_{cw}$ is still between 14.5 and 15.9 °C on its way to 30 °C, when the plan counts on a colder jacket than the plant has. Holding $\hat{\theta}$ at 1 or dropping $\hat{d}_T$ or the MTSR band moves the end of the dose by at most 0.5 min without a

Table 2.6 Shipped Dosing Strategies Against the MPC, Seed 20260926. Times in Minutes, Temperatures in Degrees Celsius

Case	Structure	Dose ends	Batch done	Highest T	Highest MTSR	Worst dev. (K)	Cuts	Above 72 °C	Travel (kg/h)
9, water at 30 °C	MPC	218.0	231.6	71.87	79.39	1.87	0	0.0	36199
	one-step optimizer	207.9	225.8	73.20	80.04	3.00	21	74.4	29060
	accumulation-limited	285.1	300.0	72.10	77.58	2.10	9	7.1	8274
	optimizer, 1 min prediction step[a]	207.3	222.9	73.46	79.99	3.09	21	76.8	26403
	temperature-limited, 220 min, cut at 71.25 °C[a]	255.1	272.6	71.99	76.71	1.59	12	0.0	8509
	accumulation-limited, 175 kg/h ceiling[a]	416.9	427.3	71.91	73.77	0.55	0	0.0	381
	MPC, no temperature ceiling	209.2	222.5	73.53	79.69	3.03	21	75.2	26501
	MPC, one 30 s step ahead	209.5	224.3	73.24	79.82	3.02	19	44.1	45037
	MPC, one move held over 5 min	290.6	290.6	71.71	77.80	2.94	0	0.0	13930
2, cascade at 70 °C	one-step optimizer	179.3	203.5	73.16	80.57	2.57	17	38.2	23758
	MPC	178.0	202.7	71.85	80.20	1.85	0	0.0	24842
6, fouled jacket	AIC-104 (shipped)	222.1	—	52.27	104.09	2.22	4	0.0	5573
	one-step optimizer	199.4	588.6	53.11	102.85	2.65	5	0.0	12579
	MPC	194.4	582.4	51.92	102.94	1.92	0	0.0	31858
7, weak catalyst	temperature-limited (shipped)	249.1	346.8	72.35	90.19	1.54	2	5.8	1800
	one-step optimizer	352.4	425.3	72.49	80.14	1.53	2	7.5	25168
	MPC	357.0	429.5	71.85	80.19	1.76	0	0.0	29168
Ch., as shipped[b]	fixed profile (shipped)	254.6	292.1	84.77	92.14	22.76	0	9.4	1800
	one-step optimizer	202.1	232.1	77.82	84.21	22.77	6	31.8	13297
	MPC	206.0	221.2	72.97	80.42	23.01	1	7.2	16608
Ch., suggested[b]	one-step optimizer (suggested)	223.0	236.7	71.55	75.19	1.61	0	0.0	10166
	MPC	221.0	233.9	71.83	75.05	1.83	0	0.0	10355
	optimizer, 525 kg/h ceiling[a]	212.8	226.3	71.87	75.33	1.87	0	0.0	11303

Highest values and the minutes above 72 °C during the dose come from the true state at every physics step. The worst deviation, the largest $|T - T_{SP}|$, runs from 45 min to the end of the dose as the scorecard computes it. Cuts count each start of a temperature cut, including a lowered cut's own, and feed travel sums the absolute change of the feed from scan to scan. A dash marks a batch not complete by 600 min. [a] Setting found by search in hindsight. [b] Every run of the challenge as shipped fails its target, and the three runs with the suggested settings pass. The MPC solves 436, 356, 389, 714, 412 and 442 plans in the six cases in table order, and no plan is rejected in any run.

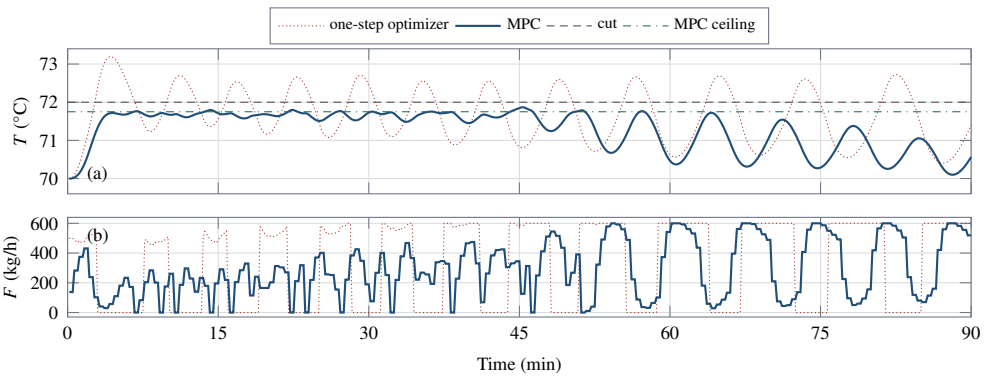


Figure 2.6 Scenario 9, the warm-water batch, over its first 90 min under the MPC (thick solid line) and under the one-step optimizer (thin dotted line) on the same seed. (a) True batch temperature with the 72 °C cut (dashed) and the MPC's 71.75 °C ceiling (dash-dotted). (b) Feed. The optimizer is cut 12 times in this window and 21 times over the dose.

cut, and holding $\hat{T}_{cw}$ at 8 °C ends it 5.1 min sooner but leaves the batch 1.42 K from its setpoint on average and is cut once on two of five other seeds. The back-off has a price of its own. With none the MPC is cut 11 times, and with 1 K it is never cut but its dose ends at 253.0 min, 35.0 min after the shipped 0.25 K.

The shipped strategies keep this batch below 72 °C only with hindsight and more slowly. The fastest such setting found by search, temperature-limited dosing over 220 min with its cut lowered to 71.25 °C, completes the batch at 272.6 min, 41.0 min after the MPC, and is still stopped 12 times by its own lowered cut. Without any cut, accumulation-limited dosing needs a 175 kg/h ceiling and completes at 427.3 min, and the optimizer needs 150 kg/h and 489.2 min.

In case 2 the MPC ends its dose 1.3 min before the optimizer (178.0 against 179.3 min) without a cut against 17. The MPC loses case 7, where it is never cut against twice but its dose ends 4.6 min after the optimizer's (357.0 against 352.4 min), and $\hat{\theta}$ is essential there, since with it held at 1 only 979 kg are fed by 600 min. In the cold cases 3 to 5 the MPC, AIC-104 and the optimizer feed within 3 kg of one another by 600 min. Case 6 fools every structure through the calorimeter. The MPC ends its dose first, at 194.4 min, but the true MTSR reaches 102.94 °C while the calorimeter reports 80.01 °C, against 104.09 °C under AIC-104 and 102.85 °C under the optimizer, and with the True UA option or the 85 % correction the MPC holds 80.18 °C and 80.17 °C. The challenge as shipped defeats every structure, and the MPC ends its dose 3.9 min after the optimizer with a worst deviation of 23.01 K from the cold start. With the suggested settings and the strategy switched to the MPC it passes and ends the dose 2.0 min before the optimizer, but the optimizer under a 525 kg/h ceiling chosen in hindsight ends it at 212.8 min, 8.2 min before the MPC, and passes as well. The MPC moves the pump more than the optimizer in cases 2, 6, 7 and 9 and in the challenge, and less only in the cold cases. A move weight of 0.1 would cut scenario 9's travel to 5189 kg/h and complete the batch 1.4 min later without a cut, but the tuning rule chose 0.001.

Every protection acts on the MPC's request. In every case run to its end no scan feeds through a cut, TAH, FSL or TAHH, and the protections stop requested feed on 1692 scans in each of cases 4 and 5 and on 92 in the challenge. Forced, a TAH at 71 °C stops 7166 scans of requests in scenario 9, a TAHH at 71.5 °C latches at 3.8 min and fires the quench, and case 4 with no MTSR margin reaches the TAHH at 203.0 min, after which no feed enters.

2.6.10 The Plan Linearized Once or Along Its Path

The iterated setting was built on the expectation that the single linearization is most wrong where the accumulation limit binds, and the runs do not bear that out. Every accepted single plan was simulated again on its own nonlinear model over every scenario, each accumulation-limited one also with each disturbance at the limits of its range. In the 46 runs whose MTSR row is active at 20 % or more of the dosing decisions, the plan under-predicts MTSR by at most 0.27 K at the 95th percentile, because MTSR is linear in the unreacted B through Equation (2.7) and the free response is exact. The temperature is where the single plan errs, in case 10 high by 0.48 K at its worst step at the median decision and by up to 1.52 K.

On the case where the accumulation limit binds with the single plan's largest MTSR error, both linearizations hold the true MTSR 0.01 K under its limit, as Table 2.7 shows. Every dose there ends before 4 h 00 min, so the target turns on the temperature band alone. On the default seed the iterated batch peaks at 71.87 °C and passes while the single one reaches 72.11 °C, but over six seeds the iterated batch peaks at 71.84 to 72.30 °C and passes twice, the single batch at 72.00 to 72.19 °C and never. The single linearization with its ceiling 1 K below the cut, the setting the first round's tuning rule picks here, peaks at 71.26 to 71.45 °C and passes on all six seeds for 1.6 min more dose. Where the limit binds the two linearizations tie, and a wider back-off wins.

Where the temperature ceiling binds the iteration helps. In case 10 the dose ends at 175.4 min against 178.0 and the batch is complete at 198.2 min against 202.7, 4.5 min sooner, both batches peaking at 71.86 and 71.85 °C without a cut. The iterated batch runs at full cooling at 98 decisions against 33, spends 88.0 min of the dose above 71.5 °C against 76.5 and moves the pump more (28 422 kg/h of travel against 24 842 kg/h), and on five other seeds it completes 3.1 to 5.6 min sooner. The decision at 120 min shows why, as shown in Figure 2.7. From the same state at full cooling, the single plan asks for 363 kg/h and predicts the batch at the ceiling by the end of the horizon, while the controller's own model under that plan never rises above 71.40 °C and ends 1.54 K lower. The

Table 2.7 The MPC Linearized Once and Along Its Plan, Seed 20260926. Times in Minutes, Temperatures in Degrees Celsius

Structure	Dose ends	Batch done	Highest T	Margin (K)	Cuts	QPs	Target
Ch., suggested, MPC, activity 150 %[a]							
one-step optimizer	204.4	213.5	72.19	−0.44	2	—	no
MPC, single	204.9	212.5	72.11	0.01	1	1	no, 0/6
MPC, iterated	204.4	212.0	71.87	0.01	0	2.40	yes, 2/6
single, 1 K back-off	206.0	213.8	71.31	0.02	0	1	yes, 6/6
one relinearization	204.4	212.0	71.87	0.01	0	2	yes, 2/6
Case 10, case 2's batch with water at 8 °C							
fixed profile (case 2)	250.1	266.1	73.56	3.80	0	—	
MPC, single	178.0	202.7	71.85	−0.20	0	1	
MPC, iterated (case 10)	175.4	198.2	71.86	−0.16	0	3.58	
one relinearization	175.5	198.4	71.86	−0.16	0	2	
Scenario 9, water at 30 °C							
MPC, single (case 9)	218.0	231.6	71.87	0.61	0	1	
MPC, iterated	217.2	229.5	71.92	0.95	0	3.81	
one relinearization	212.5	225.5	71.87	1.05	0	2	

[a] The challenge with its suggested settings, the strategy switched to the MPC (the optimizer in the first row) and the catalyst at 150 %, where the MTSR limit is 75 °C. The margin is the limit less the highest true MTSR, QPs counts the programs of a decision while dosing, and the target column gives the default seed and the count over it and five other seeds.

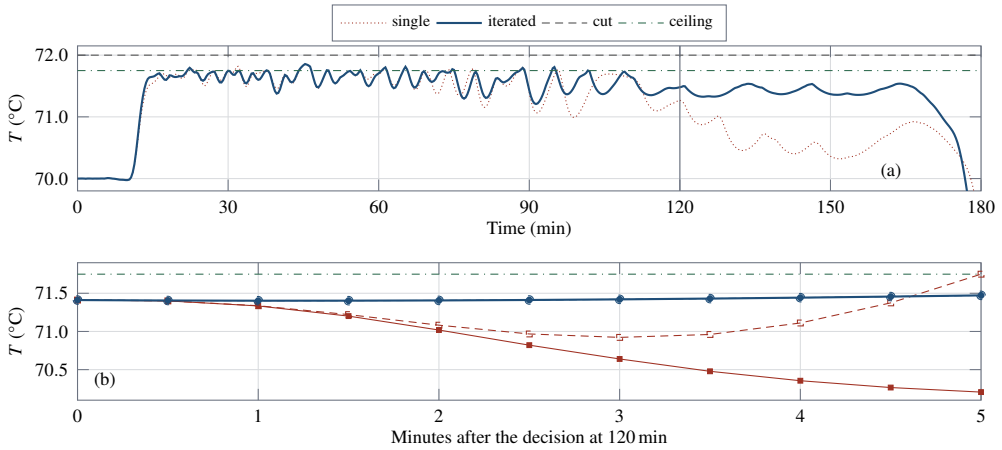


Figure 2.7 Case 10, case 2's batch under the MPC, linearized once and along its plan on the same seed. (a) True batch temperature to 180 min, linearized once (thin dotted line) and along the plan (thick solid line), with the 72 °C cut (dashed) and the 71.75 °C ceiling (dash-dotted). (b) The decision at 120 min, marked in (a), with both plans computed from the same state, squares for the single plan and circles for the iterated one. Dashed lines with open marks are each plan's predicted temperature, and solid lines with filled marks the controller's nonlinear model under that plan.

iterated plan asks for the 600 kg/h pump limit at every step and predicts 71.47 °C, what the model gives. On scenario 9's warm water the batch is complete at 229.5 min against 231.6, and 14 % of the dosing plans keep the single one. One relinearization with no test of convergence does as well on case 10 (198.4 min) and better on scenario 9 (225.5 min) at about half the cost (0.31 ms), but on one of scenario 9's five other seeds it is cut once, where neither linearization is.

Under the worst cases the application can produce, 39 cases run under every structure (every automatic scenario, each disturbance at the limits of its range on case 2's batch and the suggested challenge, the cooling lost for good at 1, 2 and 3 h, a second ten-minute failure every 30 min of the

challenge), the iterated MPC reaches no level the single one does not. Both peak at 74.64 °C, their highest true MTSR is 102.98 and 102.94 °C on case 6's fooled calorimeter, each passes the limit by more than 1 K in three runs, and neither reaches T_{D24}, the TAH or a runaway. Running closer to the ceiling, the iterated batch meets an unplanned event hotter. With the cooling lost for good at 2 h on case 2's batch it peaks at 73.16 °C against 72.74 °C, both cut once.

2.7 Technology in Practice

What Measures It

The dosing loop of this chapter controls a quantity that no instrument reads. QI-105 builds it from two temperatures, a feed flow and a heat-transfer coefficient it assumes, which is what a heat-flow calorimeter does in the laboratory. There the temperature difference between reactor and jacket measures the rate of heat production, and the accumulation of unreacted material in a semi-batch reaction can be measured [7]. Where the heat evolution can be measured or estimated online, feedback can hold it at a setpoint [26]. The two temperature elements are those of the cooled reactor of Volume 7, Chapter 2, here with lags of 15 and 6 s, and the feed meter FT-103, with its 3 s lag and 0.3 % of noise, supplies the total that ends the dosing phase.

> **TECHNOLOGY Sensor: A Coriolis Meter on the Feed**
>
> A Coriolis meter shakes one or more tubes carrying the liquid at their resonant frequency. The inertia of the flowing liquid twists the tubes in proportion to its mass flow, so the meter measures mass flow directly and, from the frequency, the density as well. It needs no straight pipe before or after it and is very accurate and dependable, which is why it is used where the flow is product being bought and sold. Its weaknesses are cost in large sizes, vibration from nearby machines, which disturbs the tubes' own motion, and gas bubbles trapped in bent tubes, which spread the mass unevenly and give errors in flow and density [27]. In this chapter the meter's total ends the dosing phase and enters the calorimeter's count of B, so a meter error passes straight into the estimated accumulation.

What Moves It

Two final elements act on the batch. The split-range pair on the jacket loop is the arrangement drawn for Volume 7, Chapter 2, here with a steeper slope on the hot side than on the cooling side (Section 2.3). The other is the dosing pump. Industrial practice often runs a semi-batch reactor isothermally and controls its temperature by the feed rate of the limiting reactant, which is hard because a larger cold feed first lowers the temperature before the faster reaction raises it [26]. The temperature-limited strategy of this chapter cuts the feed on the temperature in that spirit, while the accumulation-limited one sets it from the estimate instead.

> **TECHNOLOGY Final Element: A Metering Pump**
>
> A small dosing stream is often moved by a metering pump, a pump mechanism, motor and drive electronics in one package, whose 4 to 20 mA control signal sets the flow from zero to a maximum chosen on the pump. Many use a plunger driven at variable intervals, so the flow arrives as discrete pulses rather than a continuous stream. A positive-displacement pump discharges a fixed volume for every revolution, and with its discharge blocked it locks and overloads its motor unless a relief path returns the flow [27]. The chapter's pump P-103 delivers the requested feed up to 600 kg/h with no dynamics of its own, and the interlocks stop it directly.

Where the Controller Runs

Batch control runs on stand-alone single-loop controllers, programmable logic controllers, distributed control systems and personal computers, often in combination [26]. The batch control committee of ISA, whose standards form the ISA-88 series, describes a functional model that relates five types of control: recipe management, scheduling, sequential control, regulatory control and safety interlock

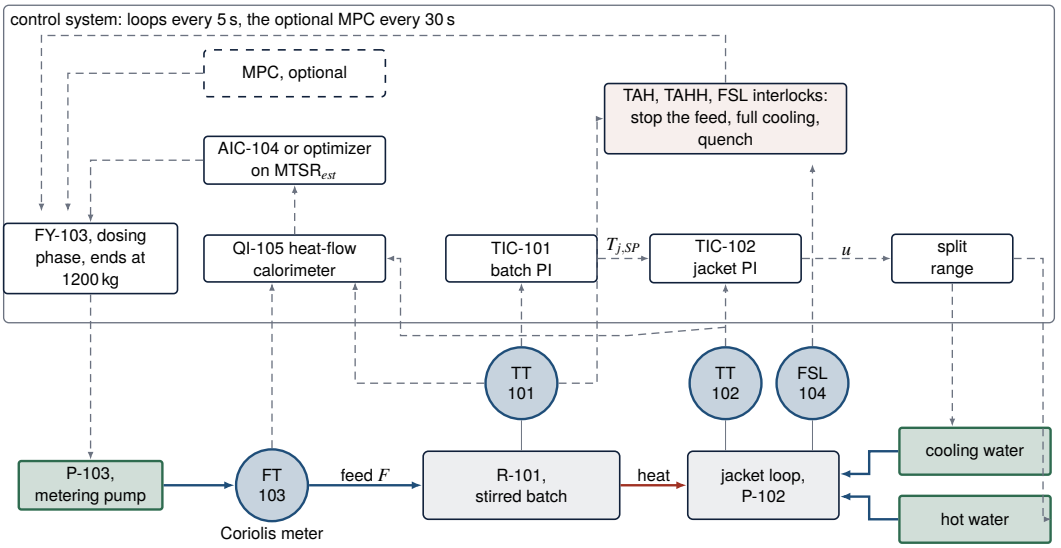


Figure 2.8 The chapter's loops on plant hardware. The cascade TIC-101 and TIC-102 reads the batch and jacket temperatures and moves the cooling-water and hot-water valves through the split range. The calorimeter QI-105 reads both temperatures and the Coriolis meter FT-103, and AIC-104, the one-step optimizer or the dashed optional MPC sets the feed request of the dosing phase, which the metering pump delivers until the measured total reaches 1200 kg. Notice that the interlocks read the same TT-101 as the cascade and act on the same feed, with no second sensor and no separate valve.

systems [28]. In those terms the cascade TIC-101 and TIC-102 is regulatory control, the dosing strategy and the measured total that ends the dose are the sequential control of the dosing phase, and TAH, TAHH and FSL-104 are the safety interlocks. A distributed control system executes such loops with cycle times from milliseconds to minutes [29]. The chapter's loops act every 5 s and the predictive controller every 30 s, as Figure 2.8 draws them on plant hardware.

IN PRACTICE Three Ways to Stop the Feed

A guide of the British safety regulator compares three ways to protect a semi-batch reactor whose temperature is safety critical. In the first an operator watches one thermocouple and shuts off the addition by hand near the upper limit. In the second a temperature controller throttles the reactant valve and an alarm calls for a manual valve to be closed. In the third an independent temperature switch, set below the upper limit, is hardwired to a separate shut-off valve in the reactant line, and sensors of different types avoid a common-mode failure. The guide points to IEC 61508 and IEC 61511 for the integrity such systems need [7].

What Is Modernizing It

Two developments are changing how such a batch is dosed. The first is optimization repeated from batch to batch. On a one-ton industrial reactor, a policy that ran isothermally and then let the reaction finish without jacket cooling, with the switching time adapted from run to run, cut the reaction time by 35 % against the isothermal policy of industrial practice. It did so after only two batches and without off-specification product, keeping a back-off from the final temperature limit for safety [26]. That is the run-to-run idea of Section 1.4.5 applied to a recipe. The second is predictive control on a process model, which the same chapter counts among the advances that have influenced industrial implementation [26], and which the MPC of this chapter brings to the dose on the balances of Equation (2.12).

2.8 Limitations

The kinetics, the heat-transfer law, the fouling and agitator factors and the decomposition parameters are illustrative, and T_{D24}, MTT, the MTSR margins and the class thresholds are teaching values, so the app demonstrates Stoessel's method [6] without assessing any real process. One synthesis and one first-order decomposition leave out induction periods, autocatalysis and side reactions, and the model ends at the rupture disc without pressure, venting or boiling.

The calorimeter keeps the declared law and constants, so its only systematic errors are those a reader introduces: fouling, the agitator, a wrong correction, an unmeasured feed strength or a changed parameter. It attributes all heat to the synthesis and cannot see a decomposition, and it is never recalibrated during a batch.

OPEN CHALLENGE Models Good Enough and Small Enough

A 2017 handbook chapter on batch control found that predictive control and optimization have influenced industrial practice, but that they depend on process models of reasonably high fidelity and modest complexity, which are difficult to achieve. It called for reduced-order models built for control, and for better sensors and analyzers, because estimates can never completely replace measurements [26]. This chapter's calorimeter and predictive controller both keep a declared heat-transfer law that a fouled jacket makes wrong. Progress would be the calorimeter the Conclusion asks for, one that recalibrates its conductance during the batch.

The heat loss and the water in the jacket give a real batch some margin that the adiabatic criterion does not count. With the circulation lost for good at 2 h 30 min (case 3 with the setpoint changed and the quench disarmed before Run, and the open-ended cooling failure fired at that time), a batch dosed at 35 °C survives with MTSR at 148.7 °C, although it passes T_{D24}, while one dosed at 34 °C, with MTSR at 151.7 °C, reaches the disc at 12 h 45 min. The criterion is conservative on purpose. The interlocks act on TT-101 with its lag and have no redundancy or voting, and the optimizer looks one step ahead.

The predictive controller's model leaves out fouling, the agitator, the feed strength, circulation below design, the decomposition and the quench mass, and its 22 s jacket lag is about four times too fast with water at 30 °C. It inherits the calorimeter's errors, and on a fouled jacket its estimates settle at $\hat{\theta} = 1.86$ and $\hat{T}_{cw} = 4.0\,^{\circ}\mathrm{C}$. Its 5 min horizon does not plan the end of the dose, and its one slack

per band cannot price a violation step by step. It trades 5.8 min of the warm-water batch for the 21 cuts it avoids, and in case 7 and the suggested challenge a classical setting finishes sooner.

The iterated linearization makes the plan exact about the controller's model, not the plant, so it removes none of this mismatch. Three relinearizations leave 2 % of case 10's dosing plans and 14 % of scenario 9's unsettled, where a fourth, outside the declared rule, would leave 0.6 and 6.6 %.

2.9 Conclusion

A semi-batch reactor stores its hazard in the reagent that has been fed and has not yet reacted, and this simulator logs that accumulation, its MTSR and Stoessel's class beside a temperature trend that cannot show them. At 70 °C the fixed four-hour dose never holds more than 234 mol, while at 30 °C the same dose holds its temperature within 0.2 K and accumulates 4398 mol by 2 h 30 min, an MTSR of 160.4 °C. Losing the circulation then runs the batch through a synthesis runaway to 157.1 °C and a decomposition to its rupture disc at 3 h 55 min, while accumulation-limited dosing meets the same failure with MTSR held at 80.0 °C and a peak of 74.6 °C. The highest MTSR of the dose is lowest, 75.6 °C, near 66 °C. A 15 % fouled jacket leaves the calorimeter reporting 1192 mol where 2054 mol are, and the challenge improves from a 22.76 K excursion to a dose finished at 3 h 43 min within 1.61 K. On warm cooling water the optional predictive controller is never cut where the one-step optimizer is cut 21 times, and it completes the batch 5.8 min later. Relinearized along its own plan it ties with one linearization where MTSR binds, because MTSR is linear in the unreacted B, and completes case 10's batch 4.5 min sooner at the same 71.9 °C peak. The kinetics and the thresholds are teaching values, and the decomposition is a single anchored reaction. A fuller model would add an autocatalytic decomposition, a vent model at the disc, and a calorimeter that recalibrates its conductance. The simulator makes the accumulation of a fed reactor visible, and with it the reason that the dosing, not the temperature loop, keeps the batch safe.

KEY POINTS

- A semi-batch reactor stores its hazard in the reagent that has been fed and has not yet reacted, and its temperature trend cannot show it.
- The same four-hour dose holds at most 234 mol unreacted at 70 °C, but 4398 mol by 2 h 30 min at 30 °C, an MTSR of 160.4 °C.
- Accumulation-limited dosing meets the same cooling failure with MTSR held at 80.0 °C and a peak of 74.6 °C.
- A jacket fouled by 15 % leaves the calorimeter reporting 1192 mol where 2054 mol are.

Where else this applies. An artificial pancreas limits the insulin already delivered whose effect has not yet appeared, the insulin on board of Volume 19, Chapter 2, as AIC-104 limits the B that has been fed and has not yet reacted.

References

[1] P. Hugo and J. Steinbach, "A comparison of the limits of safe operation of a SBR and a CSTR," *Chemical Engineering Science*, vol. 41, no. 4, pp. 1081–1087, 1986.

[2] M. Steensma and K. R. Westerterp, "Thermally safe operation of a semibatch reactor for liquid-liquid reactions. Slow reactions," *Industrial & Engineering Chemistry Research*, vol. 29, no. 7, pp. 1259–1270, 1990.

[3] K. R. Westerterp and E. J. Molga, "Safety and runaway prevention in batch and semibatch reactors—A review," *Chemical Engineering Research and Design*, vol. 84, no. 7, pp. 543–552, 2006.

[4] P. Lerena, W. Wehner, H. Weber, and F. Stoessel, "Assessment of hazards linked to accumulation in semi-batch reactors," *Thermochimica Acta*, vol. 289, no. 2, pp. 127–142, 1996.

[5] R. Gygax, "Chemical reaction engineering for safety," *Chemical Engineering Science*, vol. 43, no. 8, pp. 1759–1771, 1988.

[6] F. Stoessel, *Thermal Safety of Chemical Processes: Risk Assessment and Process Design*. Wiley, 2008.

[7] Health and Safety Executive, *Designing and Operating Safe Chemical Reaction Processes*, ser. HSG143. Sudbury, UK: HSE Books, 2000. [Online]. Available: https://www.hse.gov.uk/pubns/priced/hsg143.pdf

[8] I. S. de Buruaga, A. Echevarría, P. D. Armitage, J. C. de la Cal, J. R. Leiza, and J. M. Asua, "On-line control of a semibatch emulsion polymerization reactor based on calorimetry," *AIChE Journal*, vol. 43, no. 4, pp. 1069–1081, 1997.

[9] F. Rossi, S. Copelli, A. Colombo, C. Pirola, and F. Manenti, "Online model-based optimization and control for the combined optimal operation and runaway prediction and prevention in (fed-)batch systems," *Chemical Engineering Science*, vol. 138, pp. 760–771, 2015.

[10] H. Schuler and C.-U. Schmidt, "Calorimetric-state estimators for chemical reactor diagnosis and control: review of methods and applications," *Chemical Engineering Science*, vol. 47, no. 4, pp. 899–913, 1992.

[11] O. Ubrich, B. Srinivasan, P. Lerena, D. Bonvin, and F. Stoessel, "Optimal feed profile for a second order reaction in a semi-batch reactor under safety constraints," *Journal of Loss Prevention in the Process Industries*, vol. 12, no. 6, pp. 485–493, 1999.

[12] ——, "The use of calorimetry for on-line optimisation of isothermal semi-batch reactors," *Chemical Engineering Science*, vol. 56, no. 17, pp. 5147–5156, 2001.

[13] D. I. Townsend and J. C. Tou, "Thermal hazard evaluation by an accelerating rate calorimeter," *Thermochimica Acta*, vol. 37, no. 1, pp. 1–30, 1980.

[14] A. Keller, D. Stark, H. Fierz, E. Heinzle, and K. Hungerbühler, "Estimation of the time to maximum rate using dynamic DSC experiments," *Journal of Loss Prevention in the Process Industries*, vol. 10, no. 1, pp. 31–41, 1997.

[15] W. Regenass, "The development of heat flow calorimetry as a tool for process optimization and process safety," *Journal of Thermal Analysis*, vol. 49, no. 3, pp. 1661–1675, 1997.

[16] U.S. Chemical Safety and Hazard Investigation Board, "T2 Laboratories, Inc. runaway reaction," U.S. Chemical Safety and Hazard Investigation Board, Investigation Report 2008-3-I-FL, Sep. 2009. [Online]. Available: https://www.csb.gov/t2-laboratories-inc-reactive-chemical-explosion/

[17] ——, "Investigation report: Pesticide chemical runaway reaction pressure vessel explosion, Bayer CropScience, LP, Institute, West Virginia, August 28, 2008," U.S. Chemical Safety and Hazard Investigation Board, Washington, DC, Tech. Rep. 2008-08-I-WV, Jan. 2011. [Online]. Available: https://www.csb.gov/assets/recommendation/bayer_report_final.pdf

[18] 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.

[19] 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.

[20] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[21] A. Kummer, T. Varga, and L. Nagy, "Semi-batch reactor control with NMPC avoiding thermal runaway," *Computers & Chemical Engineering*, vol. 134, p. 106694, 2020.

[22] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[23] J. H. Lee and N. L. Ricker, "Extended kalman filter based nonlinear model predictive control," *Industrial & Engineering Chemistry Research*, vol. 33, no. 6, pp. 1530–1541, 1994.

[24] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer, "Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations," *Journal of Process Control*, vol. 12, no. 4, pp. 577–585, 2002.

[25] S. Gros, M. Zanon, R. Quirynen, A. Bemporad, and M. Diehl, "From linear to nonlinear MPC: bridging the gap via the real-time iteration," *International Journal of Control*, vol. 93, no. 1, pp. 62–80, 2020.

[26] D. Bonvin and G. François, "Control and optimization of batch chemical processes," in *Coulson and Richardson's Chemical Engineering*. Elsevier, 2017, pp. 441–503.

[27] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[28] International Society of Automation, "ISA88: Batch control, committee scope and purpose," www.isa.org, standards committees, 2026, accessed 22 November 2026.

[29] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

2.10 Exercises

2.1 Concept check · 2 min

The same four-hour dose of B is run into one batch held at 70 °C and into another held at 30 °C. Both temperature loops hold their setpoints within a fraction of a kelvin. Which statement about a cooling failure halfway through the dose is right?

(a) The cold batch is safer, because it stays 40 K further from T_{D24} during the dose

(b) The two batches are equally safe, because both hold their setpoints within a fraction of a kelvin

(c) The cold batch is more dangerous, because its reaction lags the feed and the unreacted B raises its MTSR

(d) The cold batch is more dangerous, because its jacket runs near full cooling and has no reserve left

2.2 Concept check · 2 min

Since the calorimeter QI-105 was calibrated, the jacket has fouled and lost 15 % of its heat-transfer coefficient. The calorimeter still uses the clean-jacket law, and AIC-104 doses on its estimate. What happens?

(a) The calorimeter reports too much unreacted B, so AIC-104 slows the feed more than it needs to

(b) The calorimeter reports the right unreacted B, because the error in UA cancels in the heat balance

(c) The calorimeter reports too little reaction heat, so the dose finishes later than it should

(d) The calorimeter reports too little unreacted B, so AIC-104 feeds faster than the true MTSR allows

2.3 Concept check · 1 min

In case 7 the catalyst runs at a quarter of its activity, and the dose is temperature-limited: the feed is cut while TT-101 is more than 2 K above the setpoint. Why does the cut act only twice in the whole dose?

(a) A slower reaction releases its heat later rather than hotter, so the batch seldom rises 2 K above its setpoint

(b) The cut is disabled by temperature-limited dosing while the calorimeter reports a weak catalyst

(c) The batch loop raises its own gain when the catalyst is weak, which holds the batch near the setpoint

(d) The weak catalyst lowers the heat capacity of the batch, so the same heat raises it less

2.4 Theory and calculation · 5 min

The recipe doses 8000 mol of B, with $(-\Delta H_r) = 200\,\mathrm{kJ/mol}$, into a batch whose heat capacity grows from 5.40 MJ/K for the charge to 7.56 MJ/K when the dose is complete.

(a) Find the heat the whole dose carries and the temperature rise of the finished batch if it were all released at once with no heat removed. *Checkpoint:* 1.6 GJ and 211.6 K.

(b) Find the rise of MTSR for every 1000 mol of unreacted B, in the finished batch and in the charge. *Checkpoint:* 26.5 K and 37.0 K.

(c) A batch near the end of its dose sits at 70 °C. How much unreacted B would put its MTSR at $T_{D24} = 130\,^{\circ}\mathrm{C}$? *Checkpoint:* 2268 mol.

2.5 Theory and calculation · 6 min

Use the quasi-steady estimate of Equation (2.9) with $k_{70} = 2.4 \times 10^{-6}\,\mathrm{m^3/(mol\,s)}$, $E_r = 120\,\mathrm{kJ/mol}$, a feed of 300 kg/h carrying 6.667 mol/kg, and the state at the end of the dose: $V = 4.53\,\mathrm{m^3}$, $n_A = 4234\,\mathrm{mol}$ and $mc_p = 7.56\,\mathrm{MJ/K}$.

(a) Estimate the unreacted B and the MTSR at the end of the dose at 70 °C. *Checkpoint:* About 248 mol and 76.6 °C.

(b) Repeat the estimate at 30 °C, where the rate constant is 257 times smaller, and say what the result means. *Checkpoint:* About 64 000 mol, more than the whole dose.

(c) Differentiate the quasi-steady MTSR with respect to T and show that it is lowest where $\Delta T_{ad,acc} = RT_K^2/E_r$. Evaluate this at 67 °C. *Checkpoint:* 8.0 K.

2.6 Theory and calculation · 5 min

At the start of the dose the wall conductance is $UA = 5.00\,\mathrm{kW/K}$, and the jacket circulates 2 kg/s of water with $c_{pj} = 4180\,\mathrm{J/(kg\,K)}$. Each kilogram of feed carries 6.667 mol of B at 200 kJ/mol.

(a) Find the effective conductance of wall and circulation in series, and the most heat full cooling can remove from a batch at 70 °C with cooling water at 8 °C and at 30 °C. *Checkpoint:* 3.13 kW/K, 194 kW and 125 kW.

(b) Case 8 doses the 1200 kg in three hours. Find the heat this feed makes once it reacts as it arrives, and the warmest cooling water that could still remove it at 70 °C. *Checkpoint:* 148 kW and about 22.6 °C.

(c) Find the highest steady feed the warm jacket can carry at 70 °C and the dose time it implies. *Checkpoint:* About 338 kg/h and 3.55 h.

2.7 App challenge · 9 min

Use the default seed and a speed of 60 for both scenarios.

(a) Load scenario 4, Cooling failure at the wrong moment, change no setting and run until the run stops. Read the Batch and Most unreacted B tiles. *Checkpoint:* Lost at 3 h 55 min, 4398 mol.

(b) Load scenario 5, The same failure, accumulation-limited, change no setting and run to 5 h (300 min). Read the Highest T tile, the highest class on the Safety view, and the fed mass, column `fed_kg`, at 2 h 30 min. *Checkpoint:* 74.6 °C, class 2, 276 kg.

Explain why the feed interlock FSL-104, which stops the feed in both runs when the circulation fails, does not decide the outcome.

2.8 App challenge · 8 min

Load scenario 6, Fouled jacket, fooled calorimeter. Use the default seed and a speed of 60, and export the file after each run.

(a) Run until the Dose finished tile shows a time. In the first row with column `dose_done` equal to 1, read the true and estimated unreacted B (`nB_mol`, `nB_est_mol`), and read the Highest MTSR tile. *Checkpoint:* 3 h 42 min, 2054 mol true against 1192 mol estimated, 104.1 °C.

(b) Predict the calorimeter's error at that row from $-\int (UA_{est} - UA)(T - T_j)\,\mathrm{d}t/(-\Delta H_r)$, summing over the rows with columns `UA_est_kWK`, `UA_kWK`, `T_meas_C` and `Tj_meas_C`. *Checkpoint:* About −863 mol, against −862 mol from (a).

(c) Press Reset, set UA used by the calorimeter to True UA before Run, and run until the dose finishes. Read the Dose finished and Highest MTSR tiles. *Checkpoint:* 5 h 54 min and 80.5 °C.

Accept the unreacted B within 5 mol and the predicted error within 2 %.

2.9 App challenge · 8 min

Load scenario 9, MPC: the temperature cut, planned ahead, change no setting, and run at a speed of 60 with the default seed until the run stops at the end of the batch. Read the Dose finished, Batch, and Highest T tiles, and count the FEED CUT marks on the batch temperature trend (or the starts of column `feed_cut` in the exported file). Then press Reset, set Dosing strategy to One-step optimizer before Run, and repeat. Which controller finishes sooner, and what does each pay?

2.10 Application area · 3 min

The disc of R-101 is replaced by one with a lower set pressure, at which the solvent boils at MTT = 120 °C, below $T_{D24} = 130$ °C. Classify, with the definitions of the chapter, the recipe batch at 70 °C whose MTSR never passes 76.2 °C, and the cold batch of case 3 at 2 h 30 min, with MTSR at 160.4 °C. What does the new order of the four temperatures mean for a cooling failure?

2.11 Application area · 4 min

In case 4 with the quench armed, the high-high trip adds 1000 kg of solvent at 20 °C to a batch of 3750 kg at 156.7 °C. The batch and the solvent have the same heat capacity per kilogram, the charge's density is 900 kg/m^3 and the feed's 1000 kg/m^3, and 750 kg of feed is in the batch.

(a) Find the temperature after the quench. *Checkpoint:* 127.9 °C.

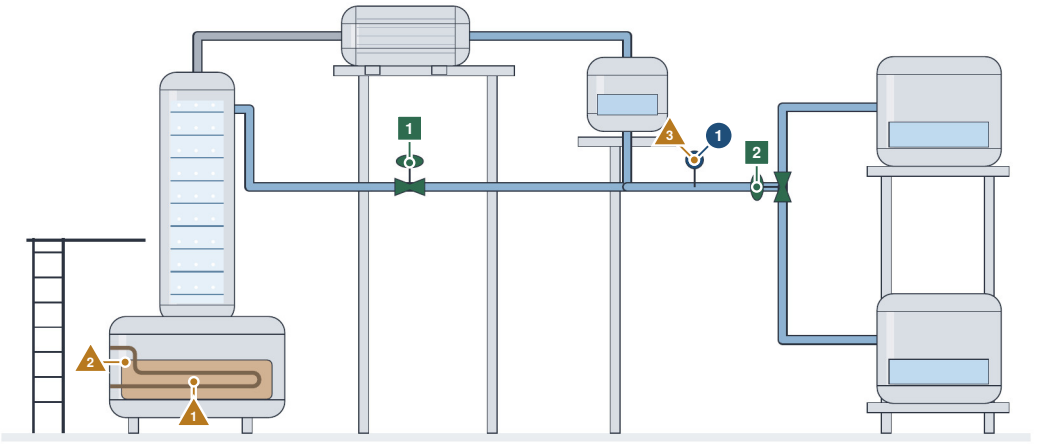
(b) Find the quench mass that would bring the batch to 120 °C, and the liquid volume with each quench in the 6 m^3 vessel. *Checkpoint:* About 1376 kg; 5.19 and 5.61 m^3.

2.12 Review the agent · 5 min

An AI assistant reviews case 6 and writes: "The accumulation-limited dose overshoots its MTSR limit by 24 K because AIC-104 is a PI loop that reacts too late. Switch the dosing strategy to MPC and set its linearization to Iterated. A converged plan predicts exactly what the nonlinear model gives for it, so the MPC will hold the true MTSR at the 80 °C limit." Find the flaw and say how you would check it.

CHAPTER 3

Reflux Policy on a Batch Column With No Steady State to Tune



The controller moves 1 reflux split, the share of condensate returned 2 receiver switch, product or slop

The controller reads 1 distillate composition analyzer

What disturbs it 1 pot running out of the light component 2 loss of heater duty 3 analyzer lag and bias

Nothing is fed during the batch, so the reflux returned at the top decides how much of the falling pot reaches the product receiver on specification. Illustration, not to scale.

IN PLAIN TERMS Getting the Most Good Product From One Pot

The main idea. A batch still boils one fixed load of a liquid mixture. The lighter liquid boils off first and is collected at the top. Part of what condenses is poured back down the column, which makes the product purer but slower to come.

The challenge. The pot holds less of the light liquid as it boils, so the same setting makes weaker product as time goes on. The whole tank of product is judged once, at the end, on its average. Very pure early product can pay for weaker product later.

Why it matters. How much is poured back, minute by minute, decides how much good product one batch gives.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history at the start of Section 3.2, the fact on energy in Section 3.7, the two try-its and the fact on analyzer lag in Section 3.6, and the key points
Study	60 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, through the off-design charge), Technology in Practice (Section 3.8), Limitations (Section 3.9) and the open challenge

The control question A batch column has no operating point, and the batch is judged once, on the amount and the cumulative purity in the receiver at the end. Which reflux policy collects the most product, and why can a composition loop that holds its setpoint closely collect less than an open-loop ramp?

Learning objectives

- Derive the total-reflux separation and explain why the top composition and the process gain fall.
- Explain how the cumulative cut rule creates a purity credit.
- Compare constant reflux, constant composition and a searched ramp on product, energy and reflux demand.
- Predict what analyzer lag and bias do to the loop and to the cut.
- Assess what a trim or re-optimization recovers on an off-design charge.

Prerequisites Sections 1.4.1, 1.4.2, 1.4.3, 1.4.4 and 1.6.2. Section 1.6.3 for the predictive option only.

The app https://apmonitor.com/apps/batch-distillation/

See also Volume 9, Chapter 12 for a time budget met by arcs at bounds and Volume 9, Chapter 6 for predictive control on an economic objective. Volume 13, Chapter 8 for the energy of a continuous separation. Volume 5, Chapter 3 and Volume 12, Chapter 4 for continuous columns through a slow analyzer and an inferred composition.

3.1 Introduction

Regulatory control assumes an operating point. A setpoint is held against disturbances, a step test is read after the response settles, and performance is measured as deviation from a fixed condition [1]. A batch distillation column provides no operating point. The pot is charged once, light component leaves through the top continuously, and no stream replenishes it. The pot composition therefore falls monotonically from the start of collection to the end of the batch, and the separation becomes more difficult with no input change. In the simulator of this chapter the distillate leaves the condenser at 0.9996 mole fraction methanol and, ninety minutes later at an unchanged reflux ratio, at 0.9599. No equipment failed in between. The pot was depleted.

This behavior changes the role of the controller. There is no disturbance in the ordinary sense, because the depletion of the pot is continuous, monotonic and inherent to the operation. There is no operating point, so a tuning identified in the first ten minutes describes a column that no longer exists in the last ten. The quantity the operation is judged on is not a composition at all. It is the moles of on-specification product in the receiver at the terminal time, measured once, on the cumulative average of the whole drum. An objective that exists only at the terminal time is the defining feature of batch operation, and it is why batch plants are run on recipes and schedules rather than on loop tuning [2].

Pause. Try Exercise 3.1 before reading on.

The class is broad. Any plant whose feed is its own inventory shares this structure, including batch crystallization to a yield (Chapter 4), a batch reactor driven to a conversion target, evaporative concentration and semi-batch polymerization. The control problem in every case is the choice of an input trajectory over a fixed horizon against a cost evaluated at the end, with path constraints in the middle, and the structure of the solution is known to alternate between constraint-seeking and sensitivity-seeking arcs rather than to hold anything constant [3]. Conclusions drawn on the batch column therefore transfer across the class.

Simulators of distillation columns are usually built around the continuous case, where a feed arrives at a constant rate and the design questions are pairing, interaction and measurement lag [1], as on the continuous column of Volume 5, Chapter 3. However, a student who has only seen continuous columns carries a specific misconception into a batch plant, which is that a composition loop holding its setpoint is performing well. Constant-composition control on the column of this chapter holds the distillate within 0.0012 of 0.990 from 40 to 120 min and finishes second of the three policies, because the nearly flat controlled variable is paid for with a reflux ratio that climbs from 2.53 to its

limit of 40 and a distillate rate that falls by a factor of about twelve. A plot of the controlled variable shows almost none of this.

The column shows the principles that separate batch from continuous control (Section 1.4.1), including an economic objective evaluated only at the terminal time, a process gain that decays over the run, an actuator whose useful range collapses as the plant is depleted, and measurement lag and bias entering a decision that cannot be reversed [1, 4]. The model rests on component material balances, vapor-liquid equilibrium at constant relative volatility as the separation driving force, and a time-scale separation of three orders of magnitude between the tray holdups and the pot. Batch distillation carries most solvent recovery in fine chemicals and pharmaceuticals, so the choice of reflux policy is an industrial decision with a measurable energy and material cost [5].

This chapter compares three reflux policies used in batch operation on one browser column, with the same charge, the same duty and the same time budget behind all three. The column model, the policies and the grid search are all standard. Section 3.2 traces the policies from the simple still to trajectory optimization. Section 3.3 specifies the nineteen-state column down to its cumulative cut rule and its analyzer, with the balances, the constants and the numerics a second implementation would need. Section 3.4 shows why the top composition and the process gain decline over the run, and that the cumulative form of the product specification is what creates the optimization problem at all, since an instantaneous specification would make constant-composition control optimal by construction. Section 3.5 writes the three policies through one discrete controller and describes the grid search behind the reflux ramp. Section 3.6 measures the policies. The optimized trajectory outperforms both classical policies open loop and then loses most of the batch when the charge differs from the one it was computed for. The same section measures analyzer lag and bias, duty, holdup and volatility, the graded challenge and a shrinking-horizon predictive option. Sections 3.7 and 3.9 take up the energy and purity costs and what the model leaves out, and Section 3.8 describes the analyzer, the reflux splitter and the control system of a real batch column.

3.2 Background

The quantitative treatment of batch separation is older than the tray column, and its first balance, for a simple still in 1902, is told in the history segment below. That balance contains the central difficulty of batch separation. The charge is the state, the vapor is always richer than the charge, and integrating the balance forward gives a composition that falls monotonically. Adding trays and a reflux stream raises the enrichment achieved at each instant and does not change the direction of the trajectory.

> ⧖ HISTORY **1902: The Residue Moves One Way**
>
> In 1902 Lord Rayleigh published experiments on a question for which information seemed almost entirely lacking: how the strength of a boiling binary mixture relates to the strength of the vapor in equilibrium with it. If the curve relating the two never crosses the diagonal, the distillate is always richer than the liquid, and the residue of a limited charge moves always in one direction toward purity. If it crosses, as aqueous alcohol does at 96 %, distillation stops having any effect there. His calculation of the residue treated the vapor's composition as the change of one ingredient over the change of the whole. For very weak alcohol in water it found that halving the liquid in the retort cut its alcohol strength in the ratio 2^{11} to 1, so the residue approached pure water rapidly [6].

Two operating policies grew out of that observation and have been the practical choices ever since. In constant reflux the ratio is set once and the distillate composition is allowed to slide until the accumulated product no longer meets the specification, which is what a column with a hand-set reflux divider can do. In constant composition an operator or a controller raises the reflux ratio continuously to hold the overhead composition, which trades distillate rate for purity and requires a measurement. The comparison between the two was posed as an optimal control problem in 1963, when the reflux policy that maximizes distillate over a fixed time was derived and shown to be neither of them [7].

The complementary time-optimal problem, the policy that reaches a specified quantity of a specified purity in the least time, followed shortly afterward and returned a reflux ratio that rises throughout the run [8]. The classical result has been the same ever since. The optimal policy holds nothing constant.

Computational cost kept the optimal policy out of practice for two decades. Rigorous stage-by-stage integration of a batch column was demonstrated in 1968 with the stiffness, the holdup dynamics and the numerical difficulties laid out explicitly [9], and it was expensive enough that the optimization literature turned instead to shortcut models. The shortcut treatment models the column at each instant as a continuous column at steady state whose feed is the current pot, which reduces the dynamics to one differential equation in the pot composition and made optimal reflux policies computable for multicomponent mixtures [10]. The cost of the simplification is holdup. Shortcut models assume none, and holdup is exactly what has to fill with on-specification material before any product can be taken, so the simplification is not neutral for the quantity being optimized [11]. The argument between shortcut and rigorous models organized the field for a generation and is surveyed with its computational history in the standard treatments [12, 13]. Solving the trajectory problem directly on a rigorous model became routine only when orthogonal collocation and nonlinear programming were brought to bear on it [14], and that is the technique the reference behind this application uses.

The measurement side of the problem has its own history. Composition control on a batch column is harder than on a continuous one for reasons that have nothing to do with tuning, because the gain the loop works against changes over the run and the analyzer that reports the overhead composition is slow, so the practical arrangements infer the composition from tray temperatures rather than waiting for the analyzer measurement [15], as the propylene splitter of Volume 12, Chapter 4 does on a continuous column. This application is built directly from a published experimental study of a laboratory column that combines these elements, where the model parameters were estimated from data with nonlinear confidence intervals and a sensitivity ranking before any trajectory was optimized [16]. That apparatus is a 38 tray, vacuum-jacketed Oldershaw column with a 600 W reboiler and a swinging bucket reflux divider that can be moved once every 5 min, charged with a 50 % by weight methanol and ethanol mixture and run to a 99 mol% methanol specification. Its base case is 30 minutes at total reflux and then a constant reflux ratio of 4, and its optimized reflux schedule starts low and rises in a nominally linear pattern, raising production by 14 %. The charge, the duty, the heating efficiency, the specification, the startup and the cut rule used here are all taken from that study.

3.3 Process Model

Column T-401 is a laboratory batch column with a reboiler that holds the charge, ten theoretical trays and a total condenser. The pot is charged with 28 mol of methanol and ethanol at 0.59 mole fraction methanol, a 600 W heater boils it, and all of the vapor leaving the top is condensed. The reflux split, the fraction of that condensate returned down the column against the fraction withdrawn as product, is the only operating decision. Every stream, instrument tag and symbol the equations below use is drawn in Figure 3.1. Time is in minutes throughout, inventories in moles, flows in mol/min and compositions in mole fraction of methanol, which is the light component.

3.3.1 Controlled, Manipulated and Disturbance Variables

Table 3.1 classifies the signals. A single composition is controlled, a single reflux ratio is manipulated, and the three disturbances are all optional, because the pot depletes with no disturbance applied.

Nineteen states are integrated: the pot inventory and composition, ten tray compositions, the condenser composition, four receiver states carrying total and light-component moles for the product and slop drums separately, the analyzer, and a cumulative reboiler energy meter. Integrating the two receivers as moles rather than as running average compositions avoids a zero divided by zero at the first drop and lets both mole balances close to integrator accuracy.

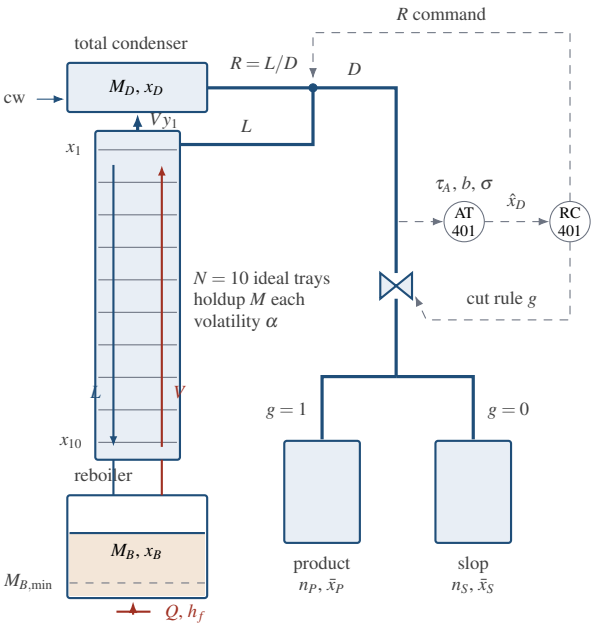


Figure 3.1 Batch column T-401 with its reboiler, ten trays, total condenser and two receivers. The reboiler holds the whole charge M_B at composition x_B and is boiled at rate V set by duty Q and heating efficiency h_f. Vapor rises through trays of holdup M against liquid L falling from above, and the total condenser holds M_D at x_D. The condensate is split at reflux ratio $R = L/D$, which is the only manipulated variable. Analyzer AT-401 reports $\hat{x}_D$ through a lag τ_A with bias b and noise σ, and RC-401 sets R and operates the routing flag g that sends the distillate to the product drum or to slop.

Table 3.1 Controlled, Manipulated, Disturbance and State Variables. Nominal Values Are Those at the Start of Product Takeoff on the Shipped Charge.

Symbol	Role	Nominal	Range
x_D (AT-401)	CV, measured	0.9996	0.90–0.999 SP
R	MV	4	0–$R_{\max}$
duty derate	DV, measured	0 %	0–60 %
b analyzer bias	DV, unmeasured	0	±0.02
σ analyzer noise	DV, unmeasured	0	0–0.01
M_B	state	27.15 mol	4–27.15 mol
x_B	state	0.5787	0–1
$x_1 \ldots x_{10}$	state	0.9993–0.7331	0–1
x_D	state	0.999 64	0–1
n_P, n_{PA}	state	0 mol	≥ 0
n_S, n_{SA}	state	0 mol	≥ 0
x_M	state	0.999 64	0–1
E	state	1080 kJ	≥ 0

3.3.2 Equilibrium and Stage Balances

Stages are ideal, so the vapor leaving a tray is in equilibrium with the liquid standing on it, and Equation (3.1) is that equilibrium written at a fixed relative volatility.

$$y_i^* = \frac{\alpha x_i}{1 + (\alpha - 1) x_i}, \qquad \frac{y^*}{1 - y^*} = \alpha \frac{x}{1 - x} \tag{3.1}$$

where α is the relative volatility, x_i is the liquid mole fraction of methanol on stage i and y_i^* is the vapor mole fraction in equilibrium with it. The second form governs the behavior of the application, because relative volatility multiplies the ratio of light to heavy and not the mole fraction itself. The enrichment available from a lean pot is therefore far smaller than the enrichment from a rich one.

Each tray is a well-mixed holdup receiving liquid from the stage above and vapor from the stage below, with constant molar overflow making L and V the same on every tray, as shown in Equation (3.2).

$$M\frac{\mathrm{d}x_i}{\mathrm{d}t} = L\,(x_{i-1} - x_i) + V\,(y^*_{i+1} - y^*_i), \qquad i = 1 \ldots N \tag{3.2}$$

where M is the liquid holdup per tray in moles, the liquid entering tray 1 arrives from the condenser so that x_{i-1} is x_D there, and the vapor entering tray N leaves the reboiler so that y^*_{N+1} is y^*_B. Both sides carry mol/min. A tray time constant is $M/V = 0.10\,\mathrm{min}$ against a batch that lasts 180 min, so the trays are very nearly quasi-steady and the slow variable is the pot.

The reboiler balances separate the batch column from the continuous case, as shown in Equations (3.3) and (3.4).

$$\frac{\mathrm{d}M_B}{\mathrm{d}t} = L - V = -D \tag{3.3}$$

$$M_B\frac{\mathrm{d}x_B}{\mathrm{d}t} = L\,(x_N - x_B) - V\,(y^*_B - x_B) \tag{3.4}$$

where M_B is the pot inventory, x_B the pot composition and y^*_B the vapor in equilibrium with it. Nothing on the right-hand side replenishes the pot, and everything leaving the top is richer in methanol than the pot is, so x_B falls monotonically for the whole batch. At a constant reflux ratio of 4 it travels from 0.5787 to 0.2337 in ninety minutes.

All the vapor condenses and both outgoing streams leave at the same composition, so the condenser balance does not contain the reflux ratio at all, as shown in Equation (3.5).

$$M_D\frac{\mathrm{d}x_D}{\mathrm{d}t} = V\,(y^*_1 - x_D), \qquad D = \frac{V}{R+1}, \qquad L = V - D \tag{3.5}$$

where M_D is the condenser and reflux drum holdup and R is the reflux ratio. Writing the distillate first and the reflux as the remainder makes $L + D$ equal V exactly, which keeps the total mole balance closed to integrator accuracy, and it accepts $R \to \infty$ for total reflux without a special case. The boil-up follows the reference energy balance in Equation (3.6).

$$V = \frac{Q\,h_f}{\Delta H_{vap}} \tag{3.6}$$

where Q is the reboiler duty expressed in kJ/min, h_f is the heating efficiency and ΔH_{vap} is the heat of vaporization. A 600 W heater, which is 36 kJ/min, at 71.9 % efficiency against 36.6 kJ/mol gives $V = 0.7072\,\mathrm{mol/min}$. Duty sets the throughput and the reflux ratio sets the split, and the two are independent, which is why losing duty costs product without costing any separation.

At total reflux $L = V$ and $D = 0$, the tray balances of (3.2) are satisfied exactly by $y^*_i = x_{i-1}$, and the ratio in (3.1) multiplies by α across each of the ten trays and again across the reboiler. The result is the Fenske relation of Equation (3.7), exact in this model rather than a shortcut.

$$\frac{x_D}{1 - x_D} = \alpha^{N+1}\,\frac{x_B}{1 - x_B} \qquad \text{at total reflux} \tag{3.7}$$

Eleven equilibrium stages at $\alpha = 2$ give a separation factor of $2^{11} = 2048$, and total reflux is the only stationary condition in the model.

3.3.3 Cumulative Cut Rule

The receiver moves through three phases in a fixed order. During the *fore cut* nothing is on specification and the distillate goes to slop. The *main cut* opens when the analyzer first reads the specification. The *tail cut* closes it when the cumulative average of everything already in the product drum falls to the specification, after which everything goes to slop. The receivers integrate as shown in Equation (3.8).

$$\frac{\mathrm{d}n_P}{\mathrm{d}t} = D\,g, \qquad \frac{\mathrm{d}n_{PA}}{\mathrm{d}t} = D\,x_D\,g, \qquad \bar{x}_P = \frac{n_{PA}}{n_P} \tag{3.8}$$
$$\frac{\mathrm{d}n_S}{\mathrm{d}t} = D\,(1-g), \qquad \frac{\mathrm{d}n_{SA}}{\mathrm{d}t} = D\,x_D\,(1-g)$$

where g is 1 during the main cut and 0 otherwise, n_P and n_{PA} are the total and light-component moles in the product drum, and $\bar{x}_P$ is the cumulative average on which the product is graded. The switching rule is Equation (3.9), and it is evaluated on the analyzer rather than on the true composition because the analyzer is the only composition information the operation has.

$$g = \begin{cases} 0 \to 1, & \hat{x}_D \geq x_{\mathrm{spec}} \quad \text{(fore to main)} \\ 1 \to 0, & \dfrac{\int D\,\hat{x}_D\,g\,\mathrm{d}t}{\int D\,g\,\mathrm{d}t} < x_{\mathrm{spec}} \quad \text{(main to tail)} \end{cases} \tag{3.9}$$

The cumulative form is the reference's own rule and it is what creates the optimization problem. Material collected above specification builds a credit, material collected below it spends the credit, and the receiver remains on specification while the credit is positive. Section 3.6 shows the searched policy finishing with an instantaneous distillate composition of 0.8285 and a receiver still at 0.9897.

3.3.4 Analyzer and Energy Meter

One online analyzer measures the distillate. Its lag is a state rather than a display filter, because the lag is large enough to change the outcome of a batch, as shown in Equation (3.10).

$$\tau_A \frac{\mathrm{d}x_M}{\mathrm{d}t} = x_D - x_M, \qquad \hat{x}_D = \mathrm{clamp}(x_M + b + n(t),\,0,\,1) \tag{3.10}$$

where x_M is the lagged composition, b is a standing bias and n is band-limited noise, formed by filtering a Gaussian sequence with a 0.5 min time constant at every controller execution and scaling the innovation so that the stationary standard deviation is σ. The bias and the noise are applied in the controller and not in (3.10), because the fourth-order integrator would otherwise sample the noise four times per step at four different values. The energy meter is Equation (3.11).

$$\frac{\mathrm{d}E}{\mathrm{d}t} = 0.06\,Q_{\mathrm{eff}}, \qquad Q_{\mathrm{eff}} = Q\left(1 - \frac{\mathrm{derate}}{100}\right) \tag{3.11}$$

where E is cumulative reboiler energy in kJ, Q_{eff} is the effective duty in watts and the coefficient 0.06 converts watts to kJ/min. The meter accrues at full duty whether or not anything leaves the top, which is why a high reflux ratio held for a long time raises the energy per mole of product.

3.3.5 Parameters, Initialization and Numerical Integration

Table 3.2 lists every constant, the plant parameters above the rule and the operator settings a scenario overrides below it. The identification applications in this series withhold the quantities a step test is meant to recover, while this application discloses every parameter, because the task is the choice of a trajectory against a plant whose parameters are given rather than the recovery of parameters from a response.

A fresh charge opens with the pot, every tray, the condenser and the analyzer all at x_0, and the pot holding $M_B(0) = M_0 - NM - M_D = 27.15$ mol. Scenario 2 instead opens on a column that has already

Table 3.2 Model Parameters and Controller Defaults With Their Adjustable Ranges

Symbol	Meaning	Value	Unit	Range
α	relative volatility	2.0	—	1.2–3.0
N	theoretical trays	10	—	fixed
M_0	charge	28	mol	12–45
x_0	charge composition	0.59	—	0.25–0.90
M	holdup per tray	0.07	mol	0.02–0.30
M_D	condenser holdup	0.15	mol	0.05–0.80
Q	reboiler duty	600	W	150–1000
h_f	heating efficiency	0.719	—	0.40–1.00
ΔH_{vap}	heat of vaporization	36.6	kJ/mol	30–45
x_{spec}	product specification	0.990	—	0.90–0.999
$D_{\min}$	production-stopped threshold	0.02	mol/min	0.002–0.1
τ_A	analyzer time constant	0.5	min	0.05–10
$R_{\max}$	maximum reflux ratio	40	—	5–100
t_{batch}	batch time budget	180	min	30–300
$M_{B,\min}$	minimum pot inventory	4	mol	0.5–15
t_{start}	total-reflux startup	30	min	0–90
$x_{D,\text{SP}}$	composition setpoint	0.990	—	0.90–0.999
R_{set}	constant reflux ratio	4.0	—	0–$R_{\max}$
R_0	hand-set ramp start	2.0	—	0–20
R_1	hand-set ramp end	10.0	—	0–$R_{\max}$
K_c	proportional gain	600	—	0–1500
τ_I	integral time	4.0	min	0.2–60
τ_D	derivative time	0	min	0–5
τ_f	derivative filter	0.2	min	0–2

been brought up at total reflux. That state is constructed from (3.7) by bisecting the light-component balance of Equation (3.12) for the pot composition, which is monotone in x_B.

$$M_B x_B + M \sum_{i=1}^{N} x_i(x_B) + M_D x_D(x_B) = M_0 x_0 \tag{3.12}$$

The result is $x_B = 0.5787$, leaner than the 0.59 charged, because the 0.85 mol of holdup in the trays and the condenser has filled with much richer material and that methanol came out of the pot. The corresponding condenser composition is 0.999 64.

The engine of Section 1.6.1 integrates the column with classical fourth-order Runge-Kutta on a fixed 0.01 min step. RC-401 executes every 0.05 min, which is 3 s of plant time, and the historian keeps ten samples per simulated minute. Noise comes from a generator seeded at 20260829, so a noisy batch is exactly reproducible. The speed setting counts simulated minutes per second of wall time, and at the default of ten a full 180 min batch completes in about eighteen seconds. Every discontinuous event, the end of the startup, the switch between slop and product, and the end of the batch, is implemented in the controller and reaches (3.2) to (3.8) only as R and g, both constant across a step, so the four Runge-Kutta stages never straddle an event.

3.3.6 Assumptions

Six simplifying assumptions define the model, and the condition under which each fails is stated with it.

Constant relative volatility, and not the reference's own value. The reference computes equilibrium from a quartic fit to a physical property database. Across the whole composition range that quartic is a constant relative volatility of 1.725 to an RMS deviation of 4.2×10^{-4} mole fraction, so nothing is lost by using (3.1) instead. What is changed deliberately is the number itself, and Section 3.5.4

gives the reasoning and the cost. Any mixture with a strongly curved equilibrium line is outside a constant-volatility model, and every azeotropic mixture is outside it completely, because the relative volatility passes through 1 at the azeotrope and no number of stages separates anything there.

Constant molar overflow. The internal flows are the same on every tray and there is no energy balance inside the column. Methanol and ethanol differ by about 9 % in molar heat of vaporization, which makes the assumption reasonable here and would not make it reasonable for a water and acetic acid split. The reference validates the same simplification against a detailed model carrying full tray energy balances and finds the difference immaterial for optimization [16].

Ideal equilibrium stages, ten of them. Stage efficiency falls with vapor rate, with fouling and with foaming, and none of that is modeled, so a column that has been in service will not reach these compositions.

Constant liquid holdup and no vapor holdup. Holdup is 0.85 mol against a 28 mol charge, which is 3.0 %, and it is fixed. Holdup is the refinement that separates a rigorous batch model from a shortcut one and it is not a small effect [11]. With no total-reflux startup, raising it from 3.0 % to 10.7 % of the charge lengthens the fore cut from 6.2 min to 21.2 min and costs 2.75 mol of product.

Boil-up follows the duty algebraically. Equation (3.6) has no thermal capacitance, the heat of vaporization is fixed, and the column pressure does not move with the vapor rate. A real reboiler has a thermal time constant of several minutes, so a duty change here appears instantly and the profile answers it within a tray time constant.

The reflux ratio is a continuous variable settable to anything from 0 to 40. The reference apparatus sets it with a swinging bucket that can only be moved every 5 min, which is why its optimizer returns a staircase rather than a smooth ramp. A column with a real bucket cannot execute a constant-composition policy at all, and that is much of why constant reflux is the classical policy.

3.4 Dynamic Analysis

3.4.1 Decline of the Top Composition and the Process Gain

The behavior that organizes the application follows from combining (3.7) with (3.4). The column multiplies the pot ratio by a factor that never changes, and the pot ratio only ever falls, so the top composition falls with it. Over ninety minutes at a constant reflux ratio of 4 the pot goes from 0.5787 to 0.2337, a ratio falling from 1.374 to 0.305 by a factor of 4.5, and the distillate follows from 0.9996 to 0.9599. No disturbance entered and no input changed.

Two consequences follow for the controller. The process gain $\partial x_D/\partial R$ decays toward zero over the run, because raising the reflux ratio buys composition by giving up distillate rate, and as the pot gets leaner the same increment buys less and costs more. And the useful actuator range collapses at the same time, since $D = V/(R+1)$ in (3.5) means the distillate rate falls hyperbolically with R. At $R = 3$ the column makes 0.177 mol/min and at $R = 39$ it makes 0.018 mol/min, so the last three quarters of the actuator's travel buy almost no product at all.

3.4.2 Reflux Demand Under Constant-Composition Control

A constant-composition controller must therefore supply an ever larger reflux ratio to hold a fixed distillate composition, and the demand it faces is set by the plant rather than by the tuning. Figure 3.2 plots the reflux ratio each of the three policies runs against time on the same charge.

The two curves that meet the same specification are separated by a factor of almost four in the reflux ratio at the end of the batch. The constant-composition demand rises without bound because the plant is running out of light component and the loop is being asked for a composition the pot can no longer support at any useful rate. That is not a tuning fault and no tuning removes it.

3.4.3 Effect of the Cumulative Cut Rule on Optimality

The optimization exists because of (3.9) and would not exist without it. Suppose the receiver were switched to slop the moment the instantaneous composition fell below x_{spec}. Then every mole

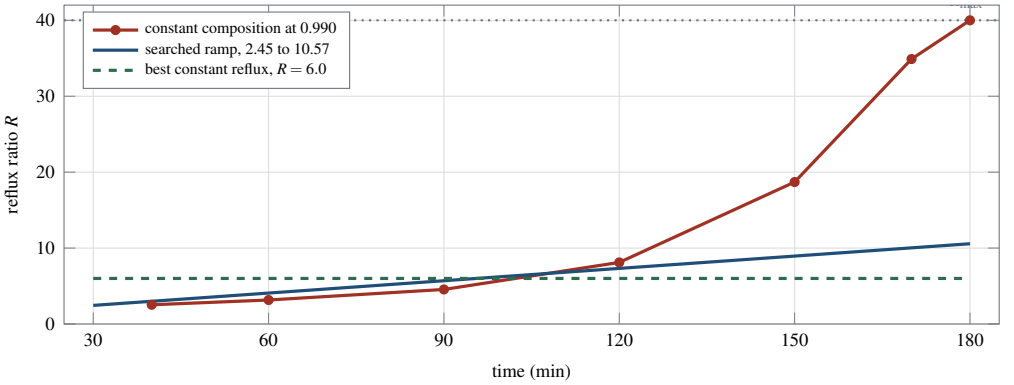


Figure 3.2 Reflux ratio against time for the three policies on the same 28 mol charge and the same 180 min budget: constant composition (solid with dots), the searched ramp (solid) and the best constant reflux ratio (dashed). The constant-composition points are measured on the running application at the shipped tuning. Holding 0.990 costs 2.53 at 40 min and the full 40 by 180 min, while the searched ramp that collects more product never asks for more than 10.57.

collected would be at or above specification, no reflux would be spent making anything purer than the customer pays for, and the best possible policy would be to hold x_D exactly at x_{spec} for as long as possible. That policy is constant-composition control at $x_{D,\text{SP}} = x_{\text{spec}}$, optimal by construction, and there would be nothing left to search for.

The cumulative rule creates a currency instead. Define the purity credit as the integral in Equation (3.13), taken over the main cut.

$$C(t) = \int_{t_{\text{main}}}^{t} D\,\left(x_D - x_{\text{spec}}\right)\,\mathrm{d}\tau \tag{3.13}$$

The receiver is on specification exactly while $C(t) \geq 0$, since $\bar{x}_P - x_{\text{spec}} = C/n_P$, and the tail cut of (3.9) fires when the credit reaches zero. In the terms of Section 1.4.4, the receiver purity is an endpoint constraint and the instantaneous composition carries no path constraint. A policy is therefore free to run the instantaneous composition below specification for as long as earlier material pays for it, and the best policy spends the credit down to exactly zero at the end of the batch. That is what the searched ramp does, and it is why it beats a loop that holds x_D near 0.990 for most of the collection.

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

3.5 Controller Design

3.5.1 Discrete PID Algorithm

RC-401 runs the velocity-form algorithm of Section 1.6.2, computing an increment to the reflux ratio each scan rather than the ratio itself. Its derivative term acts on the analyzer reading and not on the error, so a composition setpoint change produces no derivative kick at the reflux divider, as shown in Equations (3.14) and (3.15).

$$e_k = x_{D,\text{SP}} - \hat{x}_{D,k} \tag{3.14}$$

$$\begin{aligned}
\Delta R_k^P &= K_c\,(e_k - e_{k-1}) \\
\Delta R_k^I &= \frac{K_c}{\tau_I}\,e_k\,\Delta t_c \\
\Delta R_k^D &= -K_c\,\tau_D\,\frac{\hat{x}_k^f - 2\hat{x}_{k-1}^f + \hat{x}_{k-2}^f}{\Delta t_c}
\end{aligned} \tag{3.15}$$

Table 3.3 The Three Reflux Policies, Expressed Through the Same Controller

Policy	R_{ff}	K_c	τ_I
Constant reflux	—	output discarded, PID tracks	
Constant composition	0	600	4 min
Optimized ramp	Eq. (3.17)	0	∞

The controller is reverse acting because the process gain is positive. More reflux gives a richer distillate, so a measurement below setpoint, which is a positive error in (3.14), calls for a larger reflux ratio. The derivative increment is a second difference on a filtered measurement, with $\hat{x}_k^f = \alpha_f \hat{x}_{k-1}^f + (1-\alpha_f)\hat{x}_{D,k}$ and $\alpha_f = \exp(-\Delta t_c/\tau_f)$, where Δt_c is the controller period and the filter constant τ_f defaults to 0.2 min. The three increments are combined and the demand is clipped to the divider span [17], as shown in Equation (3.16).

$$\begin{aligned} s_k &= s_{k-1} + \Delta R_k^P + \Delta R_k^D \\ c_k &= R_{\mathrm{ff}}(t_k) + s_k + I_{k-1} \\ R_k &= \min(R_{\max},\ \max(0,\ c_k)) \\ I_k &= I_{k-1} + \Delta R_k^I + \frac{\Delta t_c}{\tau_I}\,(R_k - c_k) \end{aligned} \tag{3.16}$$

These are Equations (1.14) to (1.17) written for this loop. The running sum s_k is S_k, the demanded ratio c_k is v_k, the trajectory R_{ff} is the output bias u_b, and the back-calculation that corrects the integral alone has the tracking time $T_t = \tau_I$. Under the trajectory policy τ_I is infinite, so the integral increment and the back-calculation are both absent, and I carries only an offset left by a transfer, which decays with a 2 min time constant. The output is clamped to $0 \le R_k \le R_{\max}$ and carries no rate limit, because a reflux divider changes its split in one step rather than stroking through the intermediate positions. Execution is every 0.05 min, independent of the 0.01 min physics step. Transfers between MANUAL and AUTO are bumpless in the way Section 1.6.2 describes. The tuning is bounded at $0 \le K_c \le 1500$, $0.2\,\mathrm{min} \le \tau_I \le 60\,\mathrm{min}$ and $0 \le \tau_D \le 5\,\mathrm{min}$, with K_c in units of reflux ratio per unit of mole fraction.

Before any of that runs, the startup latch overrides it. For $t < t_{\mathrm{start}}$ the plant is given $R = \infty$ rather than $R_{\max}$, so that (3.5) returns $D = 0$ exactly and the pot does not drain during the startup at all. Behind the latch the loop is open, so in AUTO the controller is not updated there but tracks, as in MANUAL, the ratio its policy will open on, which is R_{set} under constant reflux, R_0 under the trajectory and the operator's reflux station under constant composition. Collection therefore opens bumplessly from that ratio, and the error standing at the opening enters through the integral, or under the trajectory through the transfer offset that decays with the 2 min time constant, rather than as a step. Section 3.6 measures the opening on the shipped analyzer.

3.5.2 The Three Reflux Policies

The three policies differ only in what R_{ff} and τ_I are, as listed in Table 3.3. The plant does not change between them and neither does the actuator.

Under constant reflux the controller output is discarded and the plant is given the operator's constant, while the PID keeps tracking so that selecting another policy later starts from the reflux ratio the column already has. Under constant composition R_{ff} is zero and the loop is an ordinary PI controller on the analyzer. Under the optimized policy R_{ff} is the searched trajectory of Equation (3.17).

$$R_{\mathrm{ff}}(t) = \mathrm{clamp}\left(R_0 + (R_1 - R_0)\,\frac{t - t_{\mathrm{start}}}{t_{\mathrm{batch}} - t_{\mathrm{start}}},\ 0,\ R_{\max}\right) \tag{3.17}$$

where the fraction is itself clamped to $[0, 1]$ so the ramp is held flat outside the collection window. Two details of that arrangement are structural rather than incidental. The trajectory enters as the controller

bias rather than being added on afterward, as the kettle's feedforward does in Section 1.4.3, which puts it inside the clamp and inside the manual tracker, and a gain of zero then gives the pure open-loop trajectory exactly. And reset is switched off, because an integral term would drive the composition onto its setpoint within a few minutes and erase the trajectory, reproducing constant-composition control. Any gain above zero is a proportional trim around the trajectory, and Section 3.6 measures what that trim is worth.

3.5.3 Grid Search for the Reflux Ramp

Selecting the searched ramp runs a forward-simulation grid search at every reset, before the first frame is drawn. Each candidate is integrated with explicit Euler at 0.02 min, a fifth of the tray time constant and a twenty-fifth of the analyzer's, on the same derivative function the plant uses, and reproduces the engine's Runge-Kutta run to within 0.01 mol of product. The search routes the distillate with the same lagged and biased analyzer the operator has, without its noise, because the lag delays the tail cut and changes the collected product.

The structure of the search follows from the objective not being smooth. The tail cut is a discrete event, so a ramp that just avoids taking one collects every mole it made while a ramp a tenth of a unit lower throws the last half hour away. The search therefore runs a diagonal sweep of every constant reflux ratio from 0.5 to 16 in steps of 0.5, which guarantees the result is never worse than the best constant policy on that grid, then a coarse two-dimensional sweep with R_0 from 0.5 to 8 and R_1 from 1.5 to 26 that skips every falling ramp, because the optimal reflux ratio does not decrease over a run in which the separation becomes monotonically harder, then a refinement around the best three candidates and a finer one around the best two, rather than around the single best point. Refining around one candidate instead costs nothing on the shipped case and 0.54 mol at a 0.995 specification, where the best flat ramp and the best rising ramp sit far apart and only one of them is near the coarse grid's winner. The whole search is deterministic, so a reset reproduces the run exactly.

On the shipped charge the search returns $R_0 = 2.45$ and $R_1 = 10.57$. The reference's own optimizer is a nonlinear program that re-picks the reflux ratio every five minutes because its apparatus can only be changed that often, and the trajectory it finds starts low and increases in a nominally linear pattern [16]. A two-parameter straight ramp is a continuous form of that trajectory.

3.5.4 Relative Volatility and Stage Count

One departure from the reference is deliberate and is documented here. The application runs at $\alpha = 2.0$ where the methanol and ethanol chemistry is 1.725. The reason is not the thermodynamics, which the constant-volatility form reproduces to an RMS of 4.2×10^{-4} mole fraction, but the apparatus. The reference column has 38 Oldershaw trays at a fitted Murphree efficiency between 0.37 and 0.45, which is roughly 14 to 17 theoretical stages against the eleven modeled here. Eleven stages at 1.725 give a total-reflux separation factor of $1.725^{11} = 402$ where eleven stages at 2.0 give $2^{11} = 2048$, a factor of 5.1 between them, and the difference at finite reflux is far larger than that. At 1.725 the ten-tray column cannot run the reference's own policy at all, since a constant reflux ratio of 4 collects 1.38 mol before it has to take a tail cut. At 2.0 it reproduces the reference base case, collecting 12.73 mol at a cumulative 0.9909 over the reference's 30 plus 90 minute schedule against the 9.5 mol at 0.992 the reference model reports. Volatility and stage count trade against each other, and here the trade was made in favor of a state vector of 19. The value 1.725 is available as a parameter setting and scenario 13 measures what it costs, so the effect of the choice is measurable inside the application.

One arithmetic correction belongs with this. The quartic equilibrium fit printed in the reference carries a coefficient of -2.016 on the fourth-order term. With that coefficient the expression returns -0.250 at $x = 0.9$ and -0.815 at $x = 1$, which are not compositions. With -0.2016 the same quartic is the constant relative volatility form of (3.1) at $\alpha = 1.725$ to an RMS deviation of 4.2×10^{-4} and a maximum deviation of 8.7×10^{-4} over $0 \le x \le 1$, and it returns 0.9999 at $x = 1$. The printed coefficient is a typographical error of one decimal place, and the model file distributed with the

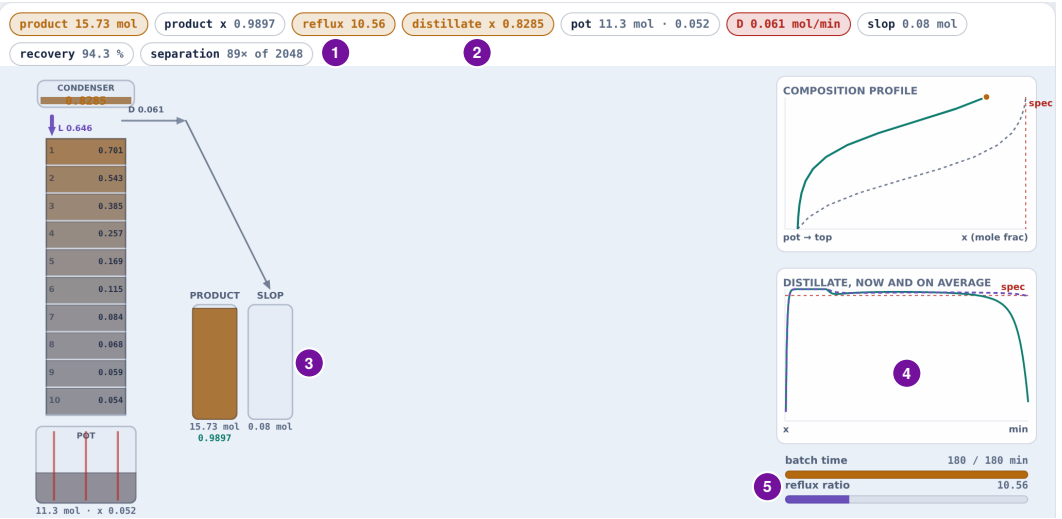


Figure 3.3 The searched reflux ramp at the end of its batch, scenario 6 at $t = 180$ min, cropped to the plant view. (1) The badges: the batch collects 15.73 mol at a cumulative composition of 0.9897, with 0.08 mol of slop and 94.3 % recovery, at 412 kJ/mol (Table 3.5), and the pot is drawn down to 11.3 mol at 0.052 mole fraction methanol. (2) The instantaneous distillate composition ends at 0.8285. (3) The product and slop receivers. (4) The distillate now and on average: the instantaneous trace falls away at the end while the average holds near the specification. (5) The batch-time and reflux-ratio bars.

reference does not use the quartic at all, computing equilibrium from component vapor pressures instead.

3.6 Results

The fifteen scenarios are listed in Table 3.4, ordered by concept rather than by difficulty. Each carries a starting mode, a reflux policy, a tuning, the parameters it overrides, the faults it fires and their timing, its own prediction, experiment, hints and explanation, and a completion rule, as Section 1.6.4 describes for every app. The Exercises at the end of the chapter use them.

All computation, the reflux search included, runs in the browser on the engine of Section 1.6.1, so a slow machine draws fewer frames and integrates the same batch. Every run exports at ten samples per simulated minute, carrying the true composition beside the analyzer reading, the cut phase, both mole balances and every parameter in force, which is enough to reconstruct the purity-credit integral offline or to audit a batch. Twenty-nine verification tests, twelve of them specific to this application, run against the application and every one passes, with both mole balances closing to better than 2×10^{-11} mol at the end of each batch of Table 3.5 against a 1×10^{-9} mol test tolerance. The receiver is graded on specification when its cumulative average is within 0.001 of x_{spec}, because the tail cut closes on the analyzer's running average, which lags a falling distillate, so a correctly run batch closes a few ten-thousandths below the specification. Figure 3.3 is the last frame of the optimized batch of scenario 6.

3.6.1 Comparison of the Three Policies

The central comparison is Table 3.5. All three policies run the same 28 mol charge at 0.59, the same 600 W duty, the same 30 min total-reflux startup and the same 180 min budget against the same 0.990 specification, and every figure is read from the application at the end of the batch. Figure 3.4 sets the product, the energy per mole and the final reflux ratio of each policy side by side.

Table 3.4 The Fifteen Scenarios and the Behavior Each One Shows

#	Scenario	What it shows
1	Total reflux, MANUAL	holdup filling at startup, Fenske
2	The settled column	the total-reflux equilibrium
3	Constant $R = 4$, 120 min	the reference base case
4	Constant $R = 3$ then a sweep	location of the constant-reflux optimum
5	Constant composition at 0.990	actuator saturation with the CV near setpoint
6	Searched reflux ramp	open-loop trajectory, purity credit
7	Hand-set ramp, 2 to 16	purity giveaway
8	Analyzer $\tau_A = 5$ min	lag in a loop and in a cut decision
9	Analyzer $+0.012$ at $t = 34$	bias that improves reported results
10	-35 % duty at $t = 60$	throughput against separation
11	Charge at $x_0 = 0.42$	open-loop policy on an off-design charge
12	Holdup 3.0 mol, no startup	purpose of the total-reflux startup
13	$\alpha = 1.725$	volatility against stage count
14	Replan at $\alpha = 1.85$	nominal predictive feedback
15	Graded challenge	control on a slow, noisy analyzer

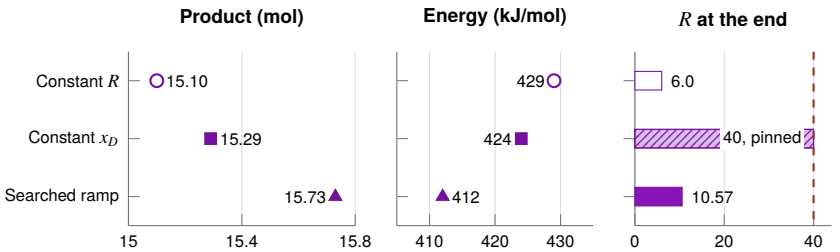


Figure 3.4 The three policies of Table 3.5 on the same charge, duty and 180 min budget: product collected and energy per mole of product (axes not starting at zero), and the reflux ratio at the end against its limit of 40 (dashed). Notice that the policy that collects most and spends least per mole ends at a reflux ratio of 10.57, while the composition loop finishes second with its actuator pinned.

Table 3.5 Terminal Results of a 180 min Batch Against a 0.990 Specification. The Constant Reflux Ratio Is the Best Flat Value Found by Sweeping in Steps of 0.05.

	Constant R $R = 6.0$	Constant x_D SP = 0.990	Searched ramp 2.45 to 10.57
Product (mol)	15.10	15.29	**15.73**
Receiver $\bar{x}_P$	0.9895	0.9899	0.9897
Recovery (%)	90.4	91.6	**94.3**
Energy (kJ/mol)	429	424	**412**
R at the end	6.0	**40, pinned**	10.57
Feedback	none	PI on $\hat{x}_D$	none

The optimized trajectory collects 15.73 mol against 15.29 mol for constant composition and 15.10 mol for the best constant reflux ratio, an improvement of 2.9 % over the loop and 4.2 % over the flat policy. It does so *open loop*, with a proportional gain of zero and a reflux ratio that never exceeds 10.57, while the loop that holds its setpoint within 0.0012 for eighty minutes finishes second with its actuator hard against a limit of 40. The reference reports a 14 % improvement from optimization on its own apparatus, and the difference between 14 % and 4.2 % is the baseline rather than the method, since the reference compares against a constant reflux ratio of 4 on a 90 minute collection and the

comparison here is against the best constant reflux ratio on a budget long enough for the trade to develop. Against a constant reflux ratio of 4 on the same 180 min budget the searched ramp collects 15.73 mol against 13.16 mol, an improvement of 19.6 %.

3.6.2 Sensitivity of the Constant-Reflux Optimum

Sweeping the constant reflux ratio gives 2.17 mol at $R = 2$, 9.16 mol at 3, 13.16 mol at 4, 14.48 mol at 5, 15.10 mol at 6, 14.14 mol at 6.5 and 11.79 mol at 8. Two different failures bound the curve. Below $R = 6$ the tail cut fires before the batch ends and the remaining production is routed to slop, with 13.99 mol of slop at $R = 3$ and a tail cut at 81.8 min. Above $R = 6$ the batch ends with the receiver above specification, and the surplus purity has no value against a 0.990 specification. The best constant reflux ratio is therefore the one whose receiver reaches specification at the moment the batch ends, so the answer moves when the budget moves and is 3.9 on the reference's own 120 min schedule rather than 6, collecting 12.95 mol against 12.73 mol at 4.

Scenario 7 quantifies the same effect on a trajectory. A plausible hand-set ramp of 2 rising to 16 collects 13.15 mol at 0.9957, giving away 16.4 % of the production for 0.0057 of purity above the specification. A hand-set 2.5 rising to 10.5, within a tenth of the searched answer, collects 15.74 mol, 0.006 mol more than the search, because that point lies between the nodes of the search grid. The value of the search is that the optimum moves whenever the charge, the duty, the specification or the batch time moves.

3.6.3 Actuator Saturation Under Constant-Composition Control

The constant-composition policy holds the composition close to setpoint for most of the collection and fails in the actuator. Collection opens on the operator's station at $R = 2$, as described in Section 3.5.1, the loop trims the ratio to 0.88 and back up to 4.66 within three minutes, and the distillate dips only to 0.9896 at 32.5 min. From 40 to 120 min the distillate stays within 0.0012 of 0.990. The reflux ratio the loop demands runs 2.53 at 40 min, 3.16 at 60, 4.55 at 90, 8.11 at 120, 18.7 at 150 and 34.9 at 170, and is hard against 40 from 174.4 min, as plotted in Figure 3.2. The distillate rate falls with it from 0.200 mol/min at 40 min to 0.017 mol/min, a factor of about twelve, and falls below the $D_{\min}$ threshold at 169.5 min.

The demand is a ramp that steepens, and a PI loop follows a ramp with a lag error of about $(\mathrm{d}R/\mathrm{d}t)\,\tau_I/K_c$ in mole fraction. At 150 min the demand rises by 0.58 per minute, which predicts 0.0038 against the 0.0034 measured, and the distillate falls behind the setpoint to 0.9866 at 150 min and 0.9834 at 170 min, and the running average closes the receiver with a tail cut at 162.7 min, leaving 0.34 mol in slop. The policy still performs well. It beats the best constant reflux ratio by 1.3 % because it never over-purifies, and it needs no trajectory and no model. It requires an analyzer and a tuning fast enough to follow a reflux demand that rises throughout the batch. At $K_c = 600$ the loop keeps up until the last hour. At $K_c = 300$ the lag error doubles, the loop takes a tail cut at 119.7 min and collects 13.33 mol, a 12.8 % loss from a tuning change that is barely visible on the controlled variable.

3.6.4 Purity Credit Trajectory of the Optimized Ramp

The searched policy starts at 2.45, near the minimum reflux the column needs, where the distillate rate is high and the pot is still rich enough that the material leaves at 0.9923 at 40 min. That interval builds the purity credit of (3.13), and the remainder of the run spends it. The instantaneous composition falls to 0.9870 at 150 min, 0.9490 at 170 and 0.8285 at the end, while the cumulative average in the receiver is still 0.9897. A policy that never let the instantaneous composition go below specification could not do this and would collect less. The gap between the instantaneous and cumulative traces is the mechanism, and the two compositions are displayed together for that reason.

Table 3.6 Constant-Composition Control at $K_c = 700$ and $\tau_I = 2$ min Against Analyzer Lag, With Noise at Zero. Every Extra Minute of Lag Costs Product, and at 8 min the Dip After the Opening Forces a Tail Cut at 39.9 min.

τ_A (min)	0.5	1	2	3	5	8
Product (mol)	15.51	15.46	15.16	14.61	12.70	1.66
Receiver $\bar{x}_P$	0.9901	0.9902	0.9896	0.9880	0.9834	0.9856

TRY IT IN THE APP **6 · The Optimized Reflux Ramp**

Load the scenario and press Run. Watch the inset of the distillate, now and on average. The instantaneous composition falls to 0.9490 at 170 min and 0.8285 at the end, while the receiver's cumulative average stays at 0.9897. The early, purer product pays for the late, weaker product.

3.6.5 Analyzer Lag and Bias

The analyzer serves two functions and only one of them is feedback. Its effect on the loop is visible. At $\tau_A = 5$ min with a tight tuning of $K_c = 700$ and $\tau_I = 2$ min and the scenario's 0.002 of analyzer noise, the loop drives the ratio from the station's 4 to zero within two minutes of the opening, the distillate falls to 0.884 before the analyzer has seen it, and from then on the ratio hunts from zero to between 25 and 40 in every half hour. Detuning to $K_c = 300$ with $\tau_I = 4$ min stops the hunt, the ratio climbing smoothly between 1.7 and 7.3 from 60 to 110 min. Without the noise the tight tuning still swings by 28.8 over the last seventy minutes. Its effect on the cut is not visible at all, because the tail cut of (3.9) is judged on the same instrument, and while the true composition is falling the analyzer reads high. Table 3.6 measures both failures on noise-free runs at the tight tuning.

At 5 min of lag in the shipped scenario the tail cut never fires at all and the receiver finishes with 13.09 mol at 0.9820, off specification, while the running average computed from the analyzer reads 0.9937. Detuning at the same lag removes the hunt and not the bias. It returns 15.36 mol at 0.9888, still off specification after a tail cut at 159.1 min judged on the same lagging reading. The lag itself is untouched and no tuning touches it.

DID YOU KNOW? **Eight Minutes of Lag Lose the Batch**

From the simulator: under constant-composition control at the tight tuning, an analyzer lag of 0.5 min collects 15.51 mol, 3 min collects 14.61 mol, and 8 min only 1.66 mol (Table 3.6). The cut that ends the collection is judged on the same late instrument.

A bias is more damaging than a lag because it improves the reported result. Drifting the analyzer 0.012 high four minutes into the collection leaves the receiver at 0.9783, which is 0.0117 low and almost exactly the bias. No trend drawn from the analyzer reveals it, because each carries the bias. The controller holds the biased reading within 0.007 of setpoint until 120 min, and the running average computed from the analyzer still reads 0.9907 at 150 min against a true 0.9790. The batch also reports better results than the unbiased one, at 16.33 mol against 15.29 mol and 96.7 % recovery against 91.6 %. Both improvements are fictitious. The extra 1.04 mol is material the cut rule should have sent to slop and the whole drum is now rework. Two measurements reveal the fault, since the true distillate composition falls to 0.8935 by the end and production stops at 151.2 min, eighteen minutes earlier than the unbiased run, with the reflux ratio at its limit of 40 from 154.9 min. Feedback cannot correct a sensor bias, because the controller works to make the biased reading true, so every defense is external to the loop. Both mole balances are recorded at every sample for that reason, and they stay below 1×10^{-9} mol of error even while the analyzer reading is biased.

Table 3.7 A Proportional Trim Around the Wrong Trajectory, on a 0.42 Charge. Re-Optimizing the Trajectory Recovers 10.69 mol on the Same Plant.

K_c	0	100	300	600	1000
Product (mol)	1.26	1.37	1.98	3.10	4.74
Receiver $\bar{x}_P$	0.9882	0.9888	0.9893	0.9896	0.9897
Slop (mol)	14.55	10.53	9.45	8.07	6.24
Recovery (%)	10.6	11.5	16.6	26.1	39.9

TRY IT IN THE APP 9 · A Biased Analyzer Quietly Makes Off-Specification Product

Load the scenario and press Run. Four minutes into the collection the analyzer drifts 0.012 high. The batch then reports 16.33 mol and 96.7 % recovery, better than the unbiased run, while the true receiver ends at 0.9783, off specification. Watch the reflux ratio reach its limit of 40 earlier than in scenario 5.

3.6.6 The Graded Challenge

Scenario 15 grades constant-composition control on an analyzer with a 1 min lag and added noise. It asks for the true distillate within 0.008 of 0.990 from 86 to 130 min and a receiver on specification within 0.001, with the model parameters and disturbances as given. With the shipped tuning, $K_c = 40$ and $\tau_I = 30$ min, the worst error from 86 to 130 min is 0.1198, about fifteen times the band. The receiver closes at 0.9893 with 3.98 mol and 13.31 mol in slop, so receiver quality passes and tracking fails. With the suggested tuning, $K_c = 500$, $\tau_I = 3$ min, $\tau_D = 0.4$ min and $\tau_f = 0.5$ min, the worst error is 0.0047 and the receiver closes at 0.9897 with 15.61 mol and no slop, about 0.0007 above its true-composition minimum of 0.989. PI at the same gain and reset also passes with a worst error of 0.0039, so the derivative is optional. Removing its filter raises the reflux-ratio travel from 997 to 4810, while a gain of 700 with a two-minute reset fails tracking at 0.0328 despite an on-specification receiver. Reset with the suggested settings reproduces the batch.

3.6.7 Sensitivity to an Off-Design Charge

An open-loop trajectory carries no information about the plant it is running on, which is the standing objection to trajectory optimization and the reason the reference spends most of its length on parameter estimation before it optimizes anything. Scenario 11 measures the objection in moles. The ramp searched for a 0.59 charge is applied to a 0.42 charge. Eleven stages multiply a ratio, and a 0.42 charge has a ratio of 0.72 where 0.59 has 1.44, so the top of the column starts a factor of two lower and 2.45 is no longer enough reflux to make specification. The batch collects 1.26 mol at 0.9882, sends 14.55 mol to slop, takes its tail cut at 36.5 min and consumes 5143 kJ/mol against 412 kJ/mol on the design charge. Re-searched against the charge it actually has, the same policy structure returns 3.40 rising to 17.83 and collects 10.69 mol at 0.9898 with 0.004 mol of slop.

Because the trajectory enters as the bias in (3.16), a proportional trim can be added around it without changing anything else, and Table 3.7 measures how much of the loss that recovers.

At $K_c = 1000$ the proportional term drives the reflux ratio to its limit of 40, holds the true distillate at 0.9562 at the end instead of 0.0440, and recovers 4.74 mol. Re-optimizing recovers 10.69 mol. More gain recovers more, because the tail cut moves from 36.5 min open loop to 37.6 min, 42.1 min, 50.8 min and 65.0 min at $K_c = 100$, 300, 600 and 1000, while part of the fall in slop comes only from the reflux ratio later climbing to 40 and the column ceasing to produce. Feedback around a wrong trajectory recovers part of the loss, and a correct trajectory recovers nearly all of it, which is why a batch operation measures its charge before executing a recipe. On the correct charge the same trim is close to free, since $K_c = 100$ collects 15.78 mol at 0.9898 against 15.73 mol open loop while pushing the reflux ratio to 20.6 by the end and putting 0.19 mol in slop against 0.08 mol, and $K_c = 300$ collects 15.73 mol. The trim protects against an off-design charge at a measurable cost in reflux and slop on the design charge.

3.6.8 Effect of Duty, Holdup and Volatility

Three parameters change the outcome in ways that resemble control faults. Losing 35 % of the reboiler duty at 60 min with the reflux ratio held at 6 drops the vapor rate from 0.707 mol/min to 0.460 mol/min and drops L and D with it, so L/D is still 6 and the column separates exactly as well as before on less throughput. Production falls from 15.10 mol to 10.91 mol and recovery from 90.4 % to 65.9 %. The loss is throughput, because the same batch time now passes 35 % less vapor, and the diagnostic is that the product comes out *purer* at 0.9970 against 0.9895. A batch column that ships material well above specification with no setpoint change is running a reflux ratio too high for its remaining throughput.

🕮 **Pause.** Try Exercise 3.2 before reading on.

The total-reflux startup exists to fill the holdup. With no startup at all at $R = 4$, the fore cut runs 6.2 min at the shipped 0.85 mol of holdup and 21.2 min at 3.00 mol, and the product falls from 12.15 mol to 9.40 mol, with $M_B(0)$ following the holdup whether it is set before Run or carried through Reset. Switching 30 min of total reflux back on erases the penalty completely, at 13.17 mol with 3.00 mol of holdup against 13.16 mol with 0.85 mol, because nothing is being thrown away while the holdup fills. That also sets how long a startup should be. At $R = 6$ on the shipped 0.85 mol column, no startup collects 14.59 mol with 3.60 mol of slop, ten minutes collects 15.10 mol with 2.08 mol, thirty minutes collects 15.10 mol with 0.06 mol, and beyond that the startup only consumes batch time, dropping to 13.64 mol at forty-five minutes and 12.12 mol at sixty. The best startup is the shortest one that leaves nothing in the slop drum.

Relative volatility decides whether the separation is possible at all. At $\alpha = 1.725$ on the same ten trays, a constant reflux ratio of 4 collects 1.38 mol and sends 19.84 mol to slop against 13.16 mol collected at 2.0, so a 14 % reduction in relative volatility costs 90 % of the production. The column remains capable but under-refluxed, because $R = 8$ collects 10.89 mol at 0.9899 and the best constant reflux ratio in steps of 0.1, 8.4, collects 11.29 mol at 0.9900. Eleven stages still make 99 mol% methanol at the reference chemistry, but only by sacrificing most of the distillate rate, and even at its best the column produces 25 % less than at $\alpha = 2.0$.

3.6.9 Replanning With an Imperfect Nominal Separator

The MPC option of Section 1.6.3 recalculates reflux every half minute, rather than preserving a ramp searched at reset. Its collection horizon ends at the current batch clock. It uses at most 150 uniformly spaced prediction intervals and twelve graduated move blocks, so the prediction spacing may exceed the decision period. Before takeoff the window spans startup end to batch end, and the first optimization at $t = 0$ describes hypothetical collection from the current estimate. Physical total reflux continues unchanged. This is a local economic MPC [18], not the reference's full nonlinear batch optimization.

The reduced model eliminates the ten tray and condenser transients. For a trial steady overhead x_D^s, its operating line marches downward:

$$y_1 = x_D^s, \quad x_i = \frac{y_i}{2 - y_i}, \quad y_{i+1} = \frac{R x_i + x_D^s}{R + 1}, \tag{3.18}$$

with the shooting condition $y_{11} = 2x_B/(1 + x_B)$. Thus relative volatility stays at two even when the simulated chemistry changes. The fixed boil-up is $V_0 = 0.707213$ mol/min. With $F = x_D^s + b$, confined to physical mole fractions, the four-state nominal balance is

$$\dot{M} = -D_0, \qquad \dot{A} = -D_0 F, \qquad D_0 = \frac{V_0}{1 + R}, \tag{3.19}$$

$$\dot{a} = (F - a)/0.5, \qquad \dot{z} = g D_0 (a - x_{\text{spec}}). \tag{3.20}$$

Here M and A are estimated pot moles and light moles, a is the analyzer, and z is the receiver's measured purity surplus. The current main-cut flag g is held across the forecast. The model is

Table 3.8 Scenario 14: Model Information and Reflux Movement

Policy	Product (mol)	True purity	Travel
Composition PI, 600/4	13.8056	.989937	22.46
Fixed nominal ramp	1.6814	.989013	8.11
Fixed ramp plus P600	11.1138	.989893	23.37
Run-model searched ramp	14.3393	.989882	5.81
MPC	14.2914	.989982	173.47

linearized once at each decision, then its exact affine hold is extrapolated over the shrinking collection window.

The inventory estimator starts from 27.15 nominal pot moles. Its initial composition uses the first analyzer sample, with a nominal Fenske inversion for an already rich settled reading. Nominal distillate flow and analyzer history update its material accounting. The receiver state is initialized from the same operator ledger that the original cut logic uses. That ledger already weights analyzer readings by the app's calculated flow. It is not a new pot or tray instrument and does not update the predictor's fixed boil-up. A filtered analyzer slope compensates a nominal half-minute lag. A gain-.1 residual filter, limited to $\pm.05$ mole fraction, adjusts b. Disturbance correction follows the motivation of offset estimation [19], without identifying volatility, actual charge size, heater performance or calibration. No hidden states, run coefficients or later events enter this predictor.

The QP rewards linearized distillate flow through the positive reflux price $hV_0/(1+\bar{R})^2$. Its other costs are $100h(a-x_{\mathrm{sp}})^2$, terminal weight 100, reflux weight $10^{-6}h$ and block-move weight $.002w$, with default $w=1$. Reflux between zero and the present drive limit, with unit changes between blocks, is hard bounded. Predicted receiver surplus and pot inventory are soft lower constraints. The analyzer floor is $x_{\mathrm{spec}}-.02$ during the main cut and $x_{\mathrm{spec}}+.0005$ otherwise, allowing measured receiver credit to fund some instantaneous dilution. Quadratic slack prices are 10^5, 10^7 and 10^4 for analyzer, surplus and pot, with linear prices 10, 100 and 100. These prices optimize a flow proxy under a local approximation. They do not certify maximum terminal product, true purity or recursive feasibility.

The existing composition PID tracks the applied finite reflux request. Any failed, nonfinite, malformed or invalid plan holds that request on the first scan, then the composition PID resumes until a valid plan returns. Historian marks and notes identify failure and recovery. Total reflux, measured cumulative cuts and the clock or pot stop retain their original authority. A tail cut remains irreversible. Ten appended CSV fields report acceptance, iterations, analyzer slack, fallback, full planning time, horizon, predicted end analyzer and receiver surplus, bias and estimated pot inventory. The badge presents the end analyzer value, not final receiver purity.

Independent operating-line reconstruction gives a maximum 6.53×10^{-11} residual against the full nominal steady balances at twelve pot/reflux points. Implicit differentiation agrees with the implemented Jacobian within 5.71×10^{-8}. A closed scalar optimization using the actual four-state linearization and its economic reflux price has zero first-move error. The reduction nevertheless supplies immediate overhead feedthrough: at $x_B=.4$, changing R from four to five changes its static overhead by .003778, while the full condenser has zero initial direct response to reflux. Tray storage and analyzer dynamics matter when receiver credit is small. A two-minute prototype allowed an early tail cut between decisions, motivating the half-minute period before the final comparison settings were frozen.

Scenario 14 puts actual volatility at 1.85 while MPC retains two. Table 3.8 compares complete batches from the same seed. The fixed nominal ramp uses 2.45 to 10.5667. Adding the original proportional trim at gain 600 recovers part of its loss, but the feedback MPC collects more. The existing Searched option uses the run's actual parameters, so its open-loop row has more model information and slightly exceeds MPC. It must not be confused with a frozen nominal ramp.

The 55 measured rows include six PI tunings, four constant ratios, two move prices, halved physics

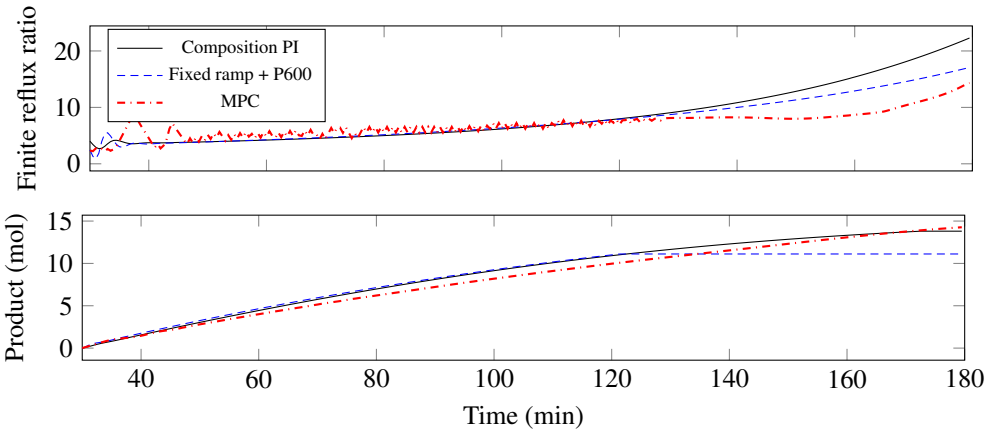


Figure 3.5 Matched volatility-1.85 batches after total reflux: composition PI (solid), fixed ramp with proportional trim (dashed) and MPC (dash-dotted). MPC improves recovery from the fixed nominal policy but requires more reflux movement. Travel excludes startup's infinite reflux and the undefined transition to a finite ratio. All five table rows spend 6480 kJ and meet the app's original receiver tolerance of .001.

steps and the original disturbance cases. The best PI in this small grid collects 13.9663 mol at .990060 with 26.69 of travel. Half-step MPC gives 14.2957 mol at .989990, retaining the comparison order. The nominal searched open-loop case still beats MPC, 15.7349 against 15.4535 mol. At volatility 1.725 the frozen MPC closes its product cut after only .9029 mol at .987859, below the receiver tolerance, while the composition PI collects 9.8613 mol. Larger holdup gives MPC 2.8451 mol against 14.3349 for PI. A positive .012 analyzer offset in scenario 14 produces .978214 true purity under MPC. Current analyzer correction cannot repair a biased reference.

On the unchanged suggested challenge, MPC collects 15.4947 mol at .989950 but fails its tracking target: maximum true error from 86 to 130 min is .01913 against .00472 for the suggested PID. Finite collection travel is 146.71 against 1364.45. The unmodified PID passes, and MPC's smaller movement does not substitute for the challenge objective. Chrome measurements across six viewport/theme captures and 300 plans per capture give .572–.627 ms per full optimization, .114–.125 percent of a core at unit playback and 6.87–7.53 percent at the highest 60-fold speed. This excludes observation, integration and rendering. The reduced horizon is computationally affordable, while its loss cases remain part of the laboratory.

3.7 Discussion

Separation is where a chemical plant spends its energy, and distillation dominates that spending, which is why separations are repeatedly identified as one of the largest available targets for reducing industrial energy use [20]. The dividing-wall column of Volume 13, Chapter 8 meets the same question on a continuous column. A batch column makes the point sharply because the reboiler spends its full duty whether or not anything is leaving the top. The energy per mole in Table 3.5 moves from 429 kJ/mol at the best constant reflux ratio to 412 kJ/mol on the searched ramp, a reduction of 4.0 % bought with no equipment and no extra instrument, and the same number reaches 5143 kJ/mol on a batch whose trajectory was written for the wrong charge.

DID YOU KNOW? A Fifth of the Energy Used in Plants

A 2005 report for the United States Department of Energy put industrial separations at about 4500 trillion Btu a year, about 22 % of all in-plant energy use in the country. Distillation alone took 49 % of the energy spent on separations [21].

The second cost is purity above specification. Scenario 7 ships 0.9957 material against a 0.990 specification and loses 16.4 % of the production doing it, and neither the application nor a plant annunciates over-purity, because no loop is asking for it. The same arithmetic drives solvent recovery, where most batch distillation capacity in fine chemicals and pharmaceuticals sits, and where solvent is the great majority of the mass a batch process handles, so recovering it rather than discarding it is the largest single decision in the material footprint of the route [5]. Measuring the distance between what was shipped and what was specified, and reporting it beside production, quantifies a loss that no control loop annunciates.

3.8 Technology in Practice

What Measures It

Every decision of this chapter's recipe rests on one composition reading. On a production column the reading may come from a process chromatograph, since the monitoring and control of distillation columns is a very common industrial use of chromatographs [22], as on the continuous column of Volume 5, Chapter 3. The laboratory column behind this chapter read refractive index on samples instead. AT-401 stands for either as a first-order lag of 0.5 min with a settable bias and noise, and the same reading closes the loop and switches the receiver.

TECHNOLOGY Sensor: Refractive Index of the Distillate

The column behind this chapter measured its distillate composition through the refractive index of the solution [16]. That index is the speed of light in a vacuum divided by its speed in the liquid, and it sets how far a beam bends on entering it [22]. A laboratory refractometer holds the sample at a controlled temperature and needs about twenty minutes to warm up [23]. A reading becomes a composition only through a calibration curve built from standards of known composition [24]. Instruments on industrial batch columns may report values with drift, noise or outliers [16]. AT-401 of this chapter reports continuously through its lag, and a bias set in a scenario shows on no trend drawn from it.

What Moves It

The only manipulated variable is the split of the condensate. The continuous column of Volume 5, Chapter 3 sets its reflux through a control valve, while a small batch column can split the condensate by time, so that the ratio of two intervals is the reflux ratio. RC-401 applies R as an exact split of V into L and D, with no travel time and no switching cycle, so the staircase of the reference's swinging bucket appears in this chapter only as a smooth ramp.

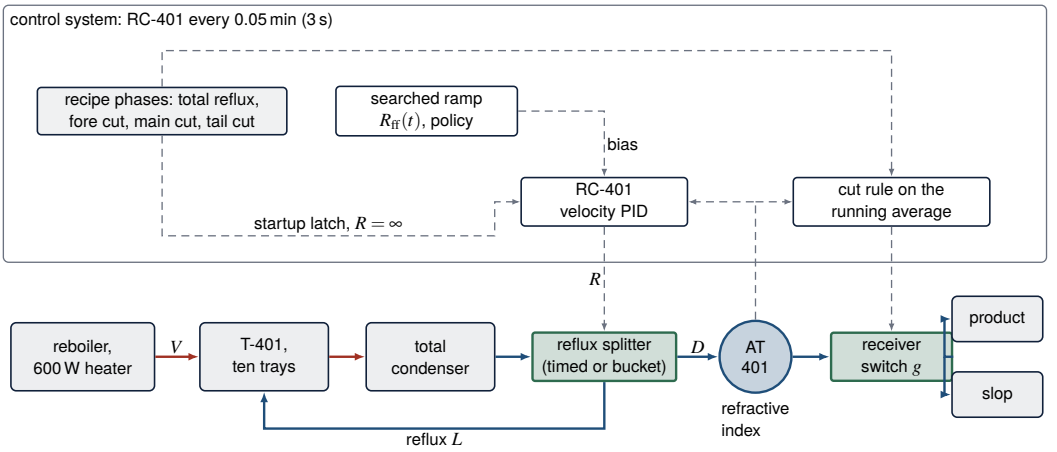


Figure 3.6 The chapter's loop on plant hardware. The analyzer AT-401, a refractive-index measurement on the reference column, reads the distillate, and the same reading drives RC-401, which sets the reflux splitter, and the cut rule, which moves the receiver switch between product and slop. The recipe phases latch total reflux during the startup and order the cuts, and the searched ramp enters RC-401 as its bias. Notice that one late or biased reading reaches both the loop and the cut.

> **TECHNOLOGY** **Final Element: A Timed Reflux Splitter**
>
> On some laboratory columns the reflux ratio is set by time rather than by a valve opening. A reflux valve switches the condensate between the column and the product receiver within a fixed cycle. On one teaching column with a 10 s cycle, a ratio of 4 to 1 sends condensate to the column for 8 s and to the product receiver for 2 s [24]. Frequent switching lets the column more closely approximate a steady state [23]. The column behind this chapter set its ratio with a swinging bucket that could be changed as often as every 5 min [16]. RC-401 instead applies any ratio from 0 to 40 at once, every 0.05 min.

Where the Controller Runs

Batch operations follow recipes developed in the laboratory: a sequence of operations in a prespecified order, repeated on a predetermined schedule to reach the production volume [25]. They are controlled from stand-alone controllers, programmable logic controllers, distributed control systems and personal computers [25]. The batch control committee of ISA describes a functional model that relates recipe management, scheduling, sequential control, regulatory control and safety interlock systems [26]. Read in those terms, the startup at total reflux, the fore cut, the main cut and the tail cut are the sequential control of one recipe, the policy and the searched ramp are recipe parameters, and RC-401 is the regulatory control. Supervisory and lower-level loops run with cycle times from milliseconds to minutes [27]. RC-401 executes every 0.05 min, which is 3 s, as Figure 3.6 draws it on plant hardware.

> **IN PRACTICE** **Running to the Cumulative Specification**
>
> On the column behind this chapter, the base case ran at total reflux for 30 min, then set a constant reflux ratio, usually between 3 and 5, and ran until the cumulative overhead composition reached 99 mol% methanol. Collection usually lasted 60 to 90 min. The instantaneous composition could be measured every 5 min and the total collected was read on a graduated cylinder, so the cumulative composition followed from both through a mass balance [16]. That is the stopping rule of Equation (3.9), kept by hand on readings five minutes apart.

What Is Modernizing It

Two developments use the repetition of batches. The first is learning from run to run. In a simulated binary batch column of ten stages, an iterative learning controller adjusted the reflux trajectory between batches to track a reference for the purity of the accumulated distillate. Its schemes converged after 25 to 30 runs, and a residual error had to be tolerated because the startup phase left the purity at the start of collection uncertain [25]. Most often the path constraints of a batch are met online and its terminal objectives run to run [25], and the receiver purity of this chapter is such a terminal objective. The second is estimation from batch data. The study behind this chapter fitted its model to experiments with an error measure insensitive to outliers, ranked which parameters the data could determine, and only then optimized the reflux schedule [16]. The shrinking-horizon MPC of Section 3.6.9 acts within one batch, and the rework of slop into the next charge, which Limitations sets aside, is where run-to-run learning would act.

3.9 Limitations

No claim about a real column's yield, its cycle time or the economics of a campaign can be founded on this model. The relative volatility was chosen to compensate for a stage count, so the column is a plausible batch column rather than the reference apparatus. There is no pressure, no temperature and no energy balance inside the column, which matters for control rather than only for accuracy, because real columns are frequently run on an inferential tray temperature with an analyzer trimming it and this model cannot show that structure at all [15]. The mixture is binary, where a real batch separates several components and a cut becomes a sequence of cuts with intermediate fractions accumulating between them. One batch is simulated in isolation, so slop is counted and then forgotten, whereas in a campaign it is reworked into the next charge, which changes that charge's composition and therefore the optimal policy for it, and the sequence of batches is the thing actually being optimized. The supervisory search integrates the same equations with the same parameters that the plant does, so it optimizes against a perfect model of a plant whose only unknown is the one a scenario changes, and real trajectory optimization is limited by parameter uncertainty rather than by search [16]. Perfect repeatability is the last simplification and the least obvious one, since a fixed step and a fixed controller period give one trajectory for a given scenario forever, whereas a real batch record from the same recipe never repeats twice. Each of the six modeling assumptions is given the condition under which it gives way in Section 3.3.6.

Two scenario disturbances are stylized. An analyzer that drifts 0.012 high in one step at a nominated minute is not how a calibration fault develops, but a step is the only form of bias whose consequence can be read off a single trend against a known reference run. A charge that is 0.42 instead of 0.59 with no measurement taken would not survive a real batch record, although it isolates what an open-loop trajectory assumes about its plant, which is the whole content of scenario 11.

The added local MPC leaves room for a full nonlinear formulation. Nonlinear Model Predictive Control could re-optimize the reflux trajectory in a receding horizon against the terminal objective, enforcing the cumulative purity constraint and the reflux limit directly rather than through the present frozen local balance or a reset-time ramp [28]. Moving horizon estimation could reconstruct the charge composition from the analyzer and the early trajectory, addressing scenario 11 directly, because a policy re-optimized on the estimated charge recovers 10.69 mol where the fixed ramp collects

1.26 mol. The same estimator structure supports soft sensing of the distillate composition between analyzer updates, which is the practical arrangement on industrial columns [15]. The reference study estimated its parameters and optimized its trajectory in APMonitor [16, 28], and the Gekko Optimization Suite provides the same nonlinear MPC and estimation framework in Python [29].

> ⚠ **OPEN CHALLENGE** **A Model That Predicts the Batch but Not Its Optimum**
>
> A 2017 handbook chapter on batch control calls model adequacy a major challenge in process optimization. A model is trained to predict the plant's outputs, not the gradients that decide optimality, so numerical optimization computes inputs that are optimal for the model and often fails to reach plant optimality. In practice, model-based optimal inputs lead to suboptimal, and often infeasible, operation [25]. Scenario 11 is a small case of it: the ramp searched for a 0.59 charge collects 1.26 mol on a 0.42 charge. Progress would be a policy that finds the plant's own optimum from its measurements as the charge changes.

3.10 Conclusion

This chapter used a ten-tray batch distillation column as an interactive laboratory for control against a plant that is consumed while the controller runs. The simulator integrates nineteen states covering the pot, ten trays, the condenser, two receivers, an analyzer and an energy meter, with equilibrium at a constant relative volatility, constant molar overflow fixing the internal flows, and a boil-up of 0.7072 mol/min set by a 600 W duty independently of the reflux ratio. Eleven equilibrium stages multiply the ratio $x/(1-x)$ by 2048 at total reflux and that factor never changes, while the pot it acts on falls from 0.5787 to 0.2337 in ninety minutes, which takes the distillate from 0.9996 to 0.9599 with no input change. On a 180 min budget against a 0.990 specification the best constant reflux ratio collects 15.10 mol, constant-composition control collects 15.29 mol with its reflux ratio pinned at 40 and its distillate rate down by a factor of about twelve, and a searched ramp from 2.45 to 10.57 collects 15.73 mol open loop. That ordering holds only because the product specification is applied to the cumulative average of the drum, since an instantaneous specification would make constant-composition control optimal by construction. The same ramp collects 1.26 mol on a charge of 0.42, where a proportional trim recovers 4.74 mol and re-optimizing recovers 10.69 mol. Further work would add multicomponent cuts, rework of slop into the following charge, a reboiler with thermal capacitance, and a search that carries parameter uncertainty rather than assuming the model. The new nominal MPC recovers product relative to a mismatched fixed ramp, but extra motion, biased instruments and strong mismatch limit that recovery. The batch column is an accessible browser laboratory for trajectory control against a terminal objective.

KEY POINTS

- A batch column has no operating point: the pot is charged once, its composition falls for the whole batch, and the batch is judged once, on the product drum at the end.
- On a 180 min budget the best constant reflux ratio collects 15.10 mol, constant-composition control 15.29 mol with its reflux ratio pinned at 40, and a searched ramp from 2.45 to 10.57 collects 15.73 mol open loop.
- That ordering holds only because the specification applies to the cumulative average of the drum.
- The same ramp collects 1.26 mol on a 0.42 charge, where a proportional trim recovers 4.74 mol and re-optimizing recovers 10.69 mol.

Where else this applies. Batch crystallization to a yield, in Chapter 4, has the same structure as this column: an input trajectory chosen over a fixed horizon against a result judged at the end, with constraints along the way.

References

[1] 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.

[2] D. Bonvin, "Optimal operation of batch reactors: a personal view," *Journal of Process Control*, vol. 8, no. 5-6, pp. 355–368, 1998.

[3] B. Srinivasan, S. Palanki, and D. Bonvin, "Dynamic optimization of batch processes: I. characterization of the nominal solution," *Computers & Chemical Engineering*, vol. 27, no. 1, pp. 1–26, 2003.

[4] B. W. Bequette, *Process Control: Modeling, Design, and Simulation*. Upper Saddle River, NJ: Prentice Hall, 2003.

[5] D. J. C. Constable, C. Jimenez-Gonzalez, and R. K. Henderson, "Perspective on solvent use in the pharmaceutical industry," *Organic Process Research & Development*, vol. 11, no. 1, pp. 133–137, 2007.

[6] L. Rayleigh, "On the distillation of binary mixtures," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 4, no. 23, pp. 521–537, 1902.

[7] A. O. Converse and G. D. Gross, "Optimal distillate-rate policy in batch distillation," *Industrial & Engineering Chemistry Fundamentals*, vol. 2, no. 3, pp. 217–221, 1963.

[8] I. Coward, "The time-optimal problem in binary batch distillation," *Chemical Engineering Science*, vol. 22, no. 4, pp. 503–516, 1967.

[9] G. P. Distefano, "Mathematical modeling and numerical integration of multicomponent batch distillation equations," *AIChE Journal*, vol. 14, no. 1, pp. 190–199, 1968.

[10] U. M. Diwekar, R. K. Malik, and K. P. Madhavan, "Optimal reflux rate policy determination for multicomponent batch distillation columns," *Computers & Chemical Engineering*, vol. 11, no. 6, pp. 629–637, 1987.

[11] I. M. Mujtaba and S. Macchietto, "Holdup issues in batch distillation-binary mixtures," *Chemical Engineering Science*, vol. 53, no. 14, pp. 2519–2530, 1998.

[12] U. Diwekar, *Batch Distillation: Simulation, Optimal Design, and Control*, 2nd ed. Boca Raton, FL: CRC Press, 2011.

[13] K.-J. Kim and U. M. Diwekar, "New era in batch distillation: Computer aided analysis, optimal design and control," *Reviews in Chemical Engineering*, vol. 17, no. 2, 2001.

[14] J. S. Logsdon and L. T. Biegler, "Accurate determination of optimal reflux policies for the maximum distillate problem in batch distillation," *Industrial & Engineering Chemistry Research*, vol. 32, no. 4, pp. 692–700, 1993.

[15] M. Barolo and F. Berto, "Composition control in batch distillation: Binary and multicomponent mixtures," *Industrial & Engineering Chemistry Research*, vol. 37, no. 12, pp. 4689–4698, 1998.

[16] S. M. Safdarnejad, J. R. Gallacher, and J. D. Hedengren, "Dynamic parameter estimation and optimization for batch distillation," *Computers & Chemical Engineering*, vol. 86, pp. 18–32, 2016.

[17] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, 2nd ed. Princeton, NJ: Princeton University Press, 2021.

[18] 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.

[19] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[20] D. S. Sholl and R. P. Lively, "Seven chemical separations to change the world," *Nature*, vol. 532, no. 7600, pp. 435–437, 2016.

[21] BCS, Incorporated and Oak Ridge National Laboratory, "Materials for separation technologies: Energy and emission reduction opportunities," U.S. Department of Energy, Industrial Technologies Program, Tech. Rep., 2005.

[22] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[23] Brigham Young University, Unit Operations Laboratory, "Batch distillation, version 16.2: startup procedure and refractometer instructions," uolab.groups.et.byu.net/files/batchdist/batchdist_manual.pdf, laboratory manual, accessed 22 November 2026.

[24] University of Utah, Department of Chemical Engineering, "UOL distillation column," www.che.utah.edu/sop/uol-distillation-column, standard operating procedure, accessed 22 November 2026.

[25] D. Bonvin and G. François, "Control and optimization of batch chemical processes," in *Coulson and Richardson's Chemical Engineering*. Elsevier, 2017, pp. 441–503.

[26] International Society of Automation, "ISA88: Batch control, committee scope and purpose," www.isa.org, standards committees, 2026, accessed 22 November 2026.

[27] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[28] J. D. Hedengren, R. Asgharzadeh Shishavan, K. M. Powell, and T. F. Edgar, "Nonlinear modeling, estimation and predictive control in APMonitor," *Computers & Chemical Engineering*, vol. 70, pp. 133–148, 2014.

[29] L. D. R. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, "GEKKO optimization suite," *Processes*, vol. 6, no. 8, p. 106, 2018.

3.11 Exercises

3.1 Concept check · 2 min

A batch of the column of this chapter is run to the end of its 180 min budget. What is the batch graded on?

(a) The largest deviation of the distillate from its 0.990 setpoint during the collection

(b) The moles in the product receiver at the end, provided their cumulative average meets 0.990

(c) The pot composition at the end, because a lean pot shows that the light component was recovered

(d) The reboiler energy spent, because the duty runs whether or not anything leaves the top

3.2 Concept check · 2 min

In scenario 10 the column runs in Manual at a reflux ratio of 6, and at 60 min the reboiler loses 35 % of its duty. What happens to the batch?

(a) The reflux ratio rises, because the reflux falls less than the distillate when the vapor rate drops

(b) The distillate grows leaner, because a lower vapor rate gives each tray less separating power

(c) Vapor, reflux and distillate fall together, so less product is made, and it comes out purer

(d) Nothing changes in the product, because the analyzer sees the fault and the loop restores the duty

3.3 Concept check · 2 min

Suppose the receiver were switched to slop the moment the instantaneous distillate composition fell below the 0.990 specification, instead of when the cumulative average of the drum falls to it. Which policy would then collect the most product?

(a) The searched ramp, because it starts near the minimum reflux and takes product fastest
(b) The best constant reflux ratio, because it needs no analyzer and cannot saturate
(c) Total reflux followed by one high reflux ratio, because every mole would then be pure
(d) Constant-composition control with its setpoint at 0.990, held as long as possible

3.4 Theory and calculation · 6 min

Use the total-reflux relation of Equation (3.7) with ten trays.

(a) At $\alpha = 2$, find the separation factor and the distillate composition when the pot holds $x_B = 0.5787$. *Checkpoint:* $2^{11} = 2048$, $x_D = 0.99964$.

(b) After ninety minutes at $R = 4$ the pot has fallen to $x_B = 0.2337$. What would total reflux give at the top now, and how does that compare with the 0.9599 the column actually delivers at $R = 4$? *Checkpoint:* $x_D = 0.9984$ at total reflux.

(c) Repeat part (a) at the reference chemistry, $\alpha = 1.725$. *Checkpoint:* Factor 402, $x_D = 0.9982$.

3.5 Theory and calculation · 5 min

The reboiler supplies 600 W at a heating efficiency of 0.719, and the heat of vaporization is 36.6 kJ/mol. A total-reflux startup of 30 min precedes a collection that runs to 180 min.

(a) Compute the boil-up V. *Checkpoint:* $V = 0.7072\,\text{mol/min}$.

(b) At a constant reflux ratio of 6, find the distillate rate and the total distillate taken over the collection, product and slop together. *Checkpoint:* $D = 0.1010\,\text{mol/min}$, 15.15 mol.

(c) Find the energy per mole of product for the best constant reflux ratio (15.10 mol) and for the searched ramp (15.73 mol). *Checkpoint:* 429 and 412 kJ/mol.

3.6 Theory and calculation · 4 min

A product drum holds 10 mol at a cumulative 0.995, and the specification is 0.990.

(a) Compute the purity credit $C = n_P(\bar{x}_P - x_{\text{spec}})$. *Checkpoint:* $C = 0.050\,\text{mol}$.

(b) Distillate now arrives at a steady 0.980. How many moles can be added before the cumulative rule of Equation (3.9) closes the drum? *Checkpoint:* 5.0 mol.

(c) How many of those moles would an instantaneous specification accept? *Checkpoint:* None.

3.7 App challenge · 8 min

Load scenario 4, Constant reflux $R = 3$: watch the composition slide. Keep Manual, the default seed and the default speed of 10, which runs a 180 min batch in about 18 s.

(a) Change nothing and run until the batch stops. Read the Product collected, Sent to slop and Batch time used tiles, and the time of the tail cut, which is the first `tail` entry in the exported CSV column `cut_phase`. *Checkpoint:* 9.16 mol, 13.99 mol of slop, stops at 161 min, tail cut at 81.8 min.

(b) Press Reset, set the Reflux ratio to 6 before pressing Run, and run to the end of the batch. Read the same tiles and the Product composition tile. *Checkpoint:* 15.10 mol at 0.9895, 0.06 mol of slop, tail cut at 179.5 min.

A ratio of 8 collects 11.79 mol at 0.9978 with no tail cut. Explain why 6 beats both 3 and 8. Give moles to within 0.02 mol and times to within 0.2 min.

3.8 App challenge · 8 min

Load scenario 5, Constant composition, and the reflux runaway. Keep Auto, the default seed, the default speed of 10 and the shipped tuning ($K_c = 600$, $\tau_I = 4\,\text{min}$).

(a) Run to the end of the batch. Read the Product collected and Sent to slop tiles, the reflux ratio at 120 min from the CSV column `mv_reflux_ratio`, and the tail-cut time from the first `tail` entry in `cut_phase`. *Checkpoint:* 15.29 mol, 0.34 mol of slop, $R = 8.11$, tail cut at 162.7 min.

(b) Press Reset, set the Proportional gain to 300 before pressing Run, and run to the end. Read the same quantities. *Checkpoint:* 13.33 mol, tail cut at 119.7 min.

(c) At 150 min the reflux demand rises by about 0.58 per minute. Estimate the ramp lag $(\mathrm{d}R/\mathrm{d}t)\,\tau_I/K_c$ at both gains and explain why the tail cut moves. *Checkpoint:* 0.0039 and 0.0077.

Give moles to within 0.02 mol, times to within 0.2 min and the reflux ratio to within 0.05.

3.9 App challenge · 9 min

Load scenario 11, A leaner charge than the ramp was optimized for. It runs in Auto with Optimal ramp By hand from 2.45 to 10.57, a Proportional gain of 0, the default seed and the default speed of 10.

(a) Run to the end. Read Product collected, Sent to slop and the tail-cut time from the first `tail` entry in the CSV column `cut_phase`. *Checkpoint:* 1.26 mol, 14.55 mol of slop, tail cut at 36.5 min.

(b) Press Reset, set the Proportional gain to 600 before pressing Run, and run to the end. Read the same quantities. *Checkpoint:* 3.10 mol, 8.07 mol of slop, tail cut at 50.8 min.

(c) Press the reload button, which puts the gain back to 0, switch to Advanced, set Optimal ramp to Searched before pressing Run, and run to the end. Read Product collected and the reflux ratio at the end. *Checkpoint:* 10.69 mol, $R = 17.83$.

What does each remedy recover, and what would a plant need to do before executing a stored recipe? Give moles to within 0.02 mol and times to within 0.2 min.

3.10 Application area · 4 min

To shorten the cycle, a production engineer proposes cutting the total-reflux startup of the shipped column from 30 min to 10 min, keeping the 180 min budget and a reflux ratio of 6. What does the chapter predict, and what would you recommend?

3.11 Application area · 4 min

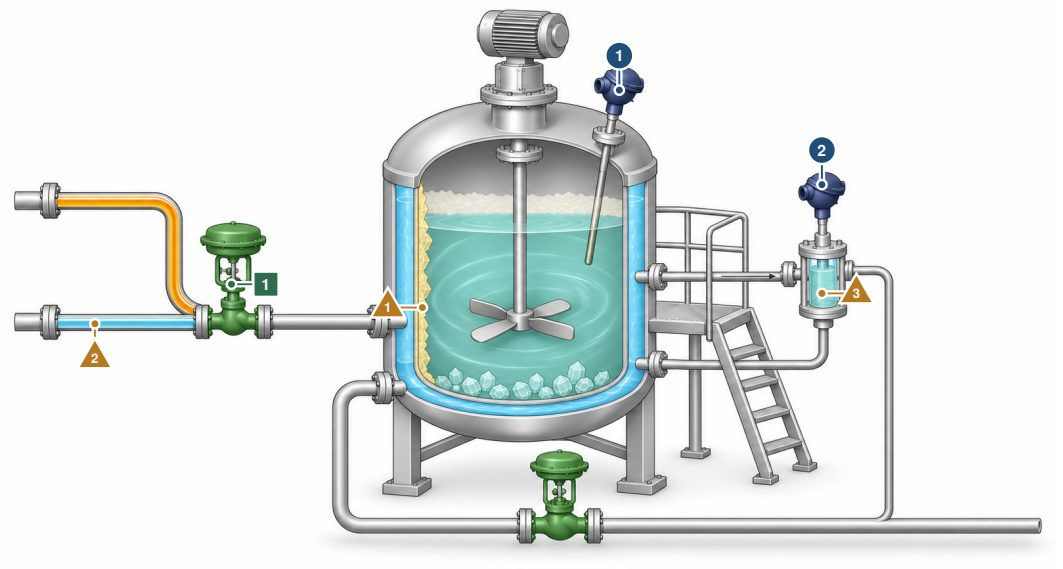
An older plant column has a swinging-bucket reflux divider that can be moved once every 5 min and an online analyzer with about 5 min of lag. The plant asks for constant-composition control at 0.990. Name two problems the chapter predicts and one alternative.

3.12 Review the agent · 5 min

An AI assistant reviews an exported scenario 9 run (the analyzer drifts 0.012 high at 34 min) and writes: “This batch beat the reference run of scenario 5: 16.33 mol against 15.29 mol and 96.7 % recovery against 91.6 %. The receiver average in column `prod_x_meas` ends at 0.9900, so the drum is on specification. Keep the new analyzer calibration.” Find the flaw and say how you would check it.

CHAPTER 4

Batch Crystallizer Simulator for Trajectory Control Inside a Metastable Zone



The controller moves 1 three-way valve blending hot and chilled utility for the jacket

The controller reads 1 batch temperature probe 2 liquor concentration analyzer

What disturbs it 1 scale growing on the cold wall 2 chilled utility warming up 3 analyzer bias

The valve cools the batch along a planned curve, but the crystals are decided by how far the liquor sits above saturation, which only the analyzer sees. Illustration, not to scale.

IN PLAIN TERMS **Growing Large Crystals by Cooling Slowly at First**

The main idea. Cooling a hot liquid brings a dissolved solid out as crystals. A few seed crystals are added so that the new solid grows on them, and the crystal size is judged once, after the batch.

The challenge. Cooled fast while the seeds are small, the liquid makes a cloud of tiny new crystals instead. The thermometer can follow its cooling plan closely all the while, so the shape of the plan matters most.

Why it matters. Tiny crystals clog the filter and dry badly, so a batch that looked fine on every chart can still fail later.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 4.2, the two try-its and the two facts in Section 4.6, and the key points
Study	50 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 4.6), Technology in Practice (Section 4.8), Limitations (Section 4.9) and the open challenge

The control question A seeded cooling crystallization is graded once, after the batch, on crystals no instrument in the loop can see. If three cooling curves on one well-tuned temperature loop give mean sizes that differ by a factor of 4.3, what should the control system be designed around: the tracking error, the trajectory, or the constraints the trajectory meets?

Learning objectives

- Compute the solubility, the relative supersaturation and the metastable limit of the batch from the solubility anchors.
- Derive the cooling rate a seed surface can absorb, and explain why the cubic curve suits a seeded batch.
- Explain why peak supersaturation orders the product and time outside the metastable zone does not.
- Compare the single loop and the cascade with a wall subcooling clamp on an encrusting vessel, by batch time and scale.
- Judge what supersaturation control and a predictive jacket supervisor gain and what each costs in analyzer dependence, fines and valve travel.

Prerequisites Sections 1.4.1 (batch against continuous), 1.4.3 (trajectory tracking), 1.4.4 (path constraints), 1.6.2 (the PI law) and 1.6.1 (engine, speed and seed). Section 1.6.3 is needed only for the predictive supervisor.

The app https://apmonitor.com/apps/batch-crystallization/

See also Volume 3, Chapter 2 for a loop that rejects only the disturbances entering its measurement path, Volume 3, Chapter 3 for the inner-loop speed a cascade requires, and Volume 3, Chapter 8 for a particle size that answers to a residence time and is measured late.

4.1 Introduction

Most plants in an introductory process control course are evaluated while they run. A temperature holds a setpoint, a level holds a band, a composition holds a specification, and the integrated error of the loop is a fair measure of how the plant was operated [1]. A seeded cooling crystallization is evaluated once, after the batch ends, by a particle size analyzer applied to the crystals on the filter. No quantity the control system measured appears in that evaluation. The loop regulates batch temperature because batch temperature has a transmitter, and the engineering content of the recipe is the choice of a temperature trajectory whose consequence is the required size distribution.

The simulator of this chapter makes the gap between the controlled variable and the product objective measurable. Three programmed cooling curves span the same 28 °C in the same 180 minutes on the same single temperature controller, and they give mass-weighted mean sizes of 86.3, 159.2 and 373.4 µm at yields of 62.5, 62.1 and 62.2 %. The temperature tracking gives the opposite ordering. The batch that follows its curve most closely, with a peak deviation of 0.142 °C, finishes at 96.3 % fines by number. The batch with the best product accumulates the largest integrated error of the three and the largest valve travel. This comparison is the central result of the chapter, and it is not available from a transfer function between valve and temperature, because the objective does not appear in that transfer function.

Crystallization determines the solid form of a pharmaceutical product. Size distribution and habit set the filtration rate, the drying time, the flow and compaction behavior in a tablet press and the dissolution rate that governs bioavailability, so the attribute is registered and the crystallizer is the unit that fixes it [2, 3]. A batch that makes 96.3 % fines by number does not fail at the crystallizer. It fails two unit operations later, at the filter and the dryer, after three hours of temperature trends that met the tracking specification.

The class of problem is broader than crystallization. It is batch trajectory tracking against a path constraint, where the objective functional is evaluated only at the final time and the controlled variable is a proxy for a quantity no online instrument reports. Section 1.4.1 sets out these properties of a batch on the kettle of Chapter 1. Batch reactors with a selectivity objective, fed-batch fermentations, polymerization with a molecular weight distribution target and curing operations share this structure.

The form of the optimal input for such problems is a documented general result [4], and the crystallization instance carries a review literature of its own [5]. Two features recur. The optimal input is usually determined by constraints rather than by a sensitivity, so most of the design work is deciding which constraint is active and when. The online measurement is chosen for the transmitter that exists, so an added measurement changes the achievable objective rather than the achievable tracking error.

The model rests on three conservation statements, a solute mass balance, energy balances on the contents and the jacket, and a population balance carried by its first five moments [6], with relative supersaturation as the single driving force for growth and nucleation [7]. The experiments built on the model exercise trajectory tracking, cascade structure justified by time-scale separation, anti-windup and output constraints, measurement lag and bias, and inferential control from a secondary measurement, all standard topics in process dynamics and control texts [1, 8].

This chapter describes a browser-hosted seeded batch crystallizer in the detail needed to rebuild it, and establishes on that plant that the shape of the trajectory decides the product while the tuning of the loop following that trajectory does not. The analysis rests on two measured results, that the ordering of the products reverses the ordering of the tracking errors, and that peak supersaturation orders the outcomes while time spent outside the metastable zone does not. The crystallization kinetics, the method of moments and the cooling profiles are all standard.

Section 4.2 traces the metastable zone, programmed cooling and in-process measurement. Section 4.3 sets down the nine states, a population balance, a wall film and an instrument chain, with their constants, in enough detail for the plant to be reproduced. Section 4.4 compares the three programmed cooling profiles with the cooling rate the crystal surface can absorb. Section 4.5 describes three classical controller architectures on one valve and a predictive supervisor of the jacket. Section 4.6 reports the fourteen scripted scenarios: the cooling curves and a tuning sweep, the ordering of the outcomes, encrustation, supersaturation control, the utility limit, the graded challenge and the predictive comparison. Section 4.8 describes the probe, the valve and the control system of a real crystallizer. Sections 4.7 and 4.9 place the results in plant practice and state the limitations, Section 4.10 summarizes, and the exercises follow.

4.2 Background

The central concept in this domain predates automatic control of it. Its first measurement, told in the history segment below, separated saturation from a further condition at which crystals appear spontaneously. The band between the two is the metastable zone, and it is the operating corridor for a seeded batch. Inside it the crystals already present grow. Outside it the solution nucleates new crystals, a few micrometers across and in numbers that dominate the population count.

> ⧖ HISTORY **1906: A Shower of Crystals Marks the Limit**
>
> In 1906 Henry Miers and Florence Isaac followed the refractive index of cooling sodium nitrate solutions, read by total reflection in a glass prism dipped into the liquid. A stirred solution first grew a few crystals slowly. At a lower temperature a shower of crystals appeared throughout the liquid, and the concentration fell suddenly. They took the two stages for Ostwald's metastable and labile conditions, and they drew a supersolubility curve, nearly parallel to the solubility curve, where a solution passes from one to the other. A crystal put into a metastable solution grows, and if the cooling is slow enough the labile condition is never reached and no cloud of crystals forms [9].

Measurement of the zone width is method dependent. The polythermal method cools a solution at a known rate and records the temperature at which crystals first appear, and the width recovered depends on the cooling rate, on the volume, on the stirring and on the thermal history of the sample [10]. Later analysis established that the width measured this way is set by how long the first nuclei take to grow to a detectable size, so it is a kinetic quantity rather than a thermodynamic boundary [11]. The upper

edge of the corridor is therefore taken from an experiment that does not resemble the batch, and no plant instrument reports its current position.

Cooling policy became a control question once the zone was understood. An unattended jacketed vessel cools on a natural exponential, fastest at the start, which places the highest demand for solute removal at the moment the vessel holds the least crystal surface. Mullin and Nývlt derived the cooling profile that keeps supersaturation constant through a seeded batch and showed it to be strongly convex, slow at the start and fast at the end, and reported that programmed cooling gives a coarser product than either natural or linear cooling at the same endpoints [12]. Jones formulated the same question as an optimal control problem on a population balance and obtained profiles of the same shape [13]. The result has been confirmed experimentally many times, and a controlled comparison on seeded potassium alum found the product mean size to increase from natural through linear to programmed cooling with the seed charge held fixed [14]. Seeding practice moved on the same evidence, from an operator habit to a seed charge and a seeding temperature chosen so that the supersaturation available at the start matches what the seed surface can absorb [7, 15].

Two elements were still missing. The first was a model that could carry a distribution through a dynamic simulation without solving a partial differential equation, which the population balance and its moment transformation supplied, and which remains the standard representation for crystallizer control studies [6]. The second was a measurement. A programmed cooling curve is designed for one seed loading, one growth rate and one solubility, and a plant that departs from any one of them makes the curve wrong. The arrival of in-situ attenuated total reflectance infrared spectroscopy, calibrated through a chemometric model to report liquor concentration during the batch, made the supersaturation itself available as a controlled variable, and closed-loop supersaturation control of a batch crystallizer was demonstrated on that basis [16]. Fujiwara and coworkers separated the two design philosophies that followed, a first-principles route that fits kinetics and optimizes a profile against them, and a direct design route that abandons the kinetic model and holds a concentration trajectory in the phase diagram [17]. Process analytical technology guidance made in-process measurement of this kind a regulatory expectation rather than a research practice [3]. Direct nucleation control went one step further by counting particles with a focused beam reflectance probe and moving the temperature to hold the count, dissolving fines deliberately when there are too many [18]. A review of the field places supersaturation control and direct nucleation control as the two approaches now in industrial use, both motivated by the absence of the product objective from the measured loop [19].

4.3 Crystallizer Model

Crystallizer CR-101 is a 100 liter jacketed vessel charged with 100 kg of water holding 14.0 kg of a dissolved organic solute, which is exactly saturation at 50 °C. The charge is cooled to 48 °C, seeded with 25 g of 50 µm crystal, and taken to 20 °C over 180 minutes. Cooling comes from a jacket fed by a three-way valve that blends a 60 °C hot utility with a 5 °C chilled one. Figure 4.1 shows the equipment, the tag numbers and every symbol the equations below use. Temperature transmitter TT-101 reports the controlled variable and analyzer AT-101 reports the liquor concentration that the third architecture needs, while valve position is the one quantity any of the three controllers can move.

Table 4.1 sorts the signals into the roles a control narrative uses. Units are those of a batch sheet. Time is in minutes, temperature in degrees Celsius, concentration in kilograms of solute per kilogram of solvent, crystal size in meters and displayed in micrometers, moments per kilogram of solvent, heat in kilojoules and valve position in percent open to the chilled utility. Nine states are integrated.

4.3.1 Solubility, Supersaturation and the Metastable Limit

Solubility is a quadratic in temperature, anchored on two measured points and a curvature rather than on three fitted coefficients, so the parameters correspond to what a solubility measurement produces.

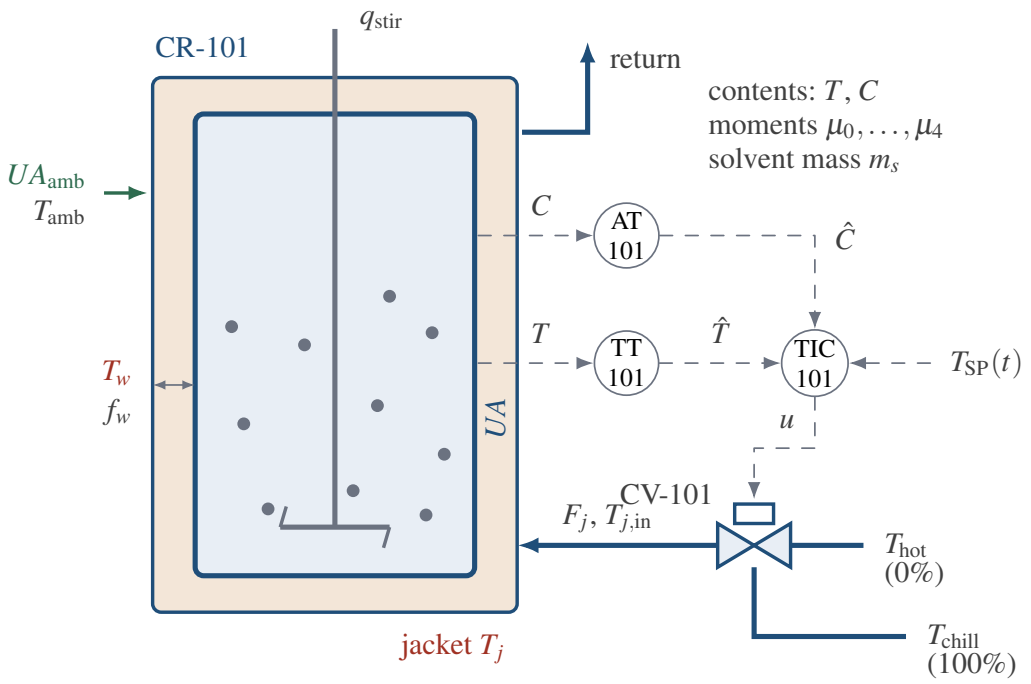


Figure 4.1 Jacketed batch crystallizer CR-101 with its three-way coolant valve. The contents hold solvent mass m_s at temperature T and dissolved solute C, with the crystal population carried as the five moments μ_0 through μ_4. Heat leaves through UA into the jacket at T_j, which is fed at F_j from valve CV-101 blending hot utility T_{hot} at 0 % with chilled utility T_{chill} at 100 %. The crystallizer-side wall sits a fraction f_w of the driving force below the bulk at T_w, which is where encrustation starts. TT-101 reports the batch temperature and AT-101 the liquor concentration, and TIC-101 follows the programmed trajectory $T_{SP}(t)$.

Table 4.1 Classification of Variables. Nominal Values Are Those of the Base Recipe at the Instant the Seed Is Charged.

Symbol	Role	Nominal	Range
T (TT-101)	CV, measured	48.0 °C	20–48 °C
σ (AT-101)	CV (SSC-101), measured	0.062	0.02–0.32 SP
u (CV-101)	MV	20.6 %	0–100 %
T_{chill}	DV, measured	5 °C	2–20 °C
probe bias	DV, unmeasured	0 °C	±2 °C
analyzer bias	DV, unmeasured	0 g/kg	±5 g/kg
T	state	48.0 °C	—
T_j	state	48.4 °C	$\geq T_{chill}$
C	state	0.1400 kg/kg	≥ 0
μ_0	state	2.57×10^6/kg	≥ 0
$\mu_1 \ldots \mu_4$	state	—	≥ 0
ϕ (fouling)	state	0	0–1

The driving force for everything that follows is the relative supersaturation of Equation (4.1).

$$C^*(T) = a_0 + a_1 T + a_2 T^2, \qquad \sigma = \frac{C - C^*(T)}{C^*(T)} \tag{4.1}$$

where C is the dissolved solute in kg/kg of solvent, C^* is the solubility at the same units and σ is dimensionless. Anchor points of 52 g/kg at 20 °C and 140 g/kg at 50 °C with $a_2 = 4.0 \times 10^{-5}\,\mathrm{kg/(kg\,^\circ C^2)}$ give $a_0 = 3.3333 \times 10^{-2}\,\mathrm{kg/kg}$ and $a_1 = 1.3333 \times 10^{-4}\,\mathrm{kg/(kg\,^\circ C)}$. Every rate law, every limit and every setpoint in the application is written in σ and not in the absolute difference $C - C^*$, which is reported for reference only.

The metastable zone width is quoted as degrees of subcooling, so the limit curve is the solubility curve shifted by ΔT_{msz} and the limit in relative supersaturation follows from Equation (4.2).

$$\sigma_{\max}(T) = \frac{C^*(T + \Delta T_{\mathrm{msz}})}{C^*(T)} - 1 \tag{4.2}$$

Because the solubility curve is quadratic rather than exponential, the limit moves with temperature. At 48 °C it is 0.175, at 40 °C 0.190 and at 20 °C 0.207, so the warm end of the batch is the tight end and the margin matters rather than the supersaturation alone.

4.3.2 Growth and Nucleation

Growth is size independent and nucleation has two mechanisms, as shown in Equations (4.3) and (4.4).

$$G = k_g\,\sigma^{g} \tag{4.3}$$

$$B = \underbrace{k_b\,\sigma^{b}\,\mu_3^{j}}_{\text{secondary}} + \underbrace{k_p\,\sigma^{p}\left[1 + \mathrm{e}^{-(\sigma - \sigma_{\max})/w}\right]^{-1}}_{\text{primary}} \tag{4.4}$$

where G is the linear growth rate in m/min and B is the nucleation rate in nuclei per kilogram of solvent per minute. Secondary nucleation is catalyzed by the crystal surface already in the vessel, so it carries the magma term μ_3^{j} and runs everywhere inside the zone as well as outside it. Primary nucleation is spontaneous and is gated off below the metastable limit. With $p = 5$ and a gate width of $w = 0.003$ the primary rate at 40 °C climbs 25 orders of magnitude between a supersaturation 0.15 below the limit and one 0.05 above it, and it multiplies by seven for every hundredth of supersaturation crossed at the limit itself. The gate is a logistic rather than a step because the fixed-step Runge-Kutta integrator evaluates the right-hand side four times per step, and a genuine discontinuity makes those four evaluations disagree about which side of the limit the batch is on. When the liquor is undersaturated, Equation (4.3) is replaced by a dissolution rate $-k_d|\sigma|$ faded linearly to zero over the last five nucleus diameters as the number mean size falls toward L_0, because the moments have no mechanism for removing a crystal from a population.

4.3.3 Population Balance by the Method of Moments

The full description of a crystal population is a number density $n(L,t)$ governed by a partial differential equation in size and time. The moments are its integrals, $\mu_k = \int_0^{\infty} L^k n(L)\,\mathrm{d}L$, and for size-independent growth the partial differential equation collapses exactly onto the ordinary differential equations of Equation (4.5) with nuclei born at a single size L_0 [6].

$$\frac{\mathrm{d}\mu_0}{\mathrm{d}t} = B, \qquad \frac{\mathrm{d}\mu_k}{\mathrm{d}t} = k\,G\,\mu_{k-1} + B\,L_0^{k}, \quad k = 1 \ldots 4 \tag{4.5}$$

Each moment has a meaning. μ_0 is the crystal count, μ_2 is the quantity proportional to total crystal surface, μ_3 is total crystal volume and therefore mass, and μ_4 is what makes the mass-weighted mean available. Equation (4.5) is not an approximation of the population balance under these kinetics. It is the same equation integrated, and the application verifies the identity $\mathrm{d}\mu_3/\mathrm{d}t = 3G\mu_2$ directly with nucleation switched off.

🕮 **Pause.** Try Exercise 4.2 before reading on.

Solute leaves the liquor exactly as crystal volume appears, which fixes the solute balance as Equation (4.6).

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\rho_c k_v \frac{\mathrm{d}\mu_3}{\mathrm{d}t} = -\rho_c k_v \left(3G\mu_2 + BL_0^3\right) \tag{4.6}$$

where ρ_c is the crystal density and k_v the volume shape factor. Writing the balance this way conserves total solute by construction rather than approximately, so the yield reported after a three hour run carries no accumulation error. Every quantity the batch is evaluated on is then arithmetic on the same five numbers, as shown in Equation (4.7).

$$L_{10} = \frac{\mu_1}{\mu_0}, \quad L_{43} = \frac{\mu_4}{\mu_3}, \quad CV = \sqrt{\frac{\mu_0 \mu_2}{\mu_1^2} - 1} \tag{4.7}$$

where L_{43} is the mass-weighted mean size a particle size analyzer reports and a formulation specification is written against, L_{10} is the number mean, which falls sharply when a nucleation event adds fines, and CV is the coefficient of variation of the number distribution. The magma density is $M_T = \rho_c k_v \mu_3$ in kilograms of crystal per kilogram of solvent. The yield is $(C_0 - C)/C_0$ on the solute originally dissolved, and the fines fraction is the nucleated crystal count $(\mu_0 - \mu_{0,\mathrm{seed}})/\mu_0$, reported by number because particle count governs filter blinding.

The seed itself is charged as a log-normal cut of declared mean size and coefficient of variation, for which $M_k = L^k (1 + v^2)^{k(k-1)/2}$, so the two numbers reported by a sieve analysis fix all five seed moments. At the base recipe the 25 g charge is 2.57×10^6 crystals per kilogram of solvent, or 2.6×10^8 in the vessel.

4.3.4 Energy Balances, the Jacket and the Wall

The contents and the jacket are each well mixed, and the three-way valve sets the jacket inlet temperature algebraically, as shown in Equation (4.8).

$$\begin{aligned} M c_p \frac{\mathrm{d}T}{\mathrm{d}t} &= -UA_{\mathrm{eff}}(T - T_j) + \Delta H_c \, m_s \rho_c k_v \frac{\mathrm{d}\mu_3}{\mathrm{d}t} \\ &\quad + UA_{\mathrm{amb}}(T_{\mathrm{amb}} - T) + q_{\mathrm{stir}} \\ m_j c_j \frac{\mathrm{d}T_j}{\mathrm{d}t} &= F_j c_j (T_{j,\mathrm{in}} - T_j) + UA_{\mathrm{eff}}(T - T_j) \\ T_{j,\mathrm{in}} &= T_{\mathrm{hot}} + \frac{u}{100}\left(T_{\mathrm{chill}} - T_{\mathrm{hot}}\right) \end{aligned} \tag{4.8}$$

where $M = m_s(1 + C^*(T_{\mathrm{sat},0})) = 114.0\,\mathrm{kg}$ is the total charge, c_p is the slurry heat capacity, ΔH_c is the heat of crystallization and u is the valve position in percent. Both sides of the first line carry kJ/min. The contents have a thermal time constant of $Mc_p/UA = 433.2/90 = 4.81\,\mathrm{min}$ and the jacket $m_j c_j/(F_j c_j + UA) = 54/234 = 0.231\,\mathrm{min}$, a separation of 21 to 1, which is the time-scale condition that justifies a cascade [8]. The fermenter of Volume 3, Chapter 3 measures how fast an inner loop must be when its lags are not separated this well. The process gain from valve position to batch temperature is negative, because 0 % is full hot utility. A percent of valve initially changes the cooling rate by 0.0703 °C/min rather than the temperature by a fixed amount, so the valve-to-temperature response is near-integrating, with self-regulation through the jacket blend and the ambient leak appearing only beyond an open-loop time constant of 7.6 min.

The crystallizer-side wall sits a fixed fraction of the driving force below the bulk, and the liquor touching it is the same liquor at a lower temperature, as shown in Equation (4.9).

$$T_w = T - f_w(T - T_j), \qquad \sigma_w = \frac{C - C^*(T_w)}{C^*(T_w)} \tag{4.9}$$

The wall supersaturation is therefore always higher than the bulk value, and a batch can encrust while the bulk marker is comfortably inside the corridor. Once the wall crosses its own limit, crystal grows

on the wall and takes the heat transfer with it, as shown in Equation (4.10).

$$\begin{aligned} \frac{\mathrm{d}\phi}{\mathrm{d}t} &= k_f \max(0,\ \sigma_w - \sigma_{\max}(T_w))\,(1-\phi) \\ UA_{\mathrm{eff}} &= UA\,(1-\lambda\phi) \end{aligned} \tag{4.10}$$

where ϕ is the scaled fraction of the surface, k_f the encrustation rate and λ the heat transfer lost at full encrustation. Encrustation is the one fault in the application that feeds itself, because scale lowers UA_{eff}, which forces a colder jacket for the same duty, which lowers T_w and grows more scale. The polymerization reactor of Volume 12, Chapter 11 also insulates itself with the product it makes, and its jacket walks colder until it reaches a floor it cannot pass.

4.3.5 Instruments

TT-101 reports the batch temperature after three modifications. A constant bias is added, colored noise whose stationary standard deviation equals the configured value is added on top of it, and the sum is taken through a probe lag τ_T. AT-101 receives the same treatment with a lag of τ_C, which stands for an in-situ probe reporting through a chemometric model about once a minute. The two lags are stepped with $\hat{x}_k = \alpha\hat{x}_{k-1} + (1-\alpha)x_k$ and $\alpha = \exp(-\Delta t/\tau)$, so neither can go unstable at any sample period. The jacket resistance temperature detector sits in a fast-moving stream and is treated as ideal, because the point of the inner cascade loop is that its measurement is the good one. The third architecture computes $\hat{\sigma} = \hat{C}/C^*(\hat{T}) - 1$ from both instruments, so an error in either propagates into it.

4.3.6 Parameters, Initialization and Numerics

The constants are listed in Table 4.2. Only one of them is fixed, the softness of the nucleation gate, which exists for the integrator rather than for the physics. The rest are adjustable inputs, because the experiments concern choosing a trajectory rather than identifying a plant.

Every batch opens in exact thermal balance. The concentration is set to $C^*(T_{\mathrm{sat},0})$, the seed moments are charged from the log-normal expressions, and the jacket temperature is solved from Equation (4.8) with $\mathrm{d}T/\mathrm{d}t = 0$ so that the vessel is not drifting before the batch starts. At the base recipe that gives $T_j(0) = 48.40\,^{\circ}\mathrm{C}$ and a valve at 20.6 %. The jacket starts warmer than the batch, because a room at 22 °C takes 39.0 kJ/min out of a vessel at 48 °C while the agitator and the growing seed together return only 3.1 kJ/min. A batch is declared finished when T reaches $T_{\mathrm{end}} + 0.3\,^{\circ}\mathrm{C}$ with $\sigma < \sigma_{\mathrm{done}}$, or when the clock reaches $t_{\max}$, and the simulation pauses at that moment so the product measures are read at their final values. Temperature alone does not end a batch, because a vessel at the final temperature that still holds supersaturation continues to grow crystal.

The engine of Section 1.6.1 carries the nine states forward by classical fourth-order Runge–Kutta on a fixed 0.002 min (0.12 s) step. Controllers run every 0.02 min (1.2 s) and the historian keeps five samples for every simulated minute. Both noise sources are drawn from one generator whose seed is 20260829 (20260901 in scenario 11) and which is reseeded at every restart, so a noisy batch repeats exactly.

4.3.7 Assumptions

The model is detailed in the population balance and the wall film and simplified elsewhere. Each simplification has a domain where it fails.

One well-mixed vessel, with one exception. The contents have a single temperature and a single concentration, where a real 100 liter crystallizer has a supersaturation gradient near the cooling surface and a slower zone under the impeller. Exactly one piece of that is restored, the wall film of Equation (4.9), because it carries the encrustation constraint. Everything else about mixing is absent, including the effect of agitation rate on secondary nucleation, which in practice is one of the more influential operating variables.

Table 4.2 Model Parameters at the Base Recipe. All Are Adjustable Except the Nucleation Gate Width.

Symbol	Meaning	Value	Unit
m_s	solvent charge	100	kg
$T_{\mathrm{sat},0}$	saturation temperature of charge	50.0	°C
T_0	temperature at seeding	48.0	°C
m_{seed}	seed charge	25	g
L_{seed}	seed mean size	50	µm
CV_{seed}	seed coefficient of variation	0.25	—
C^*_{20}	solubility at 20 °C	52	g/kg
C^*_{50}	solubility at 50 °C	140	g/kg
a_2	solubility curvature	0.040	g/(kg °C^2)
ΔT_{msz}	metastable zone width	5.5	°C
k_g	growth coefficient	22	µm/min
g	growth order	1.10	—
k_d	dissolution coefficient	200	µm/min
k_b	secondary nucleation coefficient	5.0×10^9	—
b	secondary nucleation order	2.0	—
j	magma exponent	1.0	—
k_p	primary nucleation coefficient	1.0×10^{10}	—
p	primary nucleation order	5.0	—
L_0	nucleus birth size	2.0	µm
ρ_c	crystal density	1300	kg/m^3
k_v	volume shape factor	0.50	—
UA	jacket heat transfer	90	kJ/(min °C)
c_p	slurry heat capacity	3.80	kJ/(kg °C)
m_j	jacket holdup	15	kg
c_j	coolant heat capacity	3.60	kJ/(kg °C)
F_j	coolant circulation	40	kg/min
T_{hot}	hot utility temperature	60	°C
ΔH_c	heat of crystallization	100	kJ/kg
UA_{amb}	ambient heat leak	1.5	kJ/(min °C)
T_{amb}	room temperature	22	°C
q_{stir}	agitator heat input	3.0	kJ/min
f_w	process-side share of ΔT	0.30	—
k_f	encrustation rate	0.30	1/min
λ	heat transfer lost when fully scaled	0.70	—
τ_T	temperature probe lag	0.30	min
τ_C	analyzer lag	1.00	min
$t_{\max}$	maximum batch time	300	min
σ_{done}	end-of-batch supersaturation	0.020	—
w	nucleation gate width (hidden)	0.003	—

Moments with size-independent growth. Equation (4.5) is exact for the assumed kinetics, and five numbers give every graded quantity without reconstructing a distribution. What five numbers cannot give is a shape. A batch holding coarse seed and a fresh crop of 2 µm fines is bimodal, and its first five moments are identical to those of some unimodal distribution that is not in the vessel. The application therefore reports the fines separately by number, and the gap between L_{10} and L_{43} is the other signal that the population has split. No plot in the application is a particle size distribution.

No breakage, no agglomeration, no growth rate dispersion. All three widen a real distribution, and agglomeration in particular can double an apparent mean size while degrading product quality. Their absence is why a clean batch here reaches coefficients of variation a plant would rarely see.

One polymorph and no impurities. Solubility is a quadratic and the zone width is a fixed number of degrees, neither depending on impurity level, solvent composition or crystal form. Polymorphic

transformation during a cooling crystallization is a costly failure mode the application cannot represent, and neither can it represent the impurity rejection that is often the actual purpose of the step.

The metastable limit is a smooth gate at a fixed subcooling. Real zone widths are stochastic, depend on how fast the solution was cooled to reach the condition, and narrow as the batch accumulates surface, and what a polythermal measurement returns is a nucleation and growth time rather than a boundary [10, 11]. The application treats the limit as 5.5 °C of subcooling and draws it on the phase diagram so the simplification is visible rather than hidden.

Dissolution leaves the count wrong. The fade near L_0 keeps the solute balance closed through a partial dissolution and keeps the mass right, but the crystal count afterward is not physical and neither is the coefficient of variation computed from it.

Constant physical properties and one lumped fouling number. The heat of crystallization, the ambient leak, the agitator input and the slurry heat capacity do not change as solid appears, which over a 28 °C cooling of a 114 kg batch is worth a fraction of a degree. Real scale grows unevenly and changes the mixing as well as the heat transfer, where Equation (4.10) carries one fraction.

The numerics are more repeatable than an industrial control system. With the integration step and the execution period both frozen, one trajectory appears on every machine, and no distributed control system behaves that way.

4.4 Cooling Profiles and the Admissible Cooling Rate

Three programmed profiles are available, and all three run from T_0 to T_{end} in the same batch time t_b, so the comparison between them is about shape alone, as shown in Equation (4.11).

$$T_{\mathrm{SP}}(t) = \begin{cases} T_0 - \Delta T\,\tau & \text{linear} \\ T_{\mathrm{end}} + \Delta T\,\dfrac{\mathrm{e}^{-4.605\,\tau} - 0.01}{0.99} & \text{natural} \\ T_0 - \Delta T\,\tau^3 & \text{cubic} \end{cases} \tag{4.11}$$

where $\tau = t/t_b$ and $\Delta T = T_0 - T_{\mathrm{end}}$. The natural profile is the exponential an uncontrolled vessel with a fixed jacket temperature follows, rescaled to land exactly on the end point instead of one percent short of it. Over the base recipe the three start at 0.156, 0.724 and 0.000 °C/min and finish at 0.156, 0.007 and 0.467 °C/min, as shown in Figure 4.2.

The arithmetic that separates them is available in closed form. Holding supersaturation constant means $\mathrm{d}\sigma/\mathrm{d}t = 0$, and differentiating $C = C^*(T)(1+\sigma)$ and substituting Equation (4.6) gives the cooling rate the crystals can absorb, as shown in Equation (4.12).

$$\left|\frac{\mathrm{d}T}{\mathrm{d}t}\right|_{\sigma\,\mathrm{const}} = \frac{3\,\rho_c k_v\,G\,\mu_2}{(1+\sigma)\,\mathrm{d}C^*/\mathrm{d}T} \tag{4.12}$$

where the nucleation contribution has been dropped because inside the zone it is negligible against growth. The admissible cooling rate is proportional to μ_2, which is the crystal surface, and the crystal surface is whatever the cooling has produced so far. That proportionality is the central result of this section. At the instant of seeding, $\mu_2 = 6.82 \times 10^{-3}\,\mathrm{m^2/kg}$, $\sigma = 0.062$ and $G = 1.02\,\mu\mathrm{m/min}$, so Equation (4.12) returns 0.0032 °C/min. The linear profile demands 48 times that value and the natural profile 224 times it, while the cubic profile demands essentially nothing at the start. By the end of the batch the crystal surface has grown from 0.68 to 36.2 $\mathrm{m^2}$, a factor of 53, and Equation (4.12) grows with it.

A batch crystallization therefore matches a cooling demand to a solute-removal capacity that grows during the run. The general solution is to ramp the demand as the capacity grows rather than linearly in time, which is also the structure of the solution for an autocatalytic reaction or a fed-batch fermentation, and it is the result obtained in closed form for programmed cooling [12, 13].

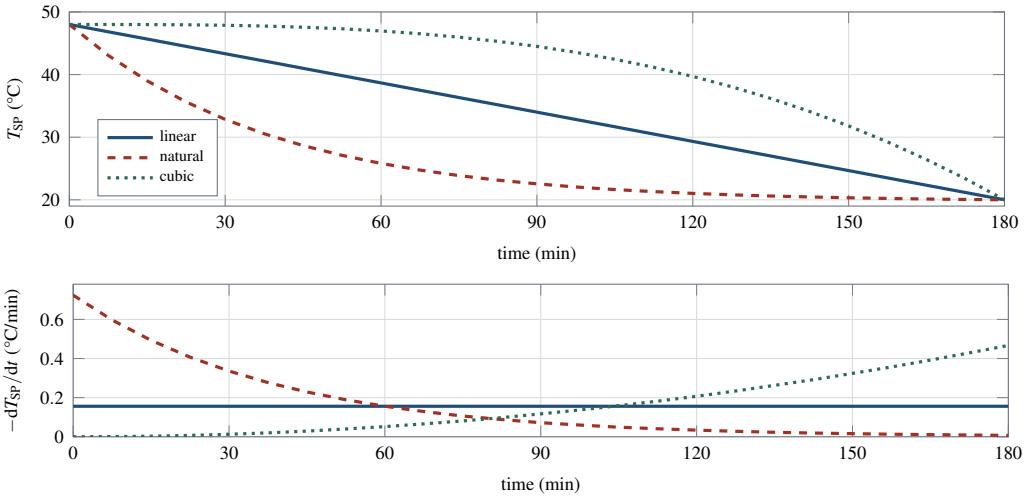


Figure 4.2 The three programmed cooling profiles of Equation (4.11) (linear solid, natural dashed, cubic dotted) and the cooling rates they demand, for the base recipe of 48 °C to 20 °C in 180 minutes. All three cover the same span in the same time. The seed can absorb 0.0032 °C/min at $t = 0$ without any rise in supersaturation, which the natural profile exceeds by a factor of 224 and the linear profile by a factor of 48.

4.5 Controller Design

4.5.1 Classical Architectures on One Valve

The three classical architectures drive the same three-way valve. The single loop is a temperature controller straight onto the valve. The cascade adds an inner jacket temperature loop, so the outer controller's output is a jacket setpoint rather than a valve position. Supersaturation control replaces the programmed trajectory entirely, so the outer controller reads $\hat{\sigma}$ and produces a temperature setpoint that the same temperature controller then follows. Every controller is the velocity-form algorithm of Section 1.6.2, written for this chapter as Equations (4.14) and (4.15).

$$e_k = s\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad s = \pm 1 \tag{4.13}$$

$$\begin{aligned} \Delta u_k^P &= K_c\,(e_k - e_{k-1}) \\ \Delta u_k^I &= \frac{K_c}{\tau_I}\, e_k\, \Delta t_c \\ \Delta u_k^D &= -\,K_c \tau_D\, s\, \frac{\mathrm{PV}_k^f - 2\mathrm{PV}_{k-1}^f + \mathrm{PV}_{k-2}^f}{\Delta t_c} \end{aligned} \tag{4.14}$$

$$\begin{aligned} w_k &= w_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= w_k + I_{k-1} \\ u_k &= \mathrm{slew}(\mathrm{clamp}(c_k, u_{\min}, u_{\max})) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{4.15}$$

where s is the action, PV^f is the measurement passed through a first-order derivative filter of time constant τ_f stepped the same way as the instruments, and T_t is the tracking time. In the notation of Section 1.6.2, s is the action a, the term w_k is the running sum S_k of proportional and derivative increments, and c_k is the output before limiting v_k, with no output bias. The integral used at execution

k was accumulated through execution $k-1$. The demand is clipped to the output limits and then slew limited to $r_{\max}\Delta t_c$ per execution, and only the supersaturation controller uses the slew limit. Anti-windup is back-calculation on the integral alone, and w_k is never truncated. The application sets $T_t = \tau_I$ on all three controllers, so at $\tau_I = 5\,\text{min}$ each execution of the single loop removes 1/250 of the clipped excess, and a loop held against a limit settles with I at the applied output rather than growing with the error. The tracking time affects the reported results where a loop runs against a limit. A recursion that absorbed the whole excess at every execution would give 366 µm at 35 % fines rather than 348 µm at 59 % on the over-tuned row of Table 4.5, and would finish the batch clamped at 1.0 °C in Table 4.8 in 256 rather than 259 minutes. An integral time at the top of its range, 60 min, is read as no reset at all. The integral gain is then zero and the back-calculation term is not applied, because there is no integral to unwind and a proportional-only controller would acquire a permanent bias.

Controller action is the setting that a change of architecture alters. The process gain from valve position to batch temperature is negative, so the single loop and the inner jacket loop both carry $s = -1$. The cascade outer loop is the exception, because its output is a jacket setpoint whose process gain onto batch temperature is positive, and a batch above setpoint needs a lower jacket setpoint, so that loop carries $s = +1$. Changing architecture at run time therefore changes what the outer controller's output means as well as the sign of its error, and carrying the previous integral across would step the valve. An architecture change is therefore treated as a transfer. For one execution every controller runs a tracking step that sets I in Equation (4.15) so that the demanded output equals the valve position already on the plant, the same tracking step that runs at every execution in Manual (Section 1.6.2). Because the integral used at the first automatic execution is the tracked one, the transfer itself moves nothing, and that execution changes the valve by the proportional increment alone. Over the three control executions the bumpless check samples, the valve moves 3.6 %, of which 2.9 % is proportional response and 0.7 % comes from two integral increments, because the measured batch temperature is falling at 4.2 °C/min during the transfer.

4.5.2 Wall Constraint as an Output Limit

The cascade serves a constraint as well as disturbance rejection. By Equation (4.9) the wall is colder than the bulk by $f_w(T - T_j)$, so the only variable that constrains the wall is the jacket temperature. A single loop commands a valve position, which offers no variable on which to state that limit. The cascade outer loop commands a jacket temperature, and its lower output limit is set every execution from the measurement, as shown in Equation (4.16).

$$T_{j,\min} = \hat{T} - \frac{\Delta T_{w,\max}}{f_w} \tag{4.16}$$

The limit therefore moves down the batch with the measured temperature, and anti-windup handles the saturation, because a constraint enforced through the controller's own output limit is seen by the back-calculation term of Equation (4.15), which holds the integral near the floor while the floor is active. Choosing an architecture partly determines which limits can be stated.

4.5.3 Supersaturation Controller

The third architecture's output is a temperature setpoint bounded below by T_{end} and above by T_0, and slew limited to a maximum cooling rate $r_{\max}$. Integral action is required. Holding a constant supersaturation means cooling continuously, and a proportional term alone cannot produce a falling output against a constant error, so the integral term generates the cooling ramp. The slew limit is equally necessary, because it prevents a noisy analyzer from demanding a step change in temperature setpoint. The default settings are $K_{c,\sigma} = 80\,^{\circ}\text{C}$ per unit of supersaturation, $\tau_{I,\sigma} = 15\,\text{min}$ and $r_{\max} = 0.80\,^{\circ}\text{C/min}$. The controller starts with the transfer of an architecture change, because the seeded liquor reads $\hat{\sigma} = 0.062$ against a setpoint of 0.13 and an ordinary first execution would add 5.5 °C of proportional demand. The setpoint starts at T_0 and the integral term opens the ramp at 0.37 °C/min.

Table 4.3 Controller Settings as Shipped. The Single Loop Is the Default and Is Used for Every Cooling Curve Comparison in Section 4.6.

Loop	Action	K_c	τ_I (min)	Output limits
TIC-101, single	−1	12	5.0	0–100 %
TIC-101, cascade	+1	5	4.0	$T_{j,\min}$–55 °C
TJC-101, inner	−1	5	0.3	0–100 %
SSC-101	−1	80	15	T_{end}–T_0

Table 4.3 lists the shipped settings. Derivative time is zero on every loop by default, because a lagged temperature probe on a plant whose setpoint is a smooth ramp offers no benefit from anticipation while amplifying measurement noise.

4.5.4 Predictive Jacket Supervision

The additional controller optimizes a jacket-temperature request rather than replacing the fast jacket loop. Section 1.6.3 describes the predictive option of the apps in general. Receding-horizon constraints are an established industrial control architecture [20]. This implementation uses three measured states $x = (T, T_j, C)$ and the declared clean-vessel constants, with the reduced dynamics

$$\begin{aligned} v &= a[\max(C/C^*(T) - 1, 0)]^g, \\ \dot{T} &= [-UA(T - T_j) + Hv + q_a(T)]/(Mc_p) + b_T, \\ \dot{T}_j &= (u - T_j)/0.3, \qquad \dot{C} = -v. \end{aligned} \tag{4.17}$$

Here u is the jacket request in degrees Celsius, a is the estimated solute-uptake coefficient, $g = 1.1$, b_T is a temperature-rate correction, $H = \Delta H_c m_s$, and $q_a(T) = UA_{\text{amb}}(T_{\text{amb}} - T) + q_{\text{stir}}$. The remaining symbols retain their energy-balance definitions. The initial $a = 3\rho_c k_v \mu_{2,0} k_g$ uses the declared seed surface. Measured concentration change updates a with an 8 minute filter when $\sigma > 0.03$ and uptake exceeds 10^{-8}, within $10^{-7} \le a \le 0.3$. The preceding temperature prediction updates b_T with gain 0.2 and a ± 2 degree-per-minute cap. Neither population moments nor actual fouling, kinetics or utility disturbances are supplied to this estimator. Instrument lags are not inverted. Bounded bias adaptation here does not establish the steady-state conditions needed for offset-free control [21].

Central differences linearize Equation (4.17) at each plan, followed by exact zero-order-hold discretization. Independent analytic Jacobians at 22, 35 and 48 degrees Celsius agree within 4.9×10^{-11} in absolute rate-derivative units. This is a derivative check rather than a plant-prediction residual. Isolated one-degree jacket steps fit a 0.17 minute lag with 0.0155 degree RMS error, while the retained nominal 0.3 minute model gives 0.0949 degree RMS. That model was frozen before the final comparisons.

A plan is computed every 0.5 minutes, or 25 controller scans. Its length in steps of 0.5 minutes is

$$N = \max\big(1,\ \min(120,\ \lceil (t_b - t)/0.5 \rceil)\big),$$

initially at most 60 minutes and shrinking near the recipe endpoint. Twelve graduated input blocks at $N = 120$ reduce the decision count. The objective is

$$\begin{aligned} \min \sum_{k=1}^{N} & \{0.03(T_k - T_e)^2 + 10(\sigma_k - \sigma_r)^2\} \\ & + 0.05 \sum_j (\Delta u_j)^2 \\ & + 0.3(T_N - T_e)^2 + 10(\sigma_N - \sigma_r)^2 \\ & + 10^7 (s_b^2 + s_w^2) + 10^4 (s_b + s_w) \\ & + 10^4 s_d^2 + 100 s_d. \end{aligned} \tag{4.18}$$

Table 4.4 Three Cooling Curves on One Controller, 48 °C to 20 °C With 25 g of Seed. Numbers Are the Loop and Product Measures at the Moment the Batch Stops.

	Linear	Natural	Cubic
How the loop performed			
Peak deviation (°C)	**0.142**	0.637	0.237
IAE (°C min)	13.43	**13.06**	15.07
Valve travel (% of span)	54	52	66
How the batch performed			
Mean size L_{43} (µm)	159.2	86.3	**373.4**
Coefficient of variation	0.293	0.310	0.308
Fines by number (%)	96.3	99.3	**13.8**
Yield (%)	62.1	62.5	62.2
Peak σ	0.213	0.353	0.150
Time outside the zone (min)	19.1	14.9	**0.0**
Batch time (min)	178.6	152.1	186.8

The final recipe temperature is T_e, and σ_r is the lesser of the operator's request and the bulk limit. Jacket requests remain within the cascade's wall-derived lower bound and hot-side upper bound, with at most three degrees per block-boundary move. Bulk and wall supersaturation limits use the nominal metastable boundary less 0.025, and the third soft limit is the requested wall subcooling. Each $s \geq 0$ is shared across its band's horizon. Existing excess receives an allowance decaying with a 6 minute constant. A fixed 10^{-8} Hessian ridge avoids an unintended absolute-input penalty from scaling the large slack weights. The wall clamp remains active above the optimizer. An invalid status, nonfinite prediction or cost, malformed plan or out-of-box move rejects the plan and restores the tracked cascade.

4.6 Results

The fourteen scenarios are ordered by the concept each one introduces: the batch cooled by hand with the valve wide open (scenario 1), the three cooling curves (2 to 4), cascade and supersaturation control (5 and 6), the recipe faults (7 to 9), the plant and instrument faults (10 to 12), the predictive supervisor (13) and the graded challenge (14). Loading a scenario sets the operating mode, the architecture, the cooling curve, the tunings, the parameter and disturbance overrides and the timed events it requires, and Section 1.6.4 describes how a case is credited. The challenge is judged on the case as given, so changing a model parameter or a disturbance voids the attempt. Reset restarts a batch from $t = 0$ with the learner's settings kept, which makes most comparisons one change and a second run, while reloading the scenario restores its own settings.

4.6.1 Cooling Curve Comparison

Table 4.4 reports both the loop and the product measures for the three cooling curves, all on the single temperature loop at $K_c = 12$ and $\tau_I = 5$ min, between the same endpoints in the same nominal 180 minutes. Integrated error accumulates over the whole batch in each case, so the observation windows are 178.6, 152.1 and 186.8 minutes and the three IAE figures are not strictly comparable. Peak deviation and valve travel do not have that problem. Peak deviations in this section are taken at the 0.002 min integration step, where the five-per-minute log reads up to 4 % lower. Integrated error is the application's own figure from that log, and valve travel is accumulated at every integration step, as the exported record carries it.

The two halves of Table 4.4 give opposite rankings. The linear batch has the smallest peak deviation in the table, 0.142 °C, and it finishes at 159.2 µm with 96.3 % fines. The natural batch has the lowest integrated error of the three, 13.06 °C min, and the worst product at 86.3 µm and 99.3 % fines. The cubic batch tracks worse than the linear one on both measures and travels the valve 22 %

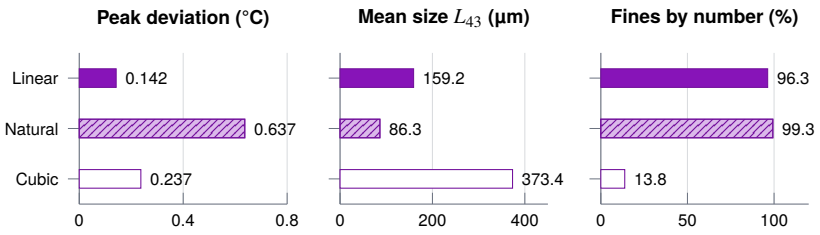


Figure 4.3 The three cooling curves of Table 4.4 on the same loop, between the same endpoints in the same nominal 180 minutes: the loop's peak deviation, the mass-weighted mean size and the fines by number. Notice that the rankings reverse: the linear curve is followed most closely, and the cubic curve, which the loop follows less closely, gives crystals two and a third times as large with a seventh of the fines.

Table 4.5 Tuning Sweep on the Cubic Curve, Scenario 4. Integrated Error Is Accumulated Over Each Batch, Whose Length Is the Last Column.

K_c	τ_I (min)	Peak dev. (°C)	IAE (°C min)	Travel (%)	L_{43} (µm)	Batch (min)
2	20	6.23	503.96	52	373.9	276.5
4	10	1.90	119.43	59	372.6	191.7
12	5	0.237	15.07	66	373.4	186.8
30	2	0.275	**5.23**	80	373.3	186.1
50	1	0.258	8.24	138	373.2	186.0
60	0.5	1.61	105.12	**9114**	348.3	185.3

further, and its crystals are two and a third times larger with a seventh of the fines. Yield differs across the three by four tenths of a point, because yield is set by the final temperature and by almost nothing else.

Figure 4.3 draws the two rankings side by side.

> **TRY IT IN THE APP** **4 · Cubic Cooling, Matched to the Growing Surface**
>
> Load the scenario and press Run. Watch the operating point on the phase diagram travel down between the solubility curve and the metastable limit without touching either line. The batch ends at 186.8 min with 373.4 µm crystals and 13.8 % fines, two and a third times the size of the linear batch, which tracked its curve better.

The cubic curve tracks worse for a structural reason. Its setpoint accelerates, from zero to 0.467 °C/min, and a controller with reset always lags a ramp whose slope is rising, while a linear setpoint has a constant slope that a proportional-plus-integral controller matches with a constant integral term and no residual error. The curve that is easiest to follow is the one that damages the product.

Pause. Try Exercise 4.1 before reading on.

Tuning does not change the result. Table 4.5 sweeps the temperature loop over a fifteen to one range in gain on the cubic curve.

Between a gain of 2 and a gain of 30 the integrated error falls by a factor of 96 and the mean size moves by 0.7 µm. Loose tuning costs batch time, because at a gain of 2 the loop never reaches its own curve and the batch runs 276.5 minutes instead of 186.8. Vessel occupancy, not product quality, is the argument for tuning this loop well.

Table 4.6 Peak Supersaturation Orders the Product and Time Outside the Zone Does Not. All Five Batches Use the Same Kinetics and the Same Vessel.

Batch	Peak σ	Outside (min)	L_{43} (µm)	Fines (%)
Cubic, 25 g seed	0.150	0.0	373.4	13.8
Linear	0.213	19.1	159.2	96.3
Cubic, 6 g seed	0.222	21.5	163.4	99.3
Natural	0.353	14.9	86.3	99.3
Manual, valve wide open	0.706	5.5	42.5	99.9

Table 4.7 Cooling Time on the Cubic Curve, Single Loop. The 150 Minute Batch Never Leaves the Metastable Zone.

t_b (min)	L_{43} (µm)	CV	Fines (%)	Peak σ	Outside (min)
75	121.4	0.376	99.0	0.303	18.8
120	190.3	0.411	97.1	0.226	19.4
150	342.1	**0.550**	**68.2**	0.188	**0.0**
180	373.4	0.308	13.8	0.150	0.0
240	373.8	0.269	10.6	0.104	0.0

The last row reverses the argument, and the damage arrives through an indirect path. At $K_c = 60$ and $\tau_I = 0.5$ min the loop hunts, the valve travels 9114 % of its span, the wall reaches 6.66 °C of subcooling, 77 % of the jacket surface scales over and the product falls to 348.3 µm at 59.1 % fines. The bulk supersaturation leaves the metastable zone for 0.9 min. Over-tuning damages this batch through the coldest surface in the vessel and through the actuator, and not through the variable the controller is regulating.

> **+ DID YOU KNOW? Ninety-Six Times the Error, the Same Crystals**
>
> From the simulator: on the cubic curve, a loop at a gain of 2 accumulates 503.96 °C min of integrated error and one at a gain of 30 only 5.23 °C min, yet their crystals measure 373.9 and 373.3 µm (Table 4.5). The loose loop costs vessel time instead, 276.5 minutes against 186.1.

4.6.2 Ordering of the Outcomes

A candidate explanation for Table 4.4 is time spent above the metastable limit, and the data reject it. Table 4.6 places five batches in order of their peak supersaturation.

Reading the product columns against the time outside the zone gives no ordering. The under-seeded batch spends 21.5 min outside the zone and finishes at 163.4 µm, while the batch cooled manually with the valve wide open spends 5.5 min outside it and finishes at 42.5 µm, four times smaller on a quarter of the exposure. Reading them against the peak supersaturation orders the fines across all five batches and the mean size in all but one pair, the linear and the under-seeded batches, whose peaks differ by 0.009 and whose mean sizes differ by 4 µm on different seed charges. The mechanism is in Equation (4.4). Primary nucleation goes as the fifth power of supersaturation, so a peak of 0.706 against a peak of 0.213 is a rate ratio of $(0.706/0.213)^5 = 400$, and a short excursion to a high peak makes far more crystals than a long excursion to a low one. Cumulative time above the limit is therefore the wrong statistic. This is the general form of any constraint violation whose penalty is a high-order power of the overshoot.

🕮 **Pause.** Try Exercise 4.3 before reading on.

A second correction runs in the opposite direction. Staying inside the zone is necessary but not sufficient. Table 4.7 shortens the batch time on the cubic curve.

Table 4.8 The Same Cubic Recipe in a Vessel With $UA = 18\,\mathrm{kJ/(min\,°C)}$, Scenario 12. The Cascade Rows Keep the Scenario's Outer-Loop Tuning, $K_c = 12$ and $\tau_I = 5\,\mathrm{min}$. The Last Column Is the Largest Departure From the Programmed Curve.

Architecture	Peak ΔT_w (°C)	Scaled (%)	L_{43} (µm)	Batch (min)	Worst error (°C)
Single loop	4.35	**31.2**	375.5	189.0	1.01
Cascade, 1.5 °C clamp	1.48	2.7	373.8	220.5	7.02
Cascade, 1.0 °C clamp	0.99	**0.0**	373.5	**259.1**	10.37

The 150 minute row carries the result. The bulk supersaturation never crosses the limit, the mean size falls only 8 % from its best value, and the batch still finishes at 68.2 % fines with a coefficient of variation of 0.550, the worst in the application. Those fines are entirely secondary nucleation, catalyzed by the seed surface and scaling as σ^2 by Equation (4.4), which runs everywhere inside the zone as well as outside it. Between 180 and 150 minutes the fines rise fivefold, and a further 30 minute reduction removes most of what remains of the product. The last 17 % of the batch time secures almost all of the product quality, a trade that is invisible wherever batch time is scheduled without product data of this kind.

> ✦ **DID YOU KNOW? Ten Minutes on the Clock**
>
> The 2004 guidance on process analytical technology observed that most pharmaceutical processes run to time-defined end points, such as blending for 10 minutes, which ignore physical differences in the raw materials [22]. A batch cooled on a fixed clock is one: the 150 minute row of Table 4.7 never leaves the metastable zone and still finishes at 68.2 % fines.

Scenario 9 separates two effects that are easy to confuse and that the seeding literature keeps apart for the same reason [14, 15]. The recipe seeds at 48 °C, two degrees below saturation, where σ is 0.062 and the seed grows from the first minute. At 50 °C the liquor is exactly saturated, the seed survives intact, and the batch still collapses to 181.3 µm at 96.4 % fines, because the curve now has 30 °C to cover instead of 28 and nothing is banked at the start. At 52 °C the liquor is undersaturated at $\sigma = -0.057$, the seed dissolves at 11 µm/min and is gone inside five minutes, and the batch crosses the limit at 111.6 min, an hour and a half after the seeding error, finishing at 113.7 µm with 98.3 % fines. Neither case produces a visible symptom at seeding time, because the vessel is clear and no instrument on the plant reports that the seed has dissolved.

4.6.3 Encrustation and the Wall Subcooling Clamp

Scenario 5 warms the chilled utility from 5 to 12 °C at $t = 130\,\mathrm{min}$. The single loop takes 0.552 °C of peak deviation and 14.2 °C min of integrated error over the remainder of the batch, and the cascade takes 0.245 °C and 9.3 °C min, because the inner loop sees the disturbance in its own measurement within seconds instead of waiting for a 114 kg charge with a 4.81 min time constant to change temperature. The product is identical either way at 373 µm and 13.8 % fines, which is the honest report. On that batch the disturbance never reached the crystals. The tank of Volume 3, Chapter 2 shows the general rule behind the faster rejection: a loop rejects the disturbances that enter its own measurement path.

Scenario 12 is the case the cascade clamp addresses. The recipe that works in the stainless development vessel is moved into a glass-lined train where $UA = 18\,\mathrm{kJ/(min\,°C)}$ instead of 90, which is five times as much temperature difference for the same duty, and nothing else is changed. Every product criterion passes while the vessel encrusts, as shown in Table 4.8.

On the single loop the bulk never comes closer than 0.04 to its limit and the product is 375.5 µm at 13.9 % fines, while the wall crosses its own limit at 116.5 min, stays over for 65 min, and 31.2 %

of the heat transfer surface is scaled over at the end of the batch. A technology transfer protocol that verifies the trajectory rather than the heat transfer coefficient verifies the part that was never at risk. The last column of Table 4.8 distinguishes an active constraint from a tuning fault. At a 1.0 °C clamp the loop falls 10.37 °C behind its own curve because the constraint is active, and the clean jacket costs 70 minutes of additional vessel time.

> **TRY IT IN THE APP** **12 · A Glass-Lined Vessel, and the Wall That Scales**
>
> Load the scenario and press Run on the single loop. Watch the wall subcooling strip, not the bulk. The bulk never comes closer than 0.04 to its limit and the crystals reach 375.5 µm, yet the wall crosses its own limit at 116.5 min and 31.2 % of the surface scales over. Then run the cascade with a 1.0 °C clamp: no scale, and 259.1 min.

4.6.4 Supersaturation Control

Supersaturation control gives 373.6 µm at 10.9 % fines in 193.3 minutes against 373.4 and 13.8 % in 186.8 for the cubic curve, so on the nominal batch the improvement is small but real and costs 6.5 minutes. The distinguishing feature is the setpoint trace rather than the product numbers, because the cooling curve is generated from a measurement instead of from an assumed kinetic model [16, 17]. The supersaturation setpoint sets the trade between batch time and fines, and the trade is monotone until the batch runs against the metastable limit. At 0.09 the batch takes 279.6 minutes for 8.1 % fines, at 0.13 it takes 193.3 for 10.9, and at 0.17 it takes 149.2 min for 13.5. A setpoint of 0.20 lies above the limit of 0.175 at the warm end, but the reset term cools too slowly to lift σ above 0.173, and the batch finishes in 126.7 minutes at 28.5 % fines. At 0.21 the bulk comes within 0.004 of the limit without crossing it and finishes at 69.9 % fines, because the nucleation gate of Equation (4.4) is 0.003 wide. The throughput optimum sits just below a limit whose position moves with temperature, which is why industrial practice holds a margin [19].

The argument for the third architecture appears when the plant differs from what the curve assumed. With the seed charge cut from 25 to 6 g, the fixed cubic curve gives 163.4 µm at 99.3 % fines and supersaturation control gives 587 µm at 47 % fines, taking 293 minutes to do it, from scenario 8 with the architecture changed or from scenario 6 with the seed changed. The controller has no measurement of the seed charge. It responds only to the supersaturation rising faster than the setpoint allows.

The cost of the third architecture is a dependency on the analyzer, and scenario 11 fails that instrument four ways against a cascade running the cubic curve, which never reads the analyzer and finishes in 187 minutes at 74 % valve travel. Each failure has its own signature. Noise of 0.5 g/kg raises valve travel from 66 to 935 % with the product unchanged at 373.6 µm, which is actuator wear attributable to the control structure. A bias of +3 g/kg makes the supersaturation read high, so the controller acts on driving force it does not have, cools too slowly, and overruns from 193.3 to 262.5 min at slightly lower valve travel, 60 against 66 %. Both together give 263.6 minutes and 1300 % of travel. An analyzer frozen at 100 min is worse than a biased one, because the reading holds at a value the controller has already reached, the error goes to zero, the cooling stops, and at the 300 minute cutoff the yield is 18.2 % instead of 62.2 with the temperature exactly on the setpoint the failed loop requested. The column of Volume 5, Chapter 3 measures what the lag of a composition analyzer alone takes from a loop.

4.6.5 Utility Temperature Limit

Scenario 10 takes the batch to 15 °C rather than 20 °C, which is worth 5.5 points of yield, 67.7 against 62.2 %, and then warms the chilled utility to 18 °C at $t = 110$ min. By Equation (4.8) the jacket inlet is bounded below by the utility temperature at any valve position, so a 15 °C batch is unreachable once the utility reaches 18 °C. The valve is pinned wide open for 133 min, the batch falls 3.9 °C behind its curve, and at the 300 minute cutoff the yield is 65.0 %. The crystals are unharmed at 378.8 µm and 14.6 % fines, because falling *behind* a cooling curve is always safe for the product. The

Table 4.9 Measured Complete-Batch Comparisons

Case	Law	Time min	L_{43} μm	Fines %	Wall excess min
Nominal	Cascade	187.0	373.3	13.8	0.0
	Supersat.	193.3	373.6	10.9	0.0
	MPC	126.9	372.5	15.6	0.0
Encrusted	Cascade	222.0	374.2	11.6	16.7
	Supersat.	195.0	375.0	11.0	0.0
	MPC	181.1	368.9	16.7	0.0
Analyzer fault	Cascade	187.0	373.3	13.8	0.0
	Supersat.	263.1	373.9	8.4	0.0
	MPC	152.6	372.5	13.2	0.0
Warm utility	Cascade	300[a]	378.8	14.6	0.0
	MPC	300[a]	378.4	15.7	0.0
Challenge	Cascade	172.4	373.5	15.1	0.0
	MPC	134.6	372.4	15.3	0.0

[a]Maximum clock, endpoint not reached. Wall excess is time above the plant's wall metastable limit, evaluated after the run.

same event on a batch taken only to 20 °C saturates the valve for 6.5 min, falls 0.56 °C behind and finishes on time at 62.2 % yield. The additional yield requires operating the jacket near its capacity, which leaves no margin for a warmer utility. A controller resting on a limit is no longer controlling, and the trend shows nothing abnormal because the batch is only running late.

4.6.6 The Graded Challenge

The graded challenge combines the full sequence of design decisions. It ships a 150 minute linear recipe on a slow single loop ($K_c = 5$, $\tau_I = 12$ min) in a vessel with half the heat transfer against a 10 °C utility, and it passes when the batch finishes within 180 minutes at a mean size of at least 320 µm, under 20 % fines and no more than 1.5 °C of wall subcooling. As shipped it finishes in 154 minutes at 148 µm, 172 µm short of the size target, and 97.1 % fines while holding its straight line to within 1.03 °C the whole way. The scenario's suggested settings, the cubic curve on a cascade with the wall clamped at 1.5 °C, stretched to a 165 minute curve, finish in 172 min at 374 µm, 15.1 % fines, a coefficient of variation of 0.32, 62.3 % yield, no time outside the zone and a peak wall subcooling of 1.44 °C. The margins on the four targets are 8 min, 54 µm, 4.9 points of fines and 0.06 °C. A single loop passes only in a narrow window, the cubic curve over 165 minutes at the shipped slow tuning, where the lag of the loop holds the wall at 1.48 °C and the batch finishes at 178 min. The same curve at $K_c = 12$ and $\tau_I = 5$ min takes the wall to 1.57 °C and fails. The order of importance is the trajectory, which decides the product, then the architecture, which decides which constraints can be stated, and the tuning last.

4.6.7 Predictive Control Comparison

Table 4.9 uses the same seed, instrument faults and endpoint rule for each structure. Cascade uses $K_c = 5$, $\tau_I = 4$ minutes and jacket $K_c = 5$, $\tau_I = 0.3$ minutes. Supersaturation control uses its shipped $\sigma_{SP} = 0.13$, $K_c^S = 80$, $\tau_I^S = 15$ minutes and 0.8 degree-per-minute rate limit. The MPC settings remain those of Section 4.5.4 in every case. Suggested challenge settings are retained before each structure change. The warm-utility rows end at the 300 minute maximum clock without reaching the required 15 degree endpoint. They are not completed batches.

The nominal gain costs valve movement and fines. Travel rises from 66.4 to 277.6% while fines rise from 13.8 to 15.6%, despite zero bulk or wall metastable-limit time under either law. Figure 4.4 shows the distinct cooling paths and cumulative movement, the cascade as a solid line and the MPC as a dashed one. On the encrusted vessel MPC eliminates the cascade's 16.7 minutes above the wall

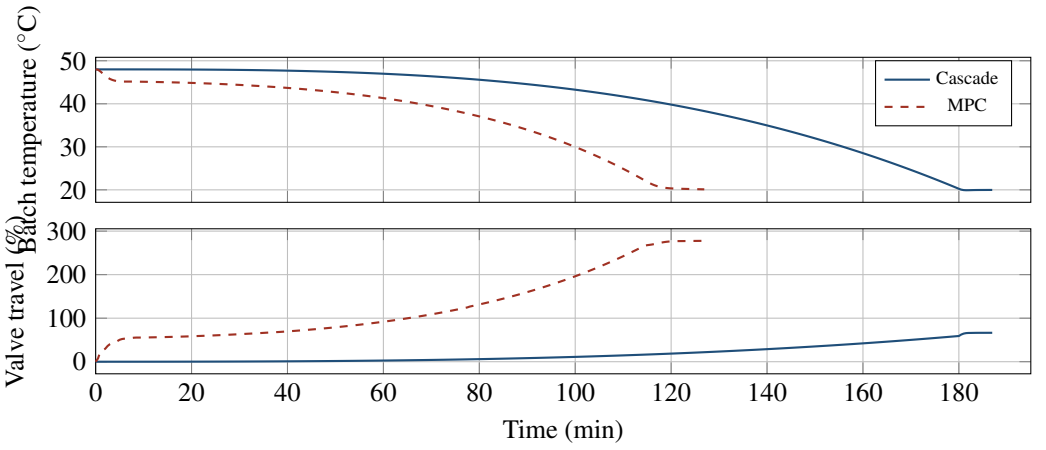


Figure 4.4 Nominal runs continued to their respective endpoints, cascade solid and MPC dashed. The 126.9 minute predictive batch moves the valve 277.6%, compared with 66.4% over the cascade's 187.0 minutes. Product quality is evaluated at those endpoints.

boundary but raises fines from 11.6 to 16.7%. The reduced model carries neither the population objective nor actual fouling, so its band constraints do not certify quality.

An eight-setting hindsight cascade grid varies recipe time over 120, 140, 160 and 180 minutes and wall subcooling over 1 and 1.5 degrees. Its fastest row finishes in 122.4 minutes with 97.1% fines. At a 160 minute recipe it finishes in 167.5 minutes with 16.7% fines. These rows delimit this grid, not the best achievable classical policy. Halving the integration step preserves the nominal ordering. Four independently stepped nominal and encrusted rows reproduce time, size and fines exactly. Across the 33-row comparison, no MPC solve is rejected. Chrome averages 0.685 milliseconds per plan over 360 solves, 0.137% of a core at unit playback and 8.22% at the maximum 60-fold playback.

4.7 Discussion

The phase diagram of Figure 4.5 is the display on which a crystallizer is read. A temperature trend shows temperature falling and a concentration trend shows concentration falling, and neither indicates whether the batch is near its limit, while the position of the operating point between the solubility curve and the metastable limit reports the margin directly. A second operating point at the same concentration and the wall temperature shows the wall crossing the limit on encrusting batches while the bulk remains inside the zone.

Crystal size distribution is not an internal number. It sets the cake resistance on the filter, the solvent retained through the dryer, the flow and compaction of the powder in a tablet press and the dissolution rate that governs bioavailability, which is why it is a registered attribute and why the crystallizer is the unit that fixes the solid form [2]. A batch making 96.3 % fines by number becomes a filtration problem, a drying problem and a formulation problem at once, hours after the crystallizer released it.

A compliant batch record does not carry that attribute. A record captures the setpoint, the measurement, the operator actions, the alarms and the deviations, every one of which is a signal the control system already had. Against a half-degree deviation limit the linear batch of Table 4.4 passes cleanly on 0.142 °C of peak deviation, is released by automated review, and produces a powder with 96.3 % fines. Nothing in the record is false and nothing in it is useful. An in-process attribute that responds to the failure has to come from somewhere else, and without a particle instrument the only candidate is concentration, which is the argument for measuring supersaturation even where it is not controlled [3, 19]. The application makes the same point about the temperature probe, where a +1 °C



Figure 4.5 Scenario 4 at the end of batch, 186.8 min after the seed was charged on the cubic cooling curve (Table 4.4), cropped to the plant view. The picture is a composite: three tiles of the app's performance panel are pasted into the empty band above the phase diagram. (1) The badges: a mean size of 373 µm, 13.8 % fines, a yield of 62.2 % and a variation coefficient of 0.31. (2) The pasted tiles, the loop and the batch side by side: 15.065 °C min of integrated error, 66.11 % of valve travel and a peak wall subcooling of 0.7 °C. (3) The vessel, its jacket and the utility valve. (4) The phase diagram shows the solubility curve, the metastable limit 5.5 °C to its left, and the path the batch took between them from 140 g/kg at 48 °C down to the endpoint at the lower left. The path never touches either line. (5) The product readout, with zero minutes outside the metastable zone.

bias makes the loop regulate the reading rather than the batch, so the batch ends at 18.9 °C instead of 19.9 and the yield rises 1.2 points with every number in the record on target.

4.8 Technology in Practice

What Measures It

Three readings close the loops of this chapter. TT-101 reads the batch temperature through a 0.3 min probe lag, the jacket detector of the cascade sits in the fast coolant stream and is treated as ideal, and AT-101 reads the liquor concentration, the one reading that sees the driving force. In situ sensors such as infrared spectrometers and laser backscattering probes became commercially available in the 1990s, and the review of the field names them first among the reasons crystallization control grew from then on [19]. Process analytical technology, in the words of the regulator, designs, analyzes and controls manufacturing through timely measurements of critical quality and performance attributes, with the goal of ensuring final product quality [22].

> **TECHNOLOGY** **Sensor: An Infrared Probe That Reads the Liquor**
>
> Ordinary infrared spectroscopy sends light through the medium, and in a slurry the crystals block the beam. An attenuated total reflection probe reflects the beam inside its own crystal, and only an evanescent wave reaches about 1 to 10 micrometers into the liquid, where few crystals intrude. A chemometric calibration relates the absorbances at many wavenumbers, together with the temperature, to the concentration. A background spectrum taken in humid laboratory air drifts, because water absorbs widely in the mid-infrared. Accuracies as high as 0.000 14 g of solute per g of solvent were reported by early 2002 [19]. AT-101 stands for such a probe as a 1 min lag with a settable bias and noise.

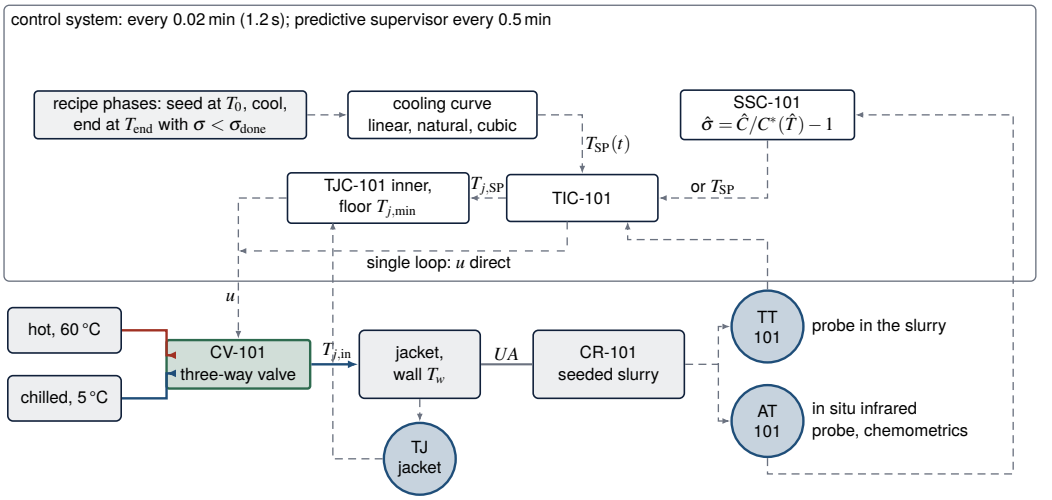


Figure 4.6 The chapter's loops on plant hardware. In the cascade, TIC-101 reads the batch probe TT-101 and sets the jacket setpoint of TJC-101, whose lower limit $T_{j,min}$ keeps the wall subcooling within its clamp, and TJC-101 moves the three-way valve CV-101. The single loop skips the inner loop and sends u to the valve directly (dashed). Supersaturation control replaces the programmed curve with a temperature setpoint computed from the infrared probe AT-101. Notice that the probe that sees the driving force enters only through SSC-101.

What Moves It

One valve moves everything. Where two streams mix, one three-way valve can replace a pair of complementary split-ranged valves [23], and CV-101 blends the hot and chilled utilities into the jacket inlet. The chilled utility's temperature is a floor that no valve position passes, which is the utility limit of scenario 10, and the valve's travel is a cost the chapter counts: analyzer noise of 0.5 g/kg raised it from 66 to 935 % of span with the product unchanged.

> **TECHNOLOGY** **Final Element: A Three-Way Mixing Valve**
>
> A three-way globe valve has three ports, and its plug sets how far two of them connect with the third, so it can split one stream or mix two into one. Mixing hot and cold water to control a temperature is a textbook use of it [23]. A pneumatic valve with excess stem friction moves in jerky steps and ignores small reversals of its signal, which is called stiction, and a loop that strokes a valve back and forth wears out its stem packing early [23]. CV-101 has neither friction nor travel time: 0 % sends hot utility and 100 % chilled. Its travel is counted, and the over-tuned row of Table 4.5 strokes it 9114 % of its span.

Where the Controller Runs

Batch operations follow recipes developed in the laboratory, a sequence of operations such as heating, cooling, reaction and crystallization performed in a pre-specified order [24], and they are controlled from stand-alone controllers, programmable logic controllers, distributed control systems and personal computers [24]. The batch control committee of ISA describes a functional model that relates recipe management, scheduling, sequential control, regulatory control and safety interlock systems [25]. Read in those terms, seeding, cooling and the end-of-batch rule are the sequential control of one recipe, the cooling curve and its time are recipe parameters, and TIC-101, TJC-101 and SSC-101 are the regulatory control. Supervisory and lower-level loops run with cycle times from milliseconds to minutes [26]. The controllers here execute every 0.02 min, which is 1.2 s, and the predictive supervisor plans every 0.5 min, as Figure 4.6 draws it on plant hardware.

IN PRACTICE **Supersaturation Control on Plant Hardware**

In concentration feedback control, the controller computes the supersaturation from the measured concentration, the temperature and the solubility curve, and sends a temperature setpoint to a lower-level temperature controller, typically a cascade of proportional-integral loops. The usual name, supersaturation control, is a misnomer, because a concentration is measured. The recipe becomes a path in the phase diagram, and the temperature profile follows from it on standard tracking control [19]. Regulatory guidance treats the end of a batch the same way: a process end point is not a fixed time but the achievement of the desired material attribute, within an evaluated window of process times [22]. SSC-101 over TIC-101 is that structure.

What Is Modernizing It

Three developments bear on this chapter. Direct nucleation control holds the number of crystals, counted in situ by focused beam reflectance, at a target through alternating phases of supersaturation and dissolution, and it adapts to variations in seed quality and to accidental seeding from crust with almost no prior model [19]. The supersaturation setpoint itself can be chosen by batch-to-batch iterative learning [19], which suits a process where remedial corrections are often impossible once off-specification crystals exist [24], and which Limitations notes this application lacks. Smart valve positioners record total stem travel and can predict when the stem packing will wear out [23], the hardware counterpart of the valve travel this chapter reports beside every batch.

4.9 Limitations

No number in this application is evidence about a real crystallization. The kinetic constants in Table 4.2 were chosen to make the mechanisms observable rather than fitted to a substance, so nothing reported here is a prediction about any material. There is no distribution shape, no agglomeration, no breakage, no growth rate dispersion, one polymorph and no impurity, which between them remove most of what makes an industrial crystallization difficult to develop. The vessel is perfectly mixed apart from a single wall film, so local supersaturation near the coolant surface, the effect of agitation rate on secondary nucleation and the suspension of the largest crystals are all absent. Encrustation is one fraction scaling one coefficient where real scale grows unevenly. AT-101 is one lag, one bias and band-limited noise, where a real in-situ probe reports through a chemometric model whose error depends on temperature, solid loading and the very supersaturation being measured, and it has no drift, no probe fouling and no recalibration. There is no particle instrument at all, so direct nucleation control, which is an industrial architecture beyond those implemented here [18], cannot be demonstrated. The predictive supervisor does not optimize the population moments, and no batch-to-batch learning of the kind in Sections 1.4.5 and 1.4.6 is implemented. The features that survived the simplification are the ones a controller has to be designed around, a setpoint that is a trajectory, a path constraint no instrument reports, a capacity that grows during the run, a constraint that binds first on an unmeasured surface, and an objective evaluated only at the end. Section 4.3.7 records where each assumption fails.

The predictive controller assumes the clean-vessel heat balance and a 0.3 minute closed-jacket response. Its three-state estimate does not reconstruct the instrument lags, crystal population or encrustation. Neither its soft constraints nor its bounded bias imply nonlinear invariance, terminal stability or product-quality guarantees.

OPEN CHALLENGE **Seeing the Crystals at Plant Solids Loadings**

Image analysis classifies crystals by size and shape, through a window or with an in situ video microscope. Most in situ imaging gives quantitative results only at low solids, typically 5 to 7 % at most, and future work aims at hardware and software that stay reliable in denser suspensions [19]. Guidance on process analytical technology names particle size and shape as attributes that are hard to sample representatively [22]. Progress would be a size measurement a loop could close on through the whole batch.

4.10 Conclusion

This chapter has used a seeded batch cooling crystallization as the example of a process that follows a trajectory rather than holding a setpoint and whose product is evaluated only after the run ends. The simulator integrates nine states, the batch and jacket temperatures, the dissolved solute, the first five moments of the crystal population under size-independent growth, and a wall encrustation fraction, with a metastable limit drawn 5.5 °C of subcooling above the solubility curve and primary nucleation gated on it at a fifth-power dependence. Three cooling curves between 48 °C and 20 °C in 180 minutes on the same single loop give 86.3, 159.2 and 373.4 µm of mass-weighted mean size at 99.3, 96.3 and 13.8 % fines, while the same three batches give peak tracking deviations of 0.637, 0.142 and 0.237 °C, so the best-tracked batch is the second-worst product. Sweeping the proportional gain from 2 to 30 moves the integrated error by a factor of 96 and the mean size by 0.7 µm, and pushing it to 60 damages the product through 9114 % of valve travel and 6.66 °C of wall subcooling rather than through the bulk supersaturation. Two measured results contradict the simplest explanation, that peak supersaturation orders the outcomes while cumulative time outside the zone does not, and that a 150 minute batch which never leaves the zone still finishes at 68.2 % fines from secondary nucleation alone. The kinetic constants are chosen to make the mechanisms observable rather than fitted to a substance, so no number reported here is a prediction about any material. The predictive jacket supervisor reduces the nominal batch from 187.0 to 126.9 minutes while increasing fines from 13.8 to 15.6% and valve travel from 66.4 to 277.6%. The app does not offer a particle count measurement, so direct nucleation control cannot be compared with supersaturation control, and it has no nonlinear model predictive control against the metastable zone path constraint, no moving horizon estimation of the growth and nucleation parameters, no solubility that depends on impurity level and no agglomeration. What the crystallizer teaches is the distinction between a controlled variable and a product objective.

KEY POINTS

- A seeded cooling crystallization is graded once, after the batch, on crystals that no instrument in the loop can see.
- Three cooling curves on the same loop give 86.3, 159.2 and 373.4 µm of mean size, and the best-tracked batch is the second-worst product.
- Peak supersaturation orders the outcomes and cumulative time outside the zone does not, and a 150 minute batch that never leaves the zone still finishes at 68.2 % fines.
- The trajectory decides the product, the architecture decides which constraints can be stated, and the tuning comes last.

Where else this applies. The autoclave of Chapter 6 also runs a recipe of temperature ramps whose product is decided inside the load: the recipe is run on the part that lags and limited by the part that leads, as this batch is limited by its coldest surface, the wall.

References

[1] 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.

[2] N. Variankaval, A. S. Cote, and M. F. Doherty, “From form to function: Crystallization of active pharmaceutical ingredients,” *AIChE Journal*, vol. 54, no. 7, pp. 1682–1688, 2008.

[3] L. X. Yu, R. A. Lionberger, A. S. Raw, R. D’Costa, H. Wu, and A. S. Hussain, “Applications of process analytical technology to crystallization processes,” *Advanced Drug Delivery Reviews*, vol. 56, no. 3, pp. 349–369, 2004.

[4] D. Bonvin, “Optimal operation of batch reactors: a personal view,” *Journal of Process Control*, vol. 8, no. 5–6, pp. 355–368, 1998.

[5] J. B. Rawlings, S. M. Miller, and W. R. Witkowski, “Model identification and control of solution crystallization processes: a review,” *Industrial & Engineering Chemistry Research*, vol. 32, no. 7, pp. 1275–1296, 1993.

[6] A. D. Randolph and M. A. Larson, *Theory of Particulate Processes: Analysis and Techniques of Continuous Crystallization*, 2nd ed. San Diego, CA: Academic Press, 1988.

[7] J. W. Mullin, *Crystallization*, 4th ed. Oxford, UK: Butterworth-Heinemann, 2001.

[8] B. W. Bequette, *Process Control: Modeling, Design, and Simulation*. Upper Saddle River, NJ: Prentice Hall, 2003.

[9] H. A. Miers and F. Isaac, "The refractive indices of crystallising solutions, with especial reference to the passage from the metastable to the labile condition," *Journal of the Chemical Society, Transactions*, vol. 89, pp. 413–454, 1906.

[10] J. Nývlt, R. Rychlý, J. Gottfried, and J. Wurzelová, "Metastable zone-width of some aqueous solutions," *Journal of Crystal Growth*, vol. 6, no. 2, pp. 151–162, 1970.

[11] N. Kubota, "A new interpretation of metastable zone widths measured for unseeded solutions," *Journal of Crystal Growth*, vol. 310, no. 3, pp. 629–634, 2008.

[12] J. W. Mullin and J. Nývlt, "Programmed cooling of batch crystallizers," *Chemical Engineering Science*, vol. 26, no. 3, pp. 369–377, 1971.

[13] A. G. Jones, "Optimal operation of a batch cooling crystallizer," *Chemical Engineering Science*, vol. 29, no. 5, pp. 1075–1087, 1974.

[14] N. Doki, N. Kubota, A. Sato, and M. Yokota, "Effect of cooling mode on product crystal size in seeded batch crystallization of potassium alum," *Chemical Engineering Journal*, vol. 81, no. 1–3, pp. 313–316, 2001.

[15] N. Kubota, N. Doki, M. Yokota, and A. Sato, "Seeding policy in batch cooling crystallization," *Powder Technology*, vol. 121, no. 1, pp. 31–38, 2001.

[16] H. Grön, A. Borissova, and K. J. Roberts, "In-process ATR-FTIR spectroscopy for closed-loop supersaturation control of a batch crystallizer producing monosodium glutamate crystals of defined size," *Industrial & Engineering Chemistry Research*, vol. 42, no. 1, pp. 198–206, 2003.

[17] M. Fujiwara, Z. K. Nagy, J. W. Chew, and R. D. Braatz, "First-principles and direct design approaches for the control of pharmaceutical crystallization," *Journal of Process Control*, vol. 15, no. 5, pp. 493–504, 2005.

[18] A. N. Saleemi, C. D. Rielly, and Z. K. Nagy, "Comparative investigation of supersaturation and automated direct nucleation control of crystal size distributions using ATR-UV/vis spectroscopy and FBRM," *Crystal Growth & Design*, vol. 12, no. 4, pp. 1792–1807, 2012.

[19] Z. K. Nagy and R. D. Braatz, "Advances and new directions in crystallization control," *Annual Review of Chemical and Biomolecular Engineering*, vol. 3, no. 1, pp. 55–75, 2012.

[20] 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.

[21] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[22] U.S. Food and Drug Administration, "Guidance for industry. PAT: A framework for innovative pharmaceutical development, manufacturing, and quality assurance," U.S. Department of Health and Human Services, Food and Drug Administration, Rockville, MD, Tech. Rep., Sep. 2004.

[23] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[24] D. Bonvin and G. François, "Control and optimization of batch chemical processes," in *Coulson and Richardson's Chemical Engineering*. Elsevier, 2017, pp. 441–503.

[25] International Society of Automation, "ISA88: Batch control, committee scope and purpose," www.isa.org, standards committees, 2026, accessed 22 November 2026.

[26] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

4.11 Exercises

4.1 Concept check · 1 min

Scenarios 2, 3 and 4 run the linear, natural and cubic cooling curves between the same endpoints in the same 180 minutes, on the same single temperature loop with the same tuning. Which statement matches the results?

(a) The linear batch tracks its curve most closely and therefore gives the coarsest crystals

(b) The cubic batch gives the coarsest crystals although it tracks its curve worse than the linear one

(c) The three batches give nearly the same crystals, because they share the endpoints and the yield

(d) The natural batch gives the coarsest crystals, because it cools fastest while the seed is fresh

4.2 Concept check · 1 min

Why can the app carry the whole crystal population with the five ordinary differential equations of Equation (4.5) instead of a partial differential equation in size and time?

(a) Nucleation is negligible inside the zone, so only growth has to be carried

(b) The size distribution is assumed log-normal for the whole batch, like the seed

(c) Growth does not depend on size and nuclei are born at one size, so the moments close

(d) The fines are counted separately by number, so the distribution needs no shape

4.3 Concept check · 2 min

Table 4.6 lists five batches with the same kinetics in the same vessel, from the cubic batch on 25 g of seed to the batch cooled by hand with the valve wide open. Which single statistic orders their fines across all five?

(a) The time the bulk supersaturation spends outside the metastable zone

(b) The integrated absolute error of the batch temperature loop

(c) The total travel of the coolant valve over the batch

(d) The peak value of the bulk relative supersaturation

4.4 Theory and calculation · 6 min

The solubility of Equation (4.1) is anchored at $C^* = 52\,\mathrm{g/kg}$ at 20 °C and 140 g/kg at 50 °C, with the curvature $a_2 = 4.0 \times 10^{-5}\,\mathrm{kg/(kg\,°C^2)}$. The charge holds $C = 0.140\,\mathrm{kg/kg}$ and is seeded at 48 °C. The metastable zone is $\Delta T_{\mathrm{msz}} = 5.5\,°\mathrm{C}$ wide.

(a) Find the coefficients a_1 and a_0. *Checkpoint:* $a_1 = 1.333 \times 10^{-4}\,\mathrm{kg/(kg\,°C)}$ and $a_0 = 3.333 \times 10^{-2}\,\mathrm{kg/kg}$.

(b) Find the solubility and the relative supersaturation at the moment of seeding. *Checkpoint:* $C^* = 0.1319\,\mathrm{kg/kg}$ and $\sigma = 0.0615$.

(c) Find the metastable limit $\sigma_{\max}$ of Equation (4.2) at 48 °C and at 20 °C, and say which end of the batch is the tight one. *Checkpoint:* 0.175 at 48 °C and 0.207 at 20 °C. The warm end is tight.

4.5 Theory and calculation · 7 min

At the instant of seeding the batch is at 48 °C with $\sigma = 0.0615$ and a seed surface $\mu_2 = 6.82 \times 10^{-3}\,\mathrm{m^2/kg}$. Take $k_g = 22\,\mu\mathrm{m/min}$, $g = 1.10$, $\rho_c = 1300\,\mathrm{kg/m^3}$, $k_v = 0.50$ and the solubility coefficients of Table 4.2 and the previous exercise.

(a) Find the growth rate G of Equation (4.3). *Checkpoint:* 1.02 µm/min.

(b) Find $\mathrm{d}C^*/\mathrm{d}T$ at 48 °C and the cooling rate the seed can absorb at constant supersaturation, Equation (4.12). *Checkpoint:* 3.973×10^{-3} kg/(kg °C) and 0.0032 °C/min.

(c) For the base recipe, 48 °C to 20 °C in 180 min, find the initial cooling rate of the linear and the natural profiles of Equation (4.11) and their ratios to the result of (b). Give the initial and final

rates of the cubic profile. *Checkpoint:* 0.156 and 0.724 °C/min, ratios 48 and 224. Cubic 0 and 0.467 °C/min.

4.6 Theory and calculation · 5 min

The charge is $M = 114.0\,\mathrm{kg}$ of slurry with $c_p = 3.80\,\mathrm{kJ/(kg\,°C)}$ behind a jacket of $m_j = 15\,\mathrm{kg}$, $c_j = 3.60\,\mathrm{kJ/(kg\,°C)}$ and $F_j = 40\,\mathrm{kg/min}$. The process-side share of the wall temperature difference is $f_w = 0.30$.

(a) With $UA = 90\,\mathrm{kJ/(min\,°C)}$, find the time constant of the contents and of the jacket, and their ratio. *Checkpoint:* 4.81 min and 0.231 min, about 21 to 1.

(b) Find how far below the measured batch temperature the cascade's jacket floor of Equation (4.16) sits for wall limits of 1.5 and 1.0 °C. *Checkpoint:* 5.0 °C and 3.33 °C.

(c) In the glass-lined vessel of scenario 12, $UA = 18\,\mathrm{kJ/(min\,°C)}$. With the jacket held on the 1.0 °C floor and a clean wall, estimate the heat removed and the cooling rate it buys, ignoring the heat of crystallization and the ambient leak. Compare it with the final rate of the cubic curve. *Checkpoint:* 60 kJ/min and about 0.14 °C/min, against 0.467 °C/min.

4.7 App challenge · 8 min

Load scenario 2, Linear cooling, the curve everybody writes first. Keep Auto, the shipped single loop and the default seed, and set the speed to 20, which runs a 180 min batch in about 9 s.

(a) Run until the batch pauses itself. Read IAE on the scorecard and the Mean size L_{43}, Fines and Batch time tiles. *Checkpoint:* IAE 13.434 °C min, 159 µm, 96.3 %, 179 min.

(b) Load scenario 4, Cubic cooling, matched to the growing surface, keep its settings and the speed of 20, and run until the batch pauses itself. Read the same four values. *Checkpoint:* IAE 15.065 °C min, 373 µm, 13.8 %, 187 min.

(c) Explain in two sentences why the batch with the larger integrated error gives the better product. *Checkpoint:* The product follows the supersaturation the trajectory imposes, not the tracking error.

Give IAE to within 0.01 °C min, sizes to within 1 µm and fines to within 0.1 point.

4.8 App challenge · 9 min

Load scenario 12, A glass-lined vessel, and the wall that scales, which ships $UA = 18\,\mathrm{kJ/(min\,°C)}$ on the single loop. Keep Auto and the default seed, and set the speed to 20.

(a) Run until the batch pauses itself. Read the Peak wall ΔT, Mean size L_{43}, Fines and Batch time tiles, the IAE on the scorecard and the last value of the column `foul_frac` in the exported CSV. *Checkpoint:* 4.3 °C, 375 µm, 13.9 %, 189 min, IAE 30.42 °C min, 0.312.

(b) Press Reset, set Control architecture to Cascade and Wall subcooling limit to 1.0 °C before pressing Run, and run until the batch pauses itself. Read the same values. *Checkpoint:* 1.0 °C, 374 µm, 9.9 %, 259 min, IAE about 625 °C min, 0.

(c) The IAE of the cascade batch is about twenty times that of the single loop. Is the cascade mistuned? *Checkpoint:* No. The clamp is active and the jacket cannot remove heat any faster.

Give temperatures to within 0.1 °C, sizes to within 1 µm, times to within 1 min and the scaled fraction to within 0.005.

4.9 App challenge · 8 min

Load scenario 13, A plan inside the metastable band, which opens on the MPC architecture. Keep Auto and the default seed, and set the speed to 20.

(a) Run until the batch pauses itself. Read the Batch time, Mean size L_{43} and Fines tiles and the last value of the column `valve_travel_pct` in the exported CSV. *Checkpoint:* 127 min, 372 µm, 15.6 %, 277.6 %.

(b) Press Reset, set Control architecture to Cascade before pressing Run, and run until the batch pauses itself. Read the same values. *Checkpoint:* 187 min, 373 µm, 13.8 %, 66.4 %.

(c) The MV travel on the scorecard of run (a) reads about 156, well below the CSV value. Explain the difference, and say which figure to quote for valve wear. *Checkpoint:* The scorecard sums the five logged samples a minute, and the CSV column accumulates every integration step. Quote the CSV.

Give times to within 1 min, fines to within 0.1 point and travel to within 1 %.

4.10 Application area · 3 min

A plant's automated review releases any batch whose temperature stayed within 0.5 °C of its programmed curve. Use the linear batch of Table 4.4 to explain why such a release can be wrong, and name the measurement on CR-101 that could have flagged the batch while it was running.

4.11 Application area · 4 min

A production planner proposes to finish every batch at 15 °C instead of 20 °C for the extra yield. The chilled utility is normally at 5 °C and can reach 18 °C on a summer afternoon.

(a) Estimate the yield at each endpoint from the solubility curve alone, with $C_0 = 0.140\,\mathrm{kg/kg}$ and the coefficients of the first theory exercise. *Checkpoint:* 62.9 % at 20 °C and 68.3 % at 15 °C.

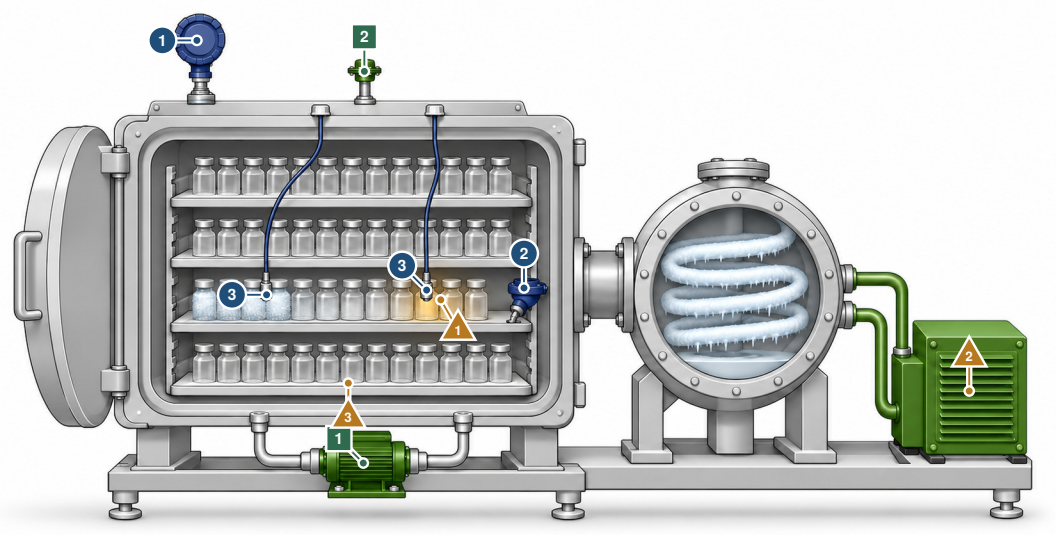
(b) Using scenario 10 as reported in the chapter, say what a warm afternoon does to the 15 °C batch and what the operator sees. *Checkpoint:* The batch cannot finish. The valve sits wide open and the trend only shows the batch running late.

4.12 Review the agent · 5 min

An AI assistant reviews scenario 4 and writes: "The cubic batch has an IAE of 15.07 °C min against 13.43 °C min for the linear batch, so the loop is too slow for the accelerating setpoint. Raise the proportional gain from 12 to 60 and cut the integral time from 5 to 0.5 min. Tighter tracking of the cubic curve will push the mean size above 373 µm." Find the flaw and say how you would check it.

CHAPTER 5

Hottest and Slowest Vials in an Endpoint-Aware Freeze-Drying Simulator



The controller moves 1 heating and cooling fluid for the shelves 2 nitrogen bleed that sets the chamber pressure

The controller reads 1 chamber pressure gauge 2 shelf temperature 3 two probes in sample vials

What disturbs it 1 vials that dry unevenly 2 condenser cooling falling short 3 poor contact under the vials

Vials on one shelf do not dry alike, so the warmest set how hot the shelf may run and the slowest set when drying may end. Illustration, not to scale.

IN PLAIN TERMS Drying Medicine in Bottles Without Spoiling Any

The main idea. Some medicines are frozen in small glass bottles. The ice then turns straight into vapor in a chamber pumped almost empty of air, while a warm shelf feeds in the heat.

The challenge. Bottles at the edge of the shelf get extra heat and can melt their dry cake, while bottles in the middle dry last. The sensors watch only two bottles and the gas, so the cycle must guard the hottest and wait for the slowest.

Why it matters. A cycle set for an average bottle spoils some bottles or wastes many hours.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 5.2, the fact in Section 5.3, the two try-its and the second fact in Section 5.6, 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 5.6), Technology in Practice (Section 5.7), Limitations (Section 5.8) and the open challenges

The control question A shelf of identical vials does not dry alike: the outer edge runs warmest and the center dries last, while the instruments see two probed vials or the whole chamber. How should a freeze-drying cycle set the shelf, end primary drying and finish secondary drying so that both the hottest and the slowest vials pass?

Learning objectives

- Explain why the hottest group limits the shelf and the slowest group decides the endpoint, and size the heat difference from the K_v fits.
- Compute the vapor the duct carries and show when the chamber pressure is lost to the vapor.
- Compare a recipe clock, product probes and the Pirani comparison as endpoints of primary drying.
- Assess what an estimator with a learned resistance gains and what a wrong gauge law costs.
- Judge whether a predictive controller's gain comes from its horizon or from the pressure it is allowed to set.

Prerequisites Sections 1.4.1, 1.4.2, 1.4.4, 1.6.2 and 1.6.4. Section 1.6.3 for the predictive option only.

The app https://apmonitor.com/apps/pharma-lyophilization/

See also Volume 5, Chapter 10 for moisture and a glass transition inferred from slow measurements, and Volume 5, Chapter 11 for a product temperature no instrument reaches. Volume 9, Chapter 4 for offset-free predictive control with a disturbance model.

5.1 Introduction

Freeze-drying removes the water from a frozen solution by subliming its ice under vacuum and then desorbing the water that never froze, and it leaves a porous cake that keeps a heat-sensitive medicine stable on the shelf. It is also slow. The standard operation of a batch freeze-dryer is protocol driven, with every phase programmed at fixed times and its process parameters held constant or interpolated between setpoints, which has been shown to be time consuming and inefficient [1]. In the pilot batch simulated here, 3066 vials each hold 3 mL of a 3 % sucrose solution, 8.90 kg of ice in all. A conservative recipe holds the shelf at −28 °C for 40 h of primary drying. The outer-edge vials are free of ice at 21.2 h and the center vials at 25.1 h, so the last 14.9 h of primary drying dry nothing, and the batch is stoppered at 50.0 h.

One belief is the target of the simulator, that a batch of identical vials on one shelf at one pressure behaves like one vial. The vial heat transfer coefficient has contributions from contact, conduction through the gas under the vial and radiation, and it depends on the chamber pressure and on the vial [2]. Vials at the front and side of an array sublime faster than those in the center because of atypical radiation from the surroundings [3]. At 10 Pa the fitted coefficient of an outer-edge vial is 24.2 W/(m^2 K) against 18.4 W/(m^2 K) for a center vial, 32 % more heat for the same shelf. The outer edge therefore limits how warm the shelf may run while the center decides when primary drying may end, and a recipe tuned on an average vial gets both decisions wrong. A transfer function from shelf to product temperature carries neither.

The class is general. A population of units in one batch, a limit set by its most exposed members and a completion set by its slowest, observed through a few sensors placed where someone could reach and through one slow aggregate signal, describes kiln and oven loads, battery formation, grain bins and other tray-loaded dryers. In Chapter 6 the coldest thermocouple sets the ramp of a composite cure while the hottest one sets its limit, and in Volume 5, Chapter 11 the temperature that decides the product is inside a brick that no instrument reaches. Freeze-drying makes the class concrete with published data. Product thermocouples sample a few vials, and the comparison of a Pirani gauge with a capacitance gauge has been found the best of several methods for detecting the end of primary drying [4]. Monitoring tools for product temperature, heat flux and sublimation flow have been reviewed for both laboratory and production scale [5], and model-based controllers for primary drying include in-line optimization of the heating fluid [6], model predictive control [7], a risk-based real-time strategy with model uncertainty [1] and multi-vial control validated on a pilot unit [8].

These studies report one dryer and one formulation at a time, and none lets a reader see the edge vials collapse while the center vials still hold ice, then change the policy, the endpoint and the secondary-drying plan and see which group fails. The simulator of this chapter is that experiment, a shelf-loaded batch of vials that keeps each group's true ice, front temperature and moisture apart from what its instruments and estimator report. It measures how a fixed recipe, probe-limited feedback and an endpoint-aware policy trade cycle time against the hottest and slowest groups. A fourth policy, model predictive control on the estimator's model, is measured against the same three, against the endpoint-aware policy retuned once by the controller's own tuning rule, and under 18 fault and mismatch conditions on two loads. The sublimation model, the vial heat transfer fits, the Pirani endpoint, the estimator, the shelf policies and the predictive controller follow the cited work. The measured results are specific to this model and calibration. Among them, an endpoint-aware policy stoppers the challenge batch 1.38 h sooner than a fixed recipe tuned on the same batch, but only 0.31 h of the gain comes from primary drying, and the same settings collapse every group when the Pirani law the estimator inverts is off by 6 %. What the predictive controller measurably adds is the chamber pressure handled without a rule, not cycle time.

The chapter runs as follows. Section 5.2 follows the domain from the first freeze-dried sera to model-based cycles. Section 5.3 sets out the batch as a five-group model whose ice and energy ledgers close to 1.2×10^{-14} relative over a whole cycle, with a quasi-steady chamber balance and the full equation set from which the simulator can be rebuilt. Section 5.4 gives the loops, the four shelf policies, the estimator, the endpoint rules and an ideal-information bound, with the predictive controller in Section 5.4.4. Section 5.5 separates code verification from agreement with the calibration study. Section 5.6 measures the nine cases and the challenge, the sensitivity to the batch and to the gauge law, and the predictive controller against the rule-based policies. Section 5.7 describes the gauges, the shelf fluid and the control system of a real dryer. Section 5.8 states the limitations, and the Exercises use the cases, each of which carries a completion rule (Section 1.6.4).

5.2 From Lyophile Serum to Model-Based Cycles

Drying biological material from the frozen state is older than its pharmaceutical use. An improved method of desiccation for biological problems was reported in 1909 [9], and in 1935 a procedure of rapid freezing and rapid dehydration from the frozen state, with the product vacuum sealed in the container in which it is stored and distributed, was described in detail for sera, antisera, enzymes and viruses, with apparatus for research laboratories, hospitals and industrial-scale production [10]. Drying in the final container, the arrangement the simulator models with vials, was part of the process from that start.

The physics of the vial came later. A 1984 study divided the resistance to vapor flow among the dried product, the semistoppered vial and the chamber, found the stopper and chamber resistances about a factor of ten below the dried product's, and described the heat reaching the vial through a heat transfer coefficient with three parallel mechanisms, each significant, whose values differ between vials of nominally the same specification while vials from one lot vary by about 4 % [2]. The product limit is a property of the formulation and of how it is measured. The collapse temperature in freeze-drying depends on the measurement methodology and on the rate of water removal from the glassy phase [11], and water plasticizes amorphous pharmaceutical solids in a way that a simplified Gordon–Taylor relationship describes [12], which is the law the simulator uses for the cake once its ice is gone.

Regulators wrote the practice down before the models were common, as the history segment below tells.

> ⧖ HISTORY **1993: An Inspection Guide Writes Down the Cycle**
>
> The United States inspection guide for freeze-dried injectables of 1993 set down what an inspector should expect of a cycle. After freezing and during primary drying, it is desirable to hold the product at least 4 to 5 degrees below its eutectic point. The dryer should control and record the shelf temperature, the product temperature, the condenser temperature, the chamber pressure and the condenser pressure, and the manufacturing directions should give time, temperature and pressure limits for the cycle. Cycles once run by cam-type recorder-controllers were giving way to microcomputer control, with the operator loading the chamber, placing probes in product vials and entering the cycle's parameters. The guide also recorded metal trays left under the vials that warped at one manufacturer and caused a moisture problem in some dosage units of a batch [13], the failure the simulator's poor-contact case represents.

The position effect was traced to its cause in 2003, when gold-coated vials and radiation shields showed that atypical radiation makes front and side vials sublime faster than center vials [3]. Feedback on the whole batch followed. Manometric temperature measurement coupled to an expert system set the shelf temperature, found drying end points and estimated residual moisture in a single laboratory run [14], and the Pirani-against-capacitance comparison was judged the best batch method of those tested for the end of primary drying, a cheap and steam-sterilizable instrument [4]. Model-based control arrived in the same decade, as in-line optimization of the heating fluid on a small industrial apparatus [6] and as model predictive control of primary drying [7]. A design space built from separately measured heat transfer coefficients and mass transfer resistances and checked by two runs [15], and a quality-by-design treatment of secondary drying that fitted Arrhenius desorption kinetics for a sucrose solution [16], carried the same models into cycle development. The current answer combines them. A risk-based strategy that propagates the uncertainty of the heat transfer and resistance fits controls the edge vials first and the center vials after, and starts secondary drying only when both the uncertainty analysis and the Pirani comparison indicate the end of primary drying [1], and multi-vial control has been validated experimentally on a pilot-scale unit [8]. The simulator's primary drying is calibrated on the first of these two.

5.3 The Batch as Modeled

The simulated dryer has four shelves, three loaded by default, each holding 1022 10R vials in a hexagonal pack, as shown in Figure 5.1. The shelves are heated and cooled by a fluid from a thermal unit, the chamber is held at its setpoint by bleeding nitrogen in, and the vapor leaves through a duct to a condenser coil and a vacuum pump. A shelf is resolved into three groups, 126 outer-edge vials that see the door and walls, 118 inner-edge vials in the second row and 778 center vials, and two more vials carry the product thermocouples. Each group is one representative vial multiplied by its count, so a batch is five vial models. Every case starts at a documented checkpoint at the end of freezing and runs through a 1 h vacuum hold, primary drying, an optional soak and secondary drying to stoppering. Freezing itself is not simulated.

The variables are classified in Table 5.1. The states are the ice $m_{i,j}$, bottom temperature $T_{b,j}$ and bound water X_j of each of the five groups, the shelf fluid temperature T_s, the condenser coil temperature T_{cd} and its ice m_{cd}, and the bleed valve position u_b, nineteen in all, plus ledgers that integrate the sublimed ice, the heat in, the sensible heat, the desorbed water, the water reaching the condenser and the thermal duties. Units are SI except temperatures, which the states carry in °C and every physical relation converts to kelvin.

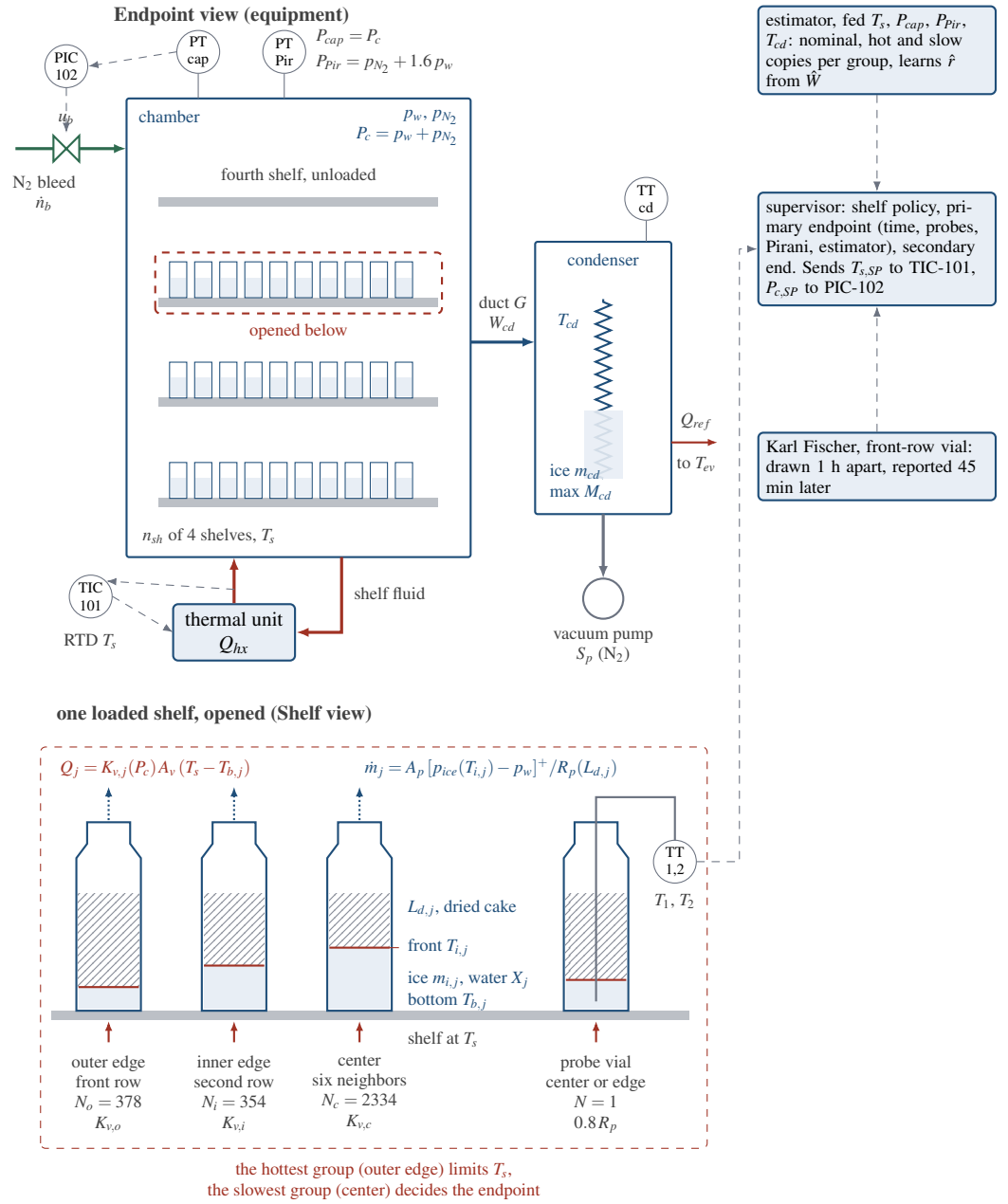


Figure 5.1 Freeze-dryer and vial groups as modeled. The chamber holds n_{sh} loaded shelves at T_s, water vapor p_w and nitrogen p_{N_2} from the bleed u_b under PIC-102, and the vapor crosses a duct of conductance G to the condenser coil at T_{cd} with ice m_{cd} and refrigeration Q_{ref}. One loaded shelf is opened below into its outer-edge, inner-edge and center groups and a probe vial, each with heat Q_j through $K_{v,j}$, a frozen layer $m_{i,j}$ below the front at $T_{i,j}$, a dried cake of thickness $L_{d,j}$ and a bottom at $T_{b,j}$. Heat flows are solid arrows, vapor dotted arrows and measurements dashed lines, and model states are labeled in blue. Process model only, not an equipment drawing. Not to scale.

Table 5.1 Roles of the Variables, With Checkpoint Values From Case 1

Variable	Role	Checkpoint	Range
T_s	CV of TIC-101 (RTD)	−40.0 °C	−55 to 50
P_c	CV of PIC-102 (capacitance)	10.0 Pa SP	5–30 Pa
$\max(T_1, T_2)$	CV of TIC-103 (probes)	−47.9 °C	—
Q_{hx}	MV, thermal-unit duty	−0.04 kW	−2.5 to 3.0 kW
u_b	MV, nitrogen bleed	20.6 %	0–100 %
$T_{s,SP}$	MV of the supervisor	−40.0 °C	ramp-limited
contact	DV, center K_v	100 %	60–110 %
r	DV, product resistance	100 %	60–160 %
b_T	DV, probe bias	0 K	±3 K
condenser	DV, refrigeration available	100 %	20–100 %
$m_{i,j}$	states, ice per vial	2.903 g	0–2.903
$T_{b,j}$	states, vial bottom	−47.2 to −47.9 °C	—
X_j	states, bound water	0.08 kg/kg	≥ 0
T_{cd}, m_{cd}	states, condenser	−75.6 °C, 0 kg	≤ 15 kg
p_w, p_{N_2}	algebraic, chamber	3.99, 6.01 Pa	—
$T_{i,j}$, $\dot{m}_j$	algebraic, front and flow	−48.3 °C (center)	—

5.3.1 Sublimation Through the Dried Layer

Ice at the front sublimes at a rate set by the difference between its vapor pressure and the chamber's water vapor, divided by the resistance of the dried layer above it, as shown in Equation (5.1).

$$\begin{aligned}
\dot{m}_j &= \frac{A_p\,\varphi_j\left[p_{ice}(T_{i,j}) - p_w\right]^+}{R_{p,j}} \\
R_{p,j} &= r\,f_j\left(R_{p0} + \frac{A_{Rp}L_{d,j}}{1 + B_{Rp}L_{d,j}}\right) \\
L_{d,j} &= \frac{m_{i0} - m_{i,j}}{\rho_{ice}\,\varepsilon\,A_p}, \qquad \varphi_j = \sqrt{\min(1,\ m_{i,j}/m_{tail})} \\
\ln p_{ice} &= 9.550426 - \frac{5723.2658}{T} + 3.53068\ln T \\
&\quad - 0.00728332\,T
\end{aligned} \tag{5.1}$$

where $\dot{m}_j$ is in kilograms per second per vial, $A_p = \pi r_i^2$ is the product area, R_p is area-normalized in Pa m^2 s/kg (which reduces to the m/s the source prints), r is the batch resistance disturbance and $f_j = 0.8$ for the two probe vials and 1 otherwise. The bracket is never negative, so no ice forms on the cake, and p_{ice} is in pascals with T in kelvin. The factor φ_j lets the last 1 % of the ice, $m_{tail} = 0.01\,m_{i0}$, leave from a shrinking area, so a vial warms before its last ice is gone and the ice reaches zero in finite time. The resistance rises from 1.51×10^4 at zero thickness to 1.043×10^5 Pa m^2 s/kg at the full 8.56 mm layer, which is why a fixed shelf runs the front warmer late in primary drying than early.

5.3.2 Heat Into the Vial

The shelf heats the vial bottom, the bottom sits above the front by the conduction drop through the remaining ice, and the vial's heat capacity is kept, as shown in Equation (5.2).

$$\begin{aligned} Q_j &= K_{v,j}(P_c)\,A_v\,(T_s - T_{b,j}), \quad K_v = \alpha + \frac{\beta P_c}{1+\gamma P_c} \\ T_{i,j} &= T_{b,j} - \frac{Q_j L_{ice,j}}{\lambda_{ice} A_p}, \qquad L_{ice,j} = \frac{m_{i,j}}{\rho_{ice}\,\varepsilon\,A_p} \\ C_j \frac{\mathrm{d}T_{b,j}}{\mathrm{d}t} &= Q_j - \Delta h_{sub}\dot{m}_j - \Delta h_{des} r_{d,j} \\ \frac{\mathrm{d}m_{i,j}}{\mathrm{d}t} &= -\dot{m}_j \\ C_j &= m_{i,j}c_{ice} + m_s c_s + m_s X_j c_w + C_{gl} \end{aligned} \tag{5.2}$$

where $A_v = \pi r_o^2$, $c_w = 4180\,\mathrm{J/(kg\,K)}$ and $r_{d,j}$ is the desorption rate of Equation (5.4). The coefficients α, β and γ are fitted separately for center and edge vials in the calibration study. The inner edge sits 35 % of the way from center to edge, $K_{v,i} = K_{v,c} + 0.35(K_{v,o} - K_{v,c})$, a teaching interpolation, and the contact disturbance scales $K_{v,c}$ alone. The coefficient is gravimetric, so it already contains the radiation the edge vials receive and no separate radiation term is added. Because K_v depends on P_c, P_c on p_w and p_w on every front temperature, the three are solved by three fixed-point passes at every evaluation.

5.3.3 The Chamber, the Condenser and the Bleed

The chamber's water vapor holds about 0.4 s of flow against a 10 s step and hours of drying, so its balance is solved at quasi-steady state, as shown in Equation (5.3).

$$\begin{aligned} W_{cd} &= \sum_j N_j(\dot{m}_j + r_{d,j}) \\ &= \frac{M_w G\,\phi_{cd}}{R\,T_{gas}}\big(p_w - p_{ice}(T_{cd})\big) \\ p_{N_2} &= \frac{u_b\,\dot{n}_b\,R\,T_{gas}}{S_p}, \qquad P_c = p_w + p_{N_2} \\ C_{cd}\frac{\mathrm{d}T_{cd}}{\mathrm{d}t} &= \Delta h_{sub} W_{cd} + UA_{lk}(T_{wall} - T_{cd}) - Q_{ref} \\ Q_{ref} &= \mathrm{clip}\big(UA_{coil}(T_{cd} - T_{ev}),\ 0,\ Q_{ref,max}\big) \\ \frac{\mathrm{d}m_{cd}}{\mathrm{d}t} &= W_{cd}, \qquad \tau_v \frac{\mathrm{d}u_b}{\mathrm{d}t} = u_{b,cmd} - u_b \\ C_{sh}\frac{\mathrm{d}T_s}{\mathrm{d}t} &= Q_{hx} - \sum_j N_j Q_j + UA_w(T_{wall} - T_s) \end{aligned} \tag{5.3}$$

where N_j is the count of group j, $\phi_{cd} = \mathrm{clip}\big((1.1M_{cd} - m_{cd})/(0.2M_{cd}),\ 0.05,\ 1\big)$ narrows the vapor path as the coil approaches its ice capacity, and C_{sh} is 50 kJ/K per loaded shelf. The first line is linear in p_w for the groups whose front is warmer than the chamber vapor, so it is solved exactly with an active set that drops a group whose ice is colder. The bleed can only add gas, so when the vapor alone exceeds the setpoint the valve shuts and the vapor load, the duct and the coil decide the pressure. At 10 Pa of vapor with the coil at −73.2 °C the declared duct carries 0.654 kg/h, the equipment edge of case 3. The chamber pressure is not a free knob in this model.

+ DID YOU KNOW? A Speed Limit in the Duct

As the sublimation rate rises, the vapor flowing to the condenser approaches the speed of sound, about 350 m/s for water vapor at room temperature, and the flow then no longer depends on the pressure on the condenser side of the duct [17]. The simulator leaves this choked flow out and lets a declared duct conductance set its equipment edge.

5.3.4 Secondary Drying and the Moving Limit

Bound water desorbs from the share of the cake that is already dry, at a rate linear in its moisture with an Arrhenius constant, and the cake's glass transition follows the simplified Gordon–Taylor law, as shown in Equation (5.4).

$$\begin{aligned} \frac{\mathrm{d}X_j}{\mathrm{d}t} &= -k_d(T_{b,j})\,X_j\,f_{dry,j}, \qquad r_{d,j} = m_s k_d X_j f_{dry,j} \\ k_d &= k_{d,0}\,\mathrm{e}^{-E_{a,d}/(RT_{b,j})}, \qquad f_{dry,j} = 1 - m_{i,j}/m_{i0} \\ T_{g,j} &= \frac{w_j T_{g,w} + K(1-w_j)T_{g,s}}{w_j + K(1-w_j)}, \qquad w_j = \frac{X_j}{1+X_j} \end{aligned} \tag{5.4}$$

The kinetic and glass-transition constants were measured on 5 % sucrose in 2R vials [16], a different formulation and vial from the primary-drying calibration, and the bound water as the ice leaves, $X_0 = 0.08\,\mathrm{kg/kg}$ (a glass transition of 25.7 °C), is a teaching value. The same law sets the sticky point of milk powder in Volume 5, Chapter 10. At 25 °C the desorption time constant is 4.1 h and at 40 °C it is 2.0 h. Secondary drying therefore carries weaker evidence than primary drying, and every result that depends on it is read that way below.

5.3.5 The Score

Each group is judged on the truth at every step boundary. While more than its last 1 % of ice remains, its limit is the collapse temperature T_c at the front, and afterward it is the glass transition of its own cake at the vial bottom. The excursion $E_j = \int \max(0, T_j - T_{lim,j})\,dt$ is accumulated, and a group is scored collapsed above $E_{max} = 10\,\mathrm{K\,min}$, a scoring convention rather than a measured damage law. A collapsed cake is a defect whose potency the model does not quantify. The batch passes only if every group passes, no group holds ice at the ramp to secondary drying, and every group ends at or below 2 % moisture. The collapse limit is a path constraint and the moisture limit an endpoint constraint in the sense of Section 1.4.4. The cycle time with the worst group is the lead metric. Energy is reported as thermal duty (shelf heating, shelf cooling and condenser refrigeration), and an electricity figure counts the heater at 1, divides the cooling duties by declared coefficients of performance, 1.5 for the shelf chiller and 0.4 for the condenser, and adds a 1.5 kW pump for the whole cycle. That figure is an estimate built on teaching values, not a measurement.

5.3.6 Instruments

The controls read only a measurement packet. The shelf RTD lags 20 s with 0.05 K of noise, the capacitance gauge lags 2 s with 0.3 % relative noise and reads P_c, and the Pirani gauge, which is calibrated for nitrogen, reads $p_{N_2} + 1.6\,p_w$ through a 30 s lag with 0.5 % noise. The factor 1.6 is the source's ratio in a vapor-filled chamber, and the linear mixture law is a teaching form. Two thermocouples read the bottoms of the probe vials through 30 s lags with 0.1 K of noise plus the bias disturbance, and the condenser thermometer lags 30 s with 0.1 K. During secondary drying a sample thief pulls an outer-edge vial from the front row every hour, starting 15 min after the ramp, and the Karl Fischer result, with 0.06 points of noise, is reported 45 min after the pull and held until the next one. Every noise comes from one seeded generator.

Table 5.2 Parameters and Their Evidence. S Read in the Calibration Study (Primary Drying) or the Secondary-Drying Study, D Derived From S Values, T Teaching Value

Symbol	Meaning	Value	Unit	Ev.	Symbol	Meaning	Value	Unit	Ev.
r_i, r_o	vial radii, inner, outer	11.0, 12.0	mm	S	X_0, w_{max}	bound water, moisture limit	0.08, 2	kg/kg, %	T
V_{fill}, w_s	fill, sucrose (m/V)	3.0, 3	mL, %	S	E_{max}	collapse exposure	10	K min	T
ρ_{ice}, ε	ice density, porosity	919.4, 0.97	kg/m^3, –	S	c_{ice}, c_s, C_{gl}	heat capacities	1900, 1250, 2.4	J/(kg K), J/K	T
λ_{ice}	ice conductivity	2.45	W/(m K)	S	N_c, N_i, N_o	vials per shelf	778, 118, 126	—	T
Δh_{sub}	sublimation, 51 136.82 J/mol	2.839×10^6	J/kg	S, D	C_{sh}, UA_w	shelf and fluid, wall leak	50, 4	kJ/K, W/K	T
α_c, β_c, γ_c	center K_v fit	3.46, 1.93, 0.0292	K_v, K_v/Pa, 1/Pa	S	Q_{hx} limits	thermal unit	−2.5, +3.0	kW	T
α_e, β_e, γ_e	outer-edge K_v fit	−1.14, 4.46, 0.0757	same	S	G	duct conductance for vapor	2.5	m^3/s	T
f_{in}	inner edge between them	0.35	—	T	S_p, $\dot{n}_b$	pump (N_2), full bleed	0.05, 6×10^{-4}	m^3/s, mol/s	T
R_{p0}	resistance at zero layer	1.51×10^4	Pa m^2 s/kg	S	τ_v	bleed valve lag	20	s	T
A_{Rp}, B_{Rp}	resistance slope, curvature	9.68×10^7, 968	Pa m s/kg, m^{-1}	S	C_{cd}, UA_{coil}	coil, coil to refrigerant	60, 200	kJ/K, W/K	T
f_j	probed-vial resistance	0.80	—	T	T_{ev}, $Q_{ref,max}$	refrigerant, duty	−78, 1.2	°C, kW	T
T_c	collapse temperature	−34.25	°C	S	UA_{lk}, M_{cd}	coil leak, ice capacity	5, 15	W/K, kg	T
$k_{d,0}$, $E_{a,d}$	desorption, other formulation	277, 37 714	s^{-1}, J/mol	S	T_{wall}, T_{gas}	wall, gas for conversions	20, 293.15	°C, K	T
K, $T_{g,w}$, $T_{g,s}$	Gordon–Taylor	0.2721, 135, 347	–, K, K	S	Δh_{des}	heat of desorption	2.84×10^6	J/kg	T

5.3.7 Parameters, Checkpoint and Numerics

Table 5.2 gives the parameters with their evidence. The primary-drying coefficients are those of the calibration study [1], read in its Table 2, and everything about the dryer is a declared teaching value for a pilot machine. At the checkpoint the shelf is at −40 °C, every vial holds its full 2.903 g of ice (90 mg of sucrose and the bound water X_0 make up the rest of the fill), and a pseudo-transient finds the hold at which each vial takes as much heat as its sublimation needs, a center bottom of −47.7 °C with the bleed at 20.6 %, 3.99 Pa of vapor and 6.01 Pa of nitrogen. The ice and the ledgers are then put back full and zero. The engine of Section 1.6.1 integrates the states with a fixed 10 s step of fourth-order Runge-Kutta, the controls execute every 10 s and the historian logs a row a minute. Events (a group's last ice, a phase change, an assay draw) happen once at a step boundary, never inside the derivatives. A 50 h cycle takes 160 ms of computing headless and 50 s at the top speed of 3600 times real time in a browser.

5.3.8 Assumptions

Each group is one vial, so the scatter inside a group, which the source puts at 7.61 % in K_v and 1.10×10^4 Pa m^2 s/kg in R_p, is absent from the plant and appears only in the estimator's copies. A group whose peak is a few tenths of a kelvin under its limit would therefore still have vials over it. The front is flat, the conduction through the ice is quasi-steady and the shrinking-area tail is a declared form. Choked flow in the vial neck and the duct, stopper resistance, freezing, nucleation and ice-crystal size are outside the model, and a product-resistance slider stands in for their effect on R_p. The chamber is lumped and the vapor balance quasi-steady, which a test justifies by integrating the dynamic balance of a 1 m^3 chamber at 1 ms and finding it within 1×10^{-6} Pa of the algebraic value. The Pirani law is linear in the mixture. Secondary drying borrows its kinetics from another formulation. The collapse temperature, the exposure score and the moisture limit are teaching values, and real ones come from the product's specification and its regulator's assessment [13]. A simulated pass is not batch release.

5.4 Controllers, Estimator and Supervisor

Three PI loops run every $\Delta t_c = 10$ s on the velocity-form PID of Section 1.6.2, Equations (1.14) to (1.17), with the error $e = \mathrm{SP} - \mathrm{PV}$, no derivative action and a tracking time equal to τ_I. TIC-101 sets the thermal-unit duty from the RTD with $K_c = 400$ W/K and $\tau_I = 10$ min between −2.5 and 3.0 kW. PIC-102 sets the bleed from the capacitance gauge with 2 % per pascal and 120 s between 0 and 100 %, and it reports the pressure lost when its output is shut and the gauge still reads more than 0.5 Pa over the setpoint. TIC-103, in the probe-limited policy only, lowers the shelf setpoint so that the hotter probe stays at the limit less the margin, with 2 K of shelf per kelvin and 30 min, its output capped at the recipe's shelf target. The supervisor's shelf target reaches TIC-101 through a ramp limit

of 0.5 K/min upward in primary drying (the secondary ramp r_2 afterward) and 1 K/min downward, and it is held at the freezing temperature through the 1 h vacuum hold. In Manual the operator sets the duty and the bleed, and every loop tracks the applied output, so transfers are bumpless.

5.4.1 Three Rule-Based Shelf Policies

The fixed recipe ramps the shelf to its target and holds it. Probe-limited feedback is TIC-103. The endpoint-aware policy sets the shelf from an estimator, by the quasi-steady closed form that holds a vial's front at $T^* = T_{lim} - \Delta T_m$, as shown in Equation (5.5).

$$\begin{aligned} \hat{T}_s &= T^* + \hat{Q}\left(\frac{L_{ice}}{\lambda_{ice} A_p} + \frac{1}{\hat{K}_v A_v}\right) \\ \hat{Q} &= \Delta h_{sub} \frac{A_p \varphi \left[p_{ice}(T^*) - \hat{p}_w\right]^+}{\hat{R}_p} \end{aligned} \tag{5.5}$$

evaluated on the estimator's hot copy of the hottest group that still holds more than its last 1 % of ice. The target moves to the next group as each is predicted dry, never exceeds the recipe's shelf target, and, while the bleed command is below 3 %, eases down at 0.05 K/min instead of rising, because the vapor is then about to take the pressure. In secondary drying it is the lower of the secondary target and the estimated glass transition of the wettest cake less the margin. The policy is not a receding-horizon optimization and is not called model predictive control.

5.4.2 The Estimator and Its Learned Resistance

For each of the three batch groups the estimator integrates Equations (5.1), (5.2) and (5.4) with the calibration's constants, driven only by the RTD, the capacitance gauge and the vapor inferred from the two gauges, $\hat{p}_w = (P_{Pir} - P_{cap})/0.6$. It runs three copies per group, nominal, hot (K_v and R_p one standard deviation high) and slow (K_v low, R_p high), with the source's spreads of 7.61 % and 1.10×10^4 Pa m^2 s/kg. The gauges and the condenser thermometer also measure the batch's sublimation through the declared duct, and with learning on a common resistance multiplier follows the ratio of the estimator's own batch sublimation $\hat{W}_p$ to it, as shown in Equation (5.6).

$$\begin{aligned} \hat{W} &= \frac{M_w G}{R T_{gas}} \left(\hat{p}_w - p_{ice}(T_{cd,m})\right)^+ \\ \hat{r}_k &= \mathrm{clip}\left(\hat{r}_{k-1} \left(\hat{W}_p / \hat{W}\right)^a,\ 0.5,\ 2\right) \\ a &= 1 - \mathrm{e}^{-\Delta t_c / \tau_r}, \qquad \tau_r = 1\,\mathrm{h} \end{aligned} \tag{5.6}$$

The update runs in primary drying only, while every nominal copy holds more than 15 % of its ice. One flow cannot separate a resistance error from a heat transfer error, so a batch that takes less heat is read as a more resistant one, the safe direction. On batches that resist 85, 100, 115, 130 and 145 % of the calibration, the multiplier settles at 0.855, 0.997, 1.143, 1.295 and 1.445.

5.4.3 Endpoints and the Ideal-Information Bound

Primary drying ends on one of four rules: a recipe time, both probes within 2 K of the shelf for 30 min, a Pirani-to-capacitance ratio at or below 1.07 for 15 min after the first hour (the source's criterion [1]), or that ratio together with every slow copy dry. Secondary drying ends on a recipe time from the ramp, an assay at or below the moisture limit, or the estimator's wettest cake at or below it after 30 min. These are the clock, measured-event and estimate guards of Section 1.4.2. The ramp leaves the ice untouched, so leftover ice keeps subliming into a warming shelf. Every policy reads the same packet. An ideal-information bound reads the teaching truth instead, sets the shelf for each group's true front by Equation (5.5) with the true K_v, R_p and p_w, ramps at the first scan with no ice left, holds secondary drying under the true lowest glass transition less its margin and stoppers when

the true moisture passes. It runs through the same loops and ramp limits. It is a bound, not a policy a plant could run.

Three models are kept apart. The plant is Section 5.3. The identified models are the calibration study's K_v and R_p fits. The controller's model is the estimator's copy of those fits with one learned multiplier, the declared duct conductance and the linear Pirani law, which knows neither a contact loss nor the batch's true resistance. The predictive controller of Section 5.4.4 plans on that same estimator model and reads nothing else of the plant.

5.4.4 A Predictive Controller for Primary Drying

The fourth shelf policy is a receding-horizon quadratic program in the standard form [18], a structure long established in industrial practice [19] and already applied to primary drying [7], set up as the option of Section 1.6.3 describes in general. It acts in primary drying while TIC-101 and PIC-102 are both automatic. Every 10 min, which is 60 control scans with the first on the priming call at $t = 0$, it plans two setpoints and applies the first move. The shelf target u_T still passes through the shipped ramp limit and vacuum hold to TIC-101, and the pressure setpoint u_P reaches PIC-102 through a 0.25 Pa/min slew. Secondary drying under this policy follows the endpoint-aware rule, and the endpoint stays the estimator's.

The controller's model is the estimator's, made quasi-steady in the product temperature, as shown in Equation (5.7).

$$
\begin{aligned}
T_s &= T_{i,q} + Q_q\Big(\frac{L_{ice,q}}{\lambda_{ice}A_p} + \frac{1}{(1+z_q\sigma_K)K_{v,j}A_v}\Big)\\
Q_q &= \Delta h_{sub}\,\frac{A_p\varphi_q\big[p_{ice}(T_{i,q}) - p_w\big]^+}{\max(\hat{r},1)R_p(L_{d,q}) + z'_q\sigma_R}\\
p_w &= p_{ice}(T_{cd,m}) + b + \frac{RT_{gas}}{M_wG}\sum_j N_j\dot{m}_j^{nom}\\
C_{sh}\frac{\mathrm{d}T_s}{\mathrm{d}t} &= K_c(u_T - T_s) + z_I - \sum_j N_jQ_j^{nom}\\
&\quad + UA_w(T_{wall} - T_s)\\
\frac{\mathrm{d}z_I}{\mathrm{d}t} &= \frac{K_c}{\tau_I}(u_T - T_s), \qquad \tau_P\frac{\mathrm{d}P_c}{\mathrm{d}t} = u_P + \hat{d}_P - P_c
\end{aligned}
\tag{5.7}
$$

where q runs over nine vial copies, the hot, nominal and slow copy of each group j with (z_q, z'_q) equal to $(1,1)$, $(0,0)$ and $(-1,1)$, $\sigma_K = 7.61\,\%$ and $\sigma_R = 1.10\times10^4$ Pa m^2 s/kg. Every constant is the platform's declared value, frozen when the program loads. The first line is Equation (5.5) solved for the front, by a safeguarded Newton iteration to 1×10^{-10} K, so a copy's only state is its ice. The vapor closes through the declared duct from the measured coil temperature $T_{cd,m}$, and the bias b aligns it with the gauges' $\hat{p}_w$, moving halfway to the new difference at each decision. The shelf is TIC-101's PI without its clamp, on $C_{sh} = 50$ kJ/K per loaded shelf, and the pressure follows its setpoint with $\tau_P = \tau_{I,P}/(K_{c,P}K_G) = 205$ s, where $K_G = \dot{n}_bRT_{gas}/S_p = 29.25$ Pa per unit of bleed opening. The input disturbance $\hat{d}_P$ is the capacitance gauge's persisting offset from that lag, filtered over one decision, so a stuck bleed appears in it within minutes. With b and $\hat{d}_P$ the plan is offset-free on the two gauges in the disturbance-model sense [20]. A disturbance model is also how the estimator of Volume 9, Chapter 4 carries a fuel bias that no instrument measures. The batch's product temperature has no measurement, and the probes are not read. The plan takes the learned multiplier only in the careful direction, $\max(\hat{r},1)$, because one flow cannot distinguish a cake that resists less from a duct that carries less than declared (Section 5.6.9).

Step tests on case 4's batch at 8 h, each run minus an unstepped run on the same seed, check the loop models. For shelf-target steps of ±1 and $+3$ K the plant reaches 63 % in 300 to 460 s and the model in 320 to 480 s, a residual of 0.86 to 0.91 % RMS of the step. For pressure steps the plant takes

150 s and the lag 220 s, 3.6 % RMS, because the lag omits the PI's proportional feedthrough (with it, 2.4 %), a difference inside one decision. The quasi-steady front drops the vial's heat capacity, so the plan adds the estimator's dynamic front minus its own as a correction ℓ_q that decays with the copy's time constant τ_q, below 1 K and 3 to 4 min. One decision ahead, the plan's warmest judged front differs from the estimator's front 10 min later by 0.220 K RMS on case 9 (118 decisions) and 0.295 K on the challenge, at worst 2.2 K at the instant the judged copy passes to the next vial.

The horizon runs in steps of $\Delta = 10$ min from now to the predicted end of primary drying, the step at which every slow copy is dry under the previous plan, at most $N = 150$ steps (25 h), and it is re-evaluated at every decision. Twelve move blocks of geometrically growing length are spread over the N steps, and below two steps the previous plan is held. The nominal trajectory is the nonlinear model under the previous plan shifted by one decision, integrated by Heun's rule with the loops' exact zero-order hold, and the program of Equation (5.8) corrects it.

$$
\begin{aligned}
\min\ & -w_W \sum_{k=1}^{N} \delta W_k + \sum_{b=1}^{12} \left(S_T \Delta u_{T,b}^2 + S_P \Delta u_{P,b}^2 \right) \\
& + \sum_{\ell} \left(\rho\, \varepsilon_\ell^2 + \sigma \varepsilon_\ell \right) \\
\text{s.t. } & \bar{T}_{i,q,k} + s_{q,k} \left(\delta T_{s,k} + \gamma_q\, \delta P_{c,k} \right) \\
& \quad + \ell_q \mathrm{e}^{-k\Delta/\tau_q} \le T^* + \varepsilon_{q,k} \\
& \bar{h}_k + \delta h_k \le -1\,\mathrm{Pa} + \varepsilon_{h,k}, \qquad h = p_w - P_c \\
& T_{s,min} \le u_T \le T_{s,1}, \qquad |\Delta u_{T,b}| \le r_1 \Delta \\
& |u_P - P_{c,SP}| \le 5\,\mathrm{Pa}, \qquad |\Delta u_{P,b}| \le 2.5\,\mathrm{Pa}
\end{aligned}
\tag{5.8}
$$

The bar marks the nominal trajectory. The deviations δT_s and δP_c follow from the moves through the three linear loop states of Equation (5.7) discretized exactly over Δ, $s_{q,k}$ is the front's sensitivity to the shelf at step k along the nominal, γ_q the ratio of its pressure to shelf sensitivity at the first step, and δW_k the batch sublimation's linearized change in kg/h. The front row applies to one copy per group, the warmest of those still holding more than their last 1 % of ice (hot, then nominal, then slow), and when no copy is icy the shelf is held at its setpoint, as the shipped policy holds it. The vapor row stands for the condenser's capacity at the planned pressure, because the bleed loses the pressure once the vapor alone reaches it. The weights are $w_W = 10$ per kg/h, $S_T = 0.05$ per K^2, $S_P = 0.05$ per Pa^2, $\rho = 10^4$ and $\sigma = 10^3$ per kelvin or pascal. Both output rows are soft, with one slack per step over the first hour and one per row beyond, a weighting that keeps the plan from paying once for a violation and riding the limit afterward [21]. The input bounds are hard: $T_{s,min} = -55\,^\circ$C, $T_{s,1}$ the primary shelf ceiling, the shelf held through the vacuum hold, u_P within 5 to 30 Pa, and per block boundary one decision of the 0.5 K/min ramp ($r_1\Delta = 5$ K) and 2.5 Pa. The pressure box of ± 5 Pa about the recipe's setpoint $P_{c,SP}$ is a design choice. A higher pressure costs nothing in this model but the Pirani's contrast, so an open box drives the plan to 30 Pa, and the box keeps the recipe's pressure a meaningful setting. The program has 24 move variables and at most 34 slacks and is solved by the series' shared active-set solver.

The plan stands only when the solver returns an optimum, the move, plan, predictions, objective and slacks are finite, every slack is non-negative, every planned input lies in its box and the first planned move equals the applied one. Otherwise the historian is marked, the shelf passes to the endpoint-aware target and the pressure to the recipe through their ramps, and the next valid plan takes control back. A test that injects ten faults rejects all ten, and a forced fallback at 4.00 h moves the shelf by at most one scan of its ramp and the pressure by one scan of its slew. The shipped protections remain above the plan, namely the ramp limit, the vacuum and fluid-excursion holds, the endpoint rules, the pressure-lost check and the endpoint-aware bleed rule. The plan moves setpoints only, so no PI is switched and no transfer needs tracking. Its input is a packet of the instruments, the estimator's nine copies and $\hat{r}$, its own setpoints and the load configuration. A source scan catches 26

Table 5.3 Code Verification and Agreement With the Calibration Study

Check	Result
Code verification	
Ice and energy per group over three whole cycles (case 1, case 7 and the challenge): $m_{i0} - m_i = \int \dot{m}\,dt$ and $\int Q\,dt = \Delta h_{sub} \int \dot{m} + \Delta h_{des} \int r_d + \int C\,dT_b$	$\leq 8.4 \times 10^{-15}$ relative
Chamber and batch water ledgers, same cycles	$\leq 1.2 \times 10^{-14}$ relative
Batch water by an independent sum, challenge as shipped	5.4×10^{-7}, the logged clip of 7.5×10^{-9} kg per vial
R_p monotone in L_d, never falls in a run, heating never creates ice	0 falls, 0 increases
Design point by hand, case 4 at 12 h	$\dot{m}$ within 2.1×10^{-5}, $\Delta h_{sub}\dot{m}/Q$ 0.999–1.000
Step 10 to 5 s, the nine original cases and the suggested challenge	cycle ≤ 0.0014 h, peaks ≤ 0.010 K, 0 verdict flips
Seeds 1 and 2 against the default, same runs	≤ 0.028 h, ≤ 0.036 K, 0 flips
Suggested challenge played in 1 min and 1 h chunks	identical
Each case loaded first, from the list and after Reset	identical, 10 of 10
Agreement with the calibration study [1]	
Edge vial dry at a fixed 1.5 K margin, 10 Pa of vapor	18.1 h against 17.9 h
Center vial dry, same policy	21.9 h against 21.5 h
The same at a 1.0 K margin	16.4 and 19.9 h
Pirani ratio while subliming, case 1, 3 to 20 h	mean 1.34 against about 1.6

of 26 doctored copies that read the plant, a replay reproduces 36 decisions bit for bit from packets recorded on a plant with 140 % resistance, 75 % contact and an 80 % duct, and a collapse temperature the controls are not told leaves 48 decisions bit-identical.

The weights come from a rule fixed before any run. On case 9 alone, the grid $w_W \in \{1, 10, 100\}$ by $S_T = S_P \in \{0.005, 0.05, 0.5\}$ is searched for the shortest primary drying that keeps every group 0.25 K under its limit at every instant (scored on the truth after the run), never loses the pressure for more than 10 min and never falls back, with ties inside 0.05 h broken by the least shelf travel. All nine settings qualify, and $w_W = 10$ with $S = 0.05$ gives 21.16 h and 31 K. A decision costs 4.2 to 6.7 ms on average headless and 5.12 ms on average in a browser (10.5 ms at the 95th percentile, 24.1 ms at worst), measured with other jobs loading the host. At 3600 times real time the engine's stepping takes 4.40 % of the wall clock under the plan against 0.82 % under the endpoint-aware policy, at 60 frames per second. Because a decision exceeds 2 ms, a frame guard ends a frame's stepping after 25 ms, which drops wall time and never a step and held no frame.

5.5 Verification and Calibration

Code verification and agreement with the literature are reported apart in Table 5.3. The first half shows that the program solves the stated equations. The second compares the model with the study whose coefficients it uses, so it is a consistency check of the calibration, not an independent validation, and the heat transfer and resistance fits come from the same equipment as the drying times compared.

The source's 17.9 and 21.5 h come from a risk-based optimizer that sampled 10 000 vials and held the predicted front at the collapse temperature with a 0.1 % risk of failure. The 17.9 h is when its uncertainty analysis had all edge vials dry, and the 21.5 h (21.48 to 21.55 h over three normal runs) is when 99.9 % of its vials were dry [1]. The simulator compares nominal vials under a fixed margin instead. The 1.5 K is a derived stand-in, because 3.09 standard deviations of the front temperature, propagated from the source's own K_v and R_p spreads, run from 1.2 to 2.3 K over the drying of the two groups. The model's times are 1.2 % and 2.0 % longer than the source's. The Pirani ratio is a disagreement. The source reads about 1.6 in a chamber filled with water vapor, while the model's declared duct leaves nitrogen in the chamber, a mean ratio of 1.34 in case 1, which rises to 1.57 under the heavier sublimation of case 2 at 12.5 h. Secondary drying has no comparison of its own, because its constants come from another formulation.

Two logging defects at $t = 0$ were corrected when the predictive controller was added. The Pirani-to-capacitance ratio was blank on the first logged row of every run although both gauges read (it now logs 1.239 on case 1, 1.281 on case 9 and 1.236 on the challenge), and each group's peak over

Table 5.4 The Ten Cases

#	Case and settings	Teaching point
1	Recipe −28 °C, 40 h, then 25 °C for 10 h	the idle hours of a clock
2	Shelf −16 °C, Pirani endpoint	the edges collapse first
3	Four shelves at −18 °C, 18 h target	the vapor takes the pressure
4	Shelf −22 °C, Pirani endpoint	hottest is not slowest
5	Case 4, 20 h clock, center contact 80 %	a recipe time outgrown
6	Probe-limited, probes in the center, shelf up to 0 °C	the probes protect themselves
7	Case 4, front-row probes end primary drying	a premature endpoint
8	Case 4, 4 h of secondary drying at 25 °C	moisture against glass transition
9	Four shelves, predictive controller, shelf up to 0 °C, Pirani endpoint	shelf and pressure planned together
10	Challenge: 115 % resistance, bleed stuck at 45 % for 1 h at 5 h	a shorter cycle every group survives

Table 5.5 Measured Outcomes. Dry Hours Are Outer, Inner and Center Groups, Excursions Are Peak and Exposure Over the Limit

#	Configuration	Outcome
1	as shipped	dry 21.2, 23.6, 25.1 h, closest 3.7 K under, stoppered 50.0 h
	Pirani endpoint	ramp 25.4 h, stoppered 35.4 h
2	−16 °C	outer +1.2 K, 367 K min, inner +0.3 K, 40 K min, center 0.2 K under, primary 16.2 h
	−20 °C, −22 °C	0.3 and 1.1 K under, primary 18.4 and 19.8 h
3	as shipped	pressure lost 737 min, peak 14.1 Pa, outer +1.6 K, 693 K min, inner 267 K min, center 77 K min
	−24 °C	pressure held, intact, primary 22.5 h
	sweep, 115 runs each	four shelves 0 feasible (best 19.1 h), three 8 (best 17.2 h)
4	as shipped	dry 16.4, 18.3, 19.6 h, outer 1.1 K under, center 2.3 K under
5	as shipped	center dry 21.6 h, 10.7 % of its ice at the 20 h ramp, +2.0 K, 30 K min
	Pirani endpoint	ramp 22.6 h, stoppered intact 32.6 h
6	as shipped	outer +2.4 K, 416 K min, inner +1.6 K, center +1.1 K, pressure lost 84 min
	one probe on the edge, 2 K	outer +0.5 K, 4 K min, primary 23.7 h
	endpoint-aware	primary 19.6 h, every group ≥ 0.9 K under
7	as shipped	probe vials dry 15.4 h, ramp 16.0 h with 18.6 % of the center's ice, center +7.2 K, 286 K min
	Pirani endpoint	ramp 19.8 h, stoppered intact 29.8 h
8	as shipped	center 4.55 % moisture at stoppering
	6 or 8 h at 25 °C	2.83 or 1.75 %
	6 h at 35 °C	1.95 %, stoppered 25.8 h
	estimate, assay end	27.4 and 28.8 h, 1.95 and 1.38 %
	1 K/min to 40 °C	every group over its glass transition, center +7.6 K
9	as shipped	primary 21.16 h, pressure held, every group ≥ 1.26 K under (Table 5.9)

its limit logged 0 K where the opening state is 12.8 to 14.7 K under. Every later row of all 60 runs checked (ten cases, six settings each) is bit-identical, and no number in this chapter moved. Four other columns stay blank by design, the probe margins after primary drying, when their collapse limit no longer applies, and the assay columns before the first report.

5.6 Ten Cases, Measured

The cases appear in Table 5.4, and the app lists them as its numbered scenarios (Section 1.6.4). All start from the pilot cycle (three shelves, −40 °C at the checkpoint, a 1 h hold, 10 Pa, a 0.5 K/min ramp to the primary shelf, the fixed-recipe policy, a 1 K margin and probes in the center) and depart only as their rows say. The numbers below come from the headless engine at the default seed, advanced step by step exactly like the browser, and a setting a row lists is entered before Run. Peak excursions are on the truth, positive above the limit. Table 5.5 collects the outcomes.

5.6.1 One Shelf, Three Groups

Under the fixed −22 °C shelf of case 4 the outer edge is dry at 16.4 h and the center at 19.6 h, 3.2 h apart, and the outer edge comes within 1.1 K of the limit while the center stays 2.3 K under, as

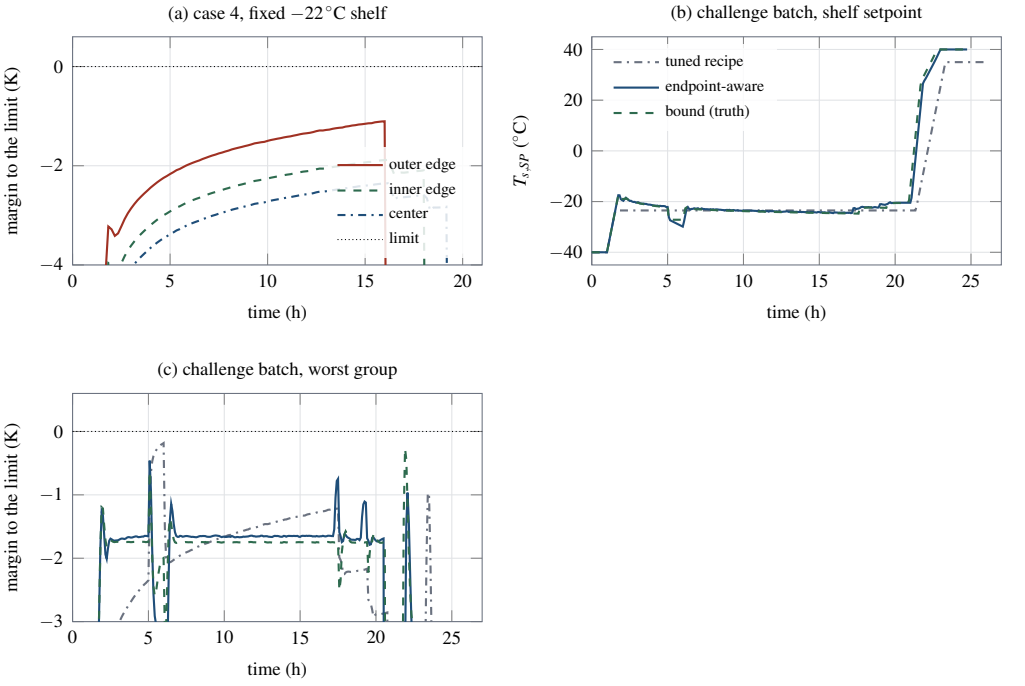


Figure 5.2 Mechanism and policy comparison from the simulator. (a) Case 4 as shipped, each group's margin to its limit (teaching truth: outer edge solid, inner edge dashed, center dash-dotted), each line leaving the plot when its group reaches its last 1 % of ice and its limit becomes the cake's glass transition. (b) The challenge batch under the strict tuned recipe (dash-dotted), the endpoint-aware settings (solid) and the ideal-information bound (dashed), shelf setpoint over the whole cycle. (c) The same three runs, the worst group's margin to its limit, the collapse temperature in primary drying and the glass transition after. Default seed, logged every 5 min in (b, c) and 10 min in (a).

shown in Figure 5.2a. The extra heat of the edge mostly goes into sublimation, but not all of it. Each margin shrinks as the dried layer thickens, the outer edge's from 2.44 K at 4 h to 1.15 K at 15 h, so a fixed shelf is set by the hottest group at the end of its drying. In case 2 the shelf at −16 °C drives the outer edge 1.2 K over the limit for 367 K min and the inner edge 0.3 K over for 40 K min, both collapsed, while the center peaks 0.2 K under. At 12.5 h the three groups stand on one screen, as shown in Figure 5.4. Two hours of primary drying (16.2 h against 18.4 h at −20 °C) buy the edges their structure.

> **TRY IT IN THE APP** **2 · Excess Shelf Heat**
>
> Load the scenario and press Run. Watch the three group cards beside the shelf map. The outer edge goes 1.2 K over its limit and the inner edge 0.3 K over, and both collapse, while the center peaks 0.2 K under. Primary drying ends at 16.2 h, two hours sooner than with the shelf at −20 °C.

Pause. Try Exercise 5.1 before reading on.

5.6.2 The Equipment Edge

Four shelves at −18 °C in case 3 make more vapor than the duct passes at 10 Pa. The bleed shuts, the pressure is lost for 737 min and peaks at 14.1 Pa, K_v rises with it, and all three groups collapse while the coil stays at −72 °C at its warmest, so the limit is the vapor path rather than the refrigeration. The

schedule's 18 h target for primary drying is infeasible on arithmetic alone. The declared duct carries 0.654 kg/h at 10 Pa of vapor, and four shelves hold 11.87 kg of ice, which is 18.1 h even if every vial sublimed at that rate from the first minute. A logged sweep of 115 four-shelf runs over every shelf policy, shelf targets of −26 °C to −16 °C, margins, ramps and pressures of 6 to 15 Pa finds none that ends primary drying by 18 h with the pressure held and every group intact (the best takes 19.1 h), while 8 of 115 three-shelf runs do. The case would not be run this way on a production dryer, but it serves to show a target that no controller can meet.

Pause. Try Exercise 5.2 before reading on.

OPEN CHALLENGE **Knowing What the Dryer Can Carry**

Every freeze-dryer supports a maximum sublimation rate at each pressure, set by its refrigeration capacity, its condenser area, the highest shelf temperature it reaches and choked flow, and the industry lacks a quantitative understanding of this equipment capability. Ice slab tests or laser absorption in the duct find the choke point, and turning the duct itself into a mass flow meter with computational fluid dynamics was an active project. Whether the critical pressure ratio can be calculated for a condenser placed below the chamber remained a research question [17]. Progress would be a capability curve known for every production dryer.

5.6.3 Measurements That Answer for Two Vials

In case 6 the probe loop holds both center probes near the limit less the margin, and every unprobed vial runs warmer than they do, because a probed vial here has 80 % of a normal vial's resistance (the direction reported for probed vials [15], the size a teaching value). All three groups collapse and the shelf drives hard enough early to lose the pressure for 84 min. One probe moved to the edge with a 2 K margin brings the outer edge to 0.5 K over for 4 K min, under the collapse score, at a primary drying of 23.7 h instead of 16.6 h, while the endpoint-aware policy, which reads no probe, ends it at 19.6 h with every group at least 0.9 K under. In case 7 the front-row probe vials are dry at 15.4 h and the probe rule ramps at 16.0 h with 18.6 % of the center's ice and 12.4 % of the inner edge's, which collapse at 7.2 K and 5.5 K over their limit. The Pirani endpoint ramps at 19.8 h, a quarter of an hour after the last center ice, and stoppers every group intact at 29.8 h. In case 5 a recipe clock written with half an hour to spare ramps at 20 h with 10.7 % of the center's ice still there, because the center takes 80 % of its heat. The Pirani does not need to know the contact, and it ramps at 22.6 h.

TRY IT IN THE APP **7 · A Premature Primary-Drying Endpoint**

Load the scenario and press Run. The front-row probe vials are dry at 15.4 h, the probe rule ramps at 16.0 h with 18.6 % of the center's ice left, and the center collapses 7.2 K over its limit. Press Reset, set End primary drying when to Pirani meets capacitance and run again: the ramp waits until 19.8 h, and every group is stoppered intact at 29.8 h.

5.6.4 Secondary Drying

Four hours at 25 °C in case 8 leave the center at 4.55 % moisture, because the ramp from −22 °C at 0.25 K/min takes 3.1 h of them. Six hours at 35 °C reach 1.95 % and stopper at 25.8 h, and ending on the assay waits for a result 45 min old and stoppers at 28.8 h with 1.38 %. A ramp of 1 K/min to 40 °C dries fastest and collapses every group with no ice left, the center 7.6 K over its glass transition, because the shelf outruns the moisture that sets the limit. These figures carry the weaker evidence of borrowed kinetics and a teaching X_0.

Table 5.6 Policies on the Challenge Batch at Six Paired Seeds. Cycle and Primary Drying in Hours, Worst Peak Over the Limit Per Group Across Seeds, Thermal Duty and Electricity Estimate as Seed Means

Policy (primary shelf, endpoint, secondary)	Pass	Cycle (h)	Primary (h)	Outer (K)	Inner (K)	Center (K)	Thermal (kWh)	Electric (kWh)
shipped: −22 °C, 19 h clock, 1 K/min to 40 °C, 3 h	0/6	22.01	19.00	+7.85	+8.37	+9.18	25.2	84.2
endpoint-aware, 1 K, learned $\hat{r}$, Pirani, est. moisture (suggested)	6/6	24.74–24.77	21.05–21.08	−0.39	−1.02	−0.96	26.5	91.6
the same without the learned $\hat{r}$ (ablation)	6/6	23.70–23.73	20.12–20.14	+0.08	−0.55	−0.18	26.3	89.0
the same with the estimator-and-gauges endpoint	6/6	25.59–25.64	22.05–22.09	−0.39	−0.68	−0.22	27.1	94.0
tuned recipe, strict: −23.5 °C, Pirani, 35 °C at 0.5 K/min	6/6	26.12–26.13	21.36	−0.06	−0.57	−0.86	26.3	94.5
tuned recipe, score only: −23 °C, same	6/6	25.68	20.93	+0.21	−0.31	−0.62	26.2	93.4
probe-limited, one probe each, 2 K, Pirani	0/6	31.60–31.62	25.08–25.11	+0.62	−0.24	−0.75	30.7	110.6
ideal-information bound, 1.75 K (reads the truth)	6/6	24.47	20.90	−0.62	−0.69	−0.22	26.2	90.7

DID YOU KNOW? Six Hours Are Not Enough at 25 Degrees

From the simulator: secondary drying at a cool shelf is slow. Held at 25 °C, the center ends at 4.55, 2.83 and 1.75 % moisture after 4, 6 and 8 h (Table 5.5), so only the longest hold meets the 2 % limit at that shelf temperature.

5.6.5 Three Policies on the Challenge Batch

The challenge batch resists the vapor 15 % more than the calibration, and the nitrogen bleed sticks at 45 % for an hour from 5 h. The target is stoppering within 30 h with every group passing and the pressure never lost for more than 10 min. Table 5.6 compares the policies at six paired seeds, each seeing the same instruments with the same delays. The fixed recipe was re-tuned on this very batch over a logged grid of 96 runs (primary shelf, secondary target and ramp), and the strict choice is the fastest that keeps every group under its limit at every instant, −23.5 °C then 35 °C at 0.5 K/min.

The shipped recipe fails at every seed in three ways at once. Its clock ramps with 4.1 % of the center's ice left, its 1 K/min ramp collapses every group on its glass transition, and 3 h at 40 °C leave the center at 2.48 %. The endpoint-aware settings stopper at 24.74 to 24.77 h with every group at least 0.39 K under its limit at every instant, and the strict tuned recipe at 26.12 h, 1.38 h later. The gain is mostly not in primary drying. The endpoint-aware policy ends primary drying at 21.05 h against 21.36 h, 0.31 h sooner, and its secondary drying takes 3.69 h against 4.76 h, because it can ramp at 1 K/min while holding the shelf under the estimated glass transition. So 1.07 h of the 1.38 h rests on the weaker half of the model, as shown in Figure 5.2b. The ideal-information bound needs a 1.75 K margin to keep every group under its limit through the bleed fault and the shelf loop's lag (at 1 K it reaches 22.96 h but goes 0.48 K over), and it stoppers at 24.47 h, so the estimator comes within 0.28 h of perfect information. Probe-limited feedback with one probe on the edge fails at every seed, 0.62 K over on the outer edge, the pressure lost for 18 min and 31.6 h of cycle. Figure 5.3 draws the cycle and the worst peak of each policy.

The learned resistance is what keeps the margin. Without it the policy is faster, 23.70 h, because the estimator then believes the cake resists less and predicts a colder front, and it touches the limit by 0.08 K (0.2 K min, inside the score). The ablation is the cheaper policy on this batch, and it fails on the next one. Thermal duty barely separates the passing policies, 26.2 to 27.1 kWh, of which the condenser's refrigeration is 18.7 kWh in the endpoint-aware run, and the electricity estimate follows the cycle time because the pump runs throughout.

5.6.6 Sensitivity to Resistance and Heat Transfer

The batches of Table 5.7 change the product resistance or the center contact, and the estimator knows neither. Over 85 to 145 % resistance the endpoint-aware settings pass every case in 22.19 to 27.41 h, the closest group 0.29 K under. The recipe tuned for 115 % keeps its time but goes 0.29 K

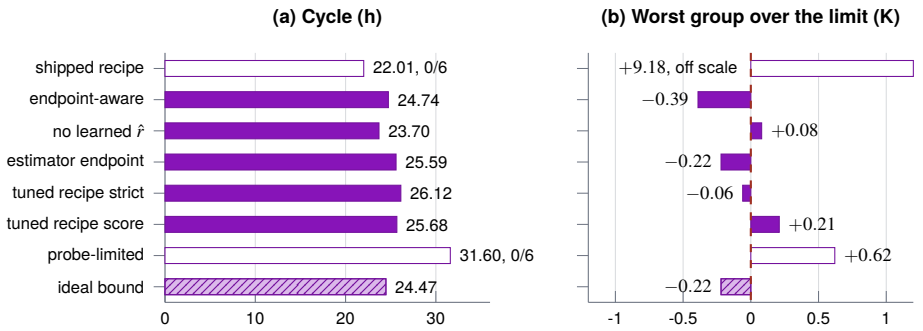


Figure 5.3 The policies of Table 5.6 on the challenge batch. (a) Cycle time, the lower end of the seed range where the table gives a range, with dark bars for policies that pass at all six seeds, white bars for those that pass at none and a light bar for the ideal-information bound. (b) The worst of the three groups' peaks over the limit, dashed line at the limit. Notice that the two fastest bars, the shipped recipe and the policy without the learned resistance, both cross the limit, and that the endpoint-aware policy ends within a fraction of an hour of the bound.

Table 5.7 Cycle (h) and Worst Peak Over the Limit (K) on Perturbed Batches, Default Seed. Bold: Fails

Batch	Tuned recipe	Endpoint-aware	Ablation	Bound
r 85 %	24.79, −0.65	22.19, −0.68	23.05, −0.99	**21.30**[a]
r 100 %	25.47, −0.44	23.44, −0.54	23.38, −0.45	22.88, −0.21
r 115 %	26.13, −0.06	24.74, −0.43	23.71, +0.03	24.47, −0.23
r 130 %	26.76, +0.29	26.07, −0.35	**24.11,** +0.47	26.04, −0.25
r 145 %	**27.39,** +0.61	27.41, −0.29	**24.46,** +1.22	27.61, −0.27
contact 85 %	**28.14**[b]	**28.08**[b]	**25.02,** +0.75[b]	25.69, −0.58
contact 70 %	**30.95**[b]	**32.58,** +0.39[b]	**26.95,** +1.76	27.02, −0.67

[a]pressure lost. [b]center over 2 % moisture. Contact rows at r 115 %.

over at 130 % (3.8 K min, a pass that depends on the score) and collapses its outer edge at 145 % (0.61 K, 18.5 K min), and the ablation collapses the outer edge at 130 % (13.6 K min). A contact loss defeats every policy that reads only the instruments. At 85 % contact the center ends at 2.00 to 2.04 % moisture under the endpoint-aware settings, the recipe and the ablation, just over the limit, because the moisture estimate does not know the contact either, and the assay samples the driest, front-row vials. Only the bound passes there. At 85 % resistance the bound itself fails, because driving the batch hard loses the pressure.

5.6.7 When the Gauge Law Is Wrong

The learned resistance is exact in this plant because the plant's Pirani obeys the linear law the estimator inverts and its duct has the declared conductance. Table 5.8 breaks each in turn. A Pirani that reads water vapor 1.7 times high instead of 1.6, a 6 % error, makes the estimator see more vapor than the batch makes and learn a cake that resists half as much ($\hat{r} = 0.52$). The endpoint-aware shelf then runs hot and every group collapses, 2.12 K over for 404 K min. A duct 20 % narrower than declared does the same, $\hat{r} = 0.50$, every group collapsed and the pressure lost for 90 min. Errors the other way are safe and slow, 4.1 to 4.6 h longer at factors of 1.4 and 1.5. A +0.3 Pa zero offset passes by 0.1 K min. The fixed recipe uses the Pirani only for its endpoint and passes every gauge case, and it fails only the narrow duct, through the plant's own pressure. The learning that buys robustness to resistance is therefore a single point of failure on the gauge law, and it is the main limitation of the method as implemented.

Pause. Try Exercise 5.3 before reading on.

Table 5.8 Plant and Estimator Mismatch on the Challenge Batch, Default Seed. Learned $\hat{r}$, Cycle (h) and Worst Peak (K)

Plant departs from the estimator	$\hat{r}$	Endpoint-aware	Verdict	Recipe
none (the app)	1.14	24.74, −0.43	pass	26.13, pass
Pirani factor 1.4	2.00	28.86, −2.02	pass	26.40, pass
Pirani factor 1.5	2.00	29.32, −1.90	pass	26.47, pass
Pirani factor 1.7	0.52	20.73, +2.12	all collapsed	26.06, pass
Pirani factor 1.8	0.50	21.05, +1.76	all collapsed	26.06, pass
zero offset −0.3 Pa	1.55	27.15, −1.43	pass	26.24, pass
zero offset +0.3 Pa	0.85	22.86, +0.56	pass, 9.9 K min	26.08, pass
duct 2.0 m^3/s	0.50	22.77, +1.55	all collapsed	26.79, fails
duct 3.0 m^3/s	1.86	28.24, −2.02	pass	26.03, pass

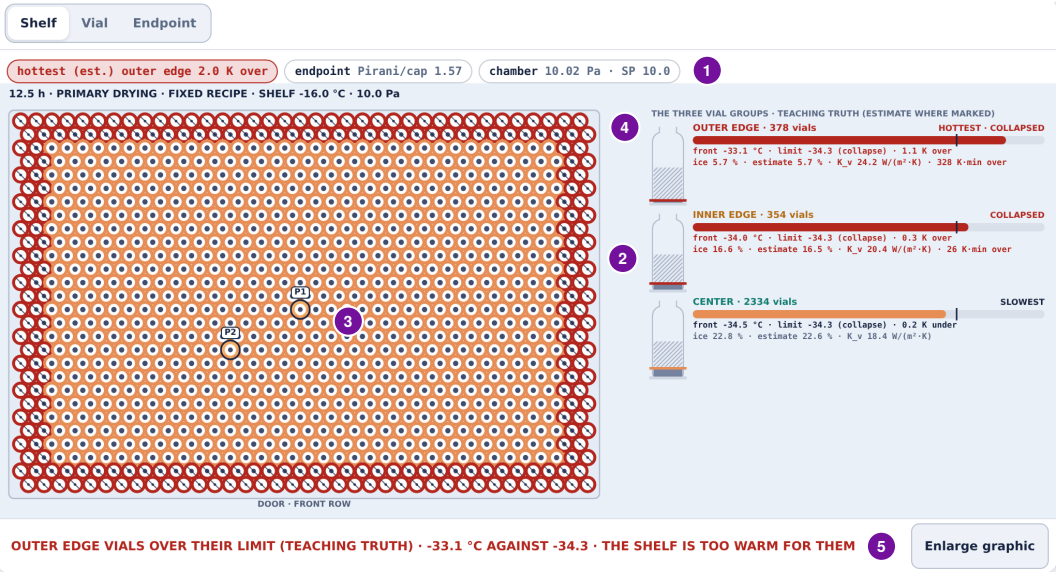


Figure 5.4 Excess shelf heat, case 2 at $t = 12.50$ h on the Shelf view, a fixed recipe at −16 °C and 10.0 Pa, cropped to the plant view. (1) The badges read the estimator's hot outer-edge vial 2.0 K over, a Pirani-to-capacitance ratio of 1.57 and the chamber at 10.02 Pa. (2) The shelf map rings every outer- and inner-edge vial red, (3) with the two probe vials in the center. (4) The group cards give the outer edge's front at −33.1 °C, 1.1 K over the limit with 5.7 % of its ice and 328 K min (collapsed and hottest), the inner edge at −34.0 °C, 0.3 K over with 16.6 % of its ice and 26 K min (collapsed), and the center at −34.5 °C, 0.2 K under with 22.8 % of its ice (slowest), with the estimator's ice within 0.2 points of each. (5) The status line names the outer edge as over its limit.

5.6.8 How Much the Score Decides

The 10 K min exposure is a convention, so its weight in the verdicts was measured. Of 48 distinct case and variant runs, 7 change a group's verdict if any excursion counts, 4 under a 1 K min score and 4 under 20 K min. In the 94 benchmark policy runs, 15 pass while over the limit at some instant, by at most 0.56 K and 9.9 K min. The endpoint-aware settings and the strict recipe never touch the limit, so the headline comparison does not depend on the threshold, while the recipe's pass at 130 %, the ablation's at 115 % and the offset case's pass do.

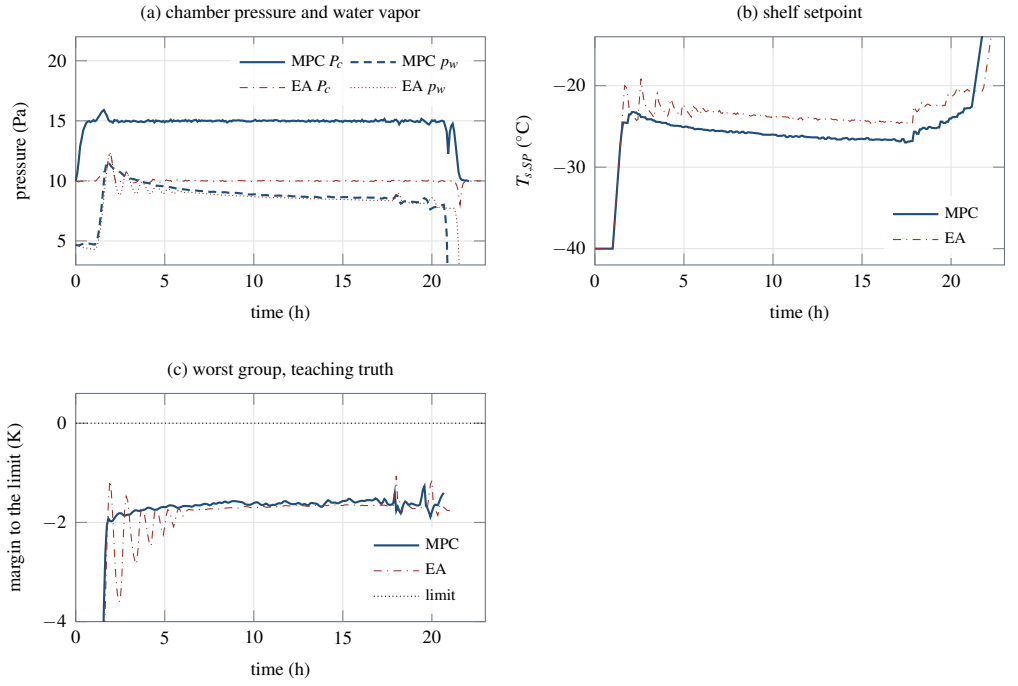


Figure 5.5 Case 9 under the predictive controller (MPC) and the shipped endpoint-aware policy (EA), default seed, logged every 5 min to an hour after primary drying. (a) Capacitance-gauge pressure (MPC solid, EA dash-dotted) and the chamber's water vapor (teaching truth, MPC dashed, EA dotted). (b) Shelf setpoint. (c) The worst group's margin to its limit (teaching truth). In (b) and (c) MPC is solid and EA dash-dotted.

5.6.9 Planning the Shelf and the Pressure Together

Case 9 loads the four shelves of case 3 (4088 vials, 11.87 kg of ice) with the shelf allowed up to 0 °C, a 1 K margin, the Pirani endpoint and the recipe's 10 Pa. It is the one shipped load on which the endpoint-aware policy loses the pressure at its recipe setpoint. Table 5.9 compares the policies on it, on case 1's batch with the same ceiling and endpoint, and on the challenge with its suggested settings, each run truth-scored to stoppering at the default seed with the same instruments. Its third group of rows answers the controller. The endpoint-aware policy retuned once by the controller's own rule over every knob it has (pressure 6 to 30 Pa, ramp 0.5 or 1 K/min, learning on or off, ceiling −10 or 0 °C) is the third row, and the strongest classical forms give the endpoint-aware policy the controller's pressure, either the top of its box or a high select on the measured vapor plus 2 Pa inside the box.

Under the controller, case 9's primary drying ends at 21.16 h against 21.81 h, the pressure is never lost against 36.7 min, the closest group stays 1.26 K under its limit against 0.91 K, and the shelf setpoint travels 31 K against 482 K, as shown in Figure 5.5. The plan raises the pressure setpoint to 15 Pa inside the vacuum hold (reached at 0.53 h) and holds a mean of 14.95 Pa through primary drying, where K_v is higher, while the vapor peaks at 11.62 Pa below it. Under the endpoint-aware policy the vapor passes 10 Pa at 1.68 h and first takes the pressure at 1.73 h, peaking at 12.36 Pa. The shelf trace of that policy saws because it recomputes the closed form every 10 s on noisy gauges, while the plan's moves are priced. On case 9 the product ceiling is active at 124 of 126 decisions with no positive slack, and the vapor row never binds (its closest predicted room is 1.16 Pa against the 1 Pa limit).

The gain is the pressure, not the horizon. Given the same 15 Pa, the endpoint-aware policy ends at 21.14 h, 19.08 h and 20.48 h on the three cases, a tie with the controller (21.16, 19.09 and 20.42 h).

Table 5.9 Predictive Controller Against the Rule-Based Policies, Default Seed, Truth-Scored to Stoppering. Closest Is the Least Margin of Any Group to Its Limit (Center in Parentheses), Negative Over It

	Case 9, four shelves				Case 1's batch		Challenge, suggested settings			
Policy	Primary (h)	Closest (K)	Lost (min)	Travel (K)	Primary (h)	Closest (K)	Primary (h)	Stoppered (h)	Closest (K)	Lost (min)
fixed recipe[a]	22.50	1.37 (2.48)	0	16	25.37	3.69	19.00	22.01	−9.17[b]	5.2
probe-limited, one each, 2 K	26.02	−1.43[c] (−0.41)	78.5	810	23.66	−0.50	—	—	—	—
endpoint-aware, 10 Pa (shipped)	21.81	0.91 (1.61)	36.7	482	19.63	0.90	21.05	24.74	0.43	0.5
predictive controller	21.16	1.26 (1.40)	0	31	19.09	1.11	20.42	24.13	0.85	0.2
endpoint-aware retuned, 30 Pa, 1 K/min	19.97	0.47 (0.93)	0	808	17.88	0.56	19.15	22.93	0.28	0
same rule, grid stopped at 25 Pa	20.24	0.52 (1.09)	0	791	18.15	0.62	19.44	23.20	0.33	0
endpoint-aware at the box top, 15 Pa	21.14	1.01 (1.58)	0	567	19.08	1.32	20.48	24.21	0.91	0.2
endpoint-aware, vapor high select	21.58	0.79 (1.55)	0	562	19.61	0.90	21.03	24.73	0.42	0.3
predictive controller, box 20 ± 5 Pa	20.37	1.45 (1.60)	0	28	—	—	—	—	—	—

[a] Case 9 −24 °C, case 1 −28 °C with the Pirani endpoint, challenge as shipped. [b] Three groups collapsed. [c] One group collapsed.
The controller solves 127, 115 and 123 times with no fallback.

Table 5.10 Worst Cases Under Every Structure. Groups Collapsed and K min Summed, Closest Approach to the Limit (K, Negative Over It), Pressure Lost (min). Load 3 Is Case 4's Batch, Load 4 Case 9's

Load, condition	Recipe	Probe, 2 K	Endpoint-aware	Same at 15 Pa	Retuned, 30 Pa	Controller, first build	Controller
3, none	0/0, 1.09, 0	0/4, −0.50, 24	0/0, 0.90, 0	0/0, 1.31, 0	0/0, 0.56, 0	0/0, 1.11, 0	0/0, 1.11, 0
3, resistance 160 %	1/238, −0.69, 0	1/11, −0.95, 0	0/0, 0.42, 0	0/0, 0.27, 0	0/2, −0.39, 0	0/0, 0.78, 0	0/0, 0.78, 0
3, condenser 30 %[a]	3/81590, −10.18, 3048	3/342, −2.57, 4700	0/0, 0.87, 2565	0/0, 0.81, 3166	0/0, 0.18, 0	0/0, 0.46, 66	0/0, 0.46, 66
3, duct 20 % narrower	0/0, 0.66, 0	1/14, −1.09, 64	3/91, −1.56, 82	3/580, −1.59, 0	3/333, −0.92, 0	3/641, −1.87, 0	0/0, 0.49, 0
3, T_c 1.75 K lower	1/194, −0.66, 0	3/237, −2.25, 24	1/22, −0.85, 0	1/63, −0.44, 0	2/172, −1.19, 0	1/60, −0.64, 0	1/57, −0.64, 0
4, none	0/0, 1.37, 0	1/28, −1.43, 78	0/0, 0.91, 37	0/0, 0.97, 0	0/0, 0.47, 0	0/0, 1.26, 0	0/0, 1.26, 0
4, condenser trip at 3 h	3/17528, −5.49, 1380	1/28, −1.43, 1337	0/0, 1.05, 849	0/0, 0.97, 0	0/0, 0.47, 0	0/0, 1.01, 0	0/0, 1.01, 0
4, resistance 60 %	0/0, 2.83, 23	1/15, −0.99, 728	0/0, 1.81, 168	0/0, 1.74, 0	0/0, 1.28, 0	0/0, 1.78, 0	0/0, 2.78, 0
4, resistance 160 %	1/21, −0.17, 0	2/40, −1.78, 30	0/0, 0.07, 16	0/0, 0.02, 0	0/3, −0.38, 0	0/0, 0.81, 0	0/0, 0.81, 0
4, contact 110 %	0/0, 1.26, 18	1/38, −1.55, 90	0/0, 0.41, 60	0/0, 0.43, 0	0/0, 0.36, 0[b]	0/0, 0.40, 0	0/0, 0.91, 0
4, condenser 30 %[a]	3/101438, −8.72, 4288	3/387, −3.58, 4706	0/0, 0.43, 3109	0/0, 0.19, 3277	0/1, −0.16, 0	0/0.3, −0.05, 97	0/0.3, −0.05, 97
4, duct 20 % narrower	0/0, 0.78, 311	3/52, −1.90, 152	1/16, −0.82, 371	3/429, −1.23, 0	3/192, −0.67, 0	3/588, −2.31, 0	0/0, 0.62, 0
4, T_c 1.75 K lower	1/91, −0.38, 0	3/298, −3.18, 78	2/79, −0.84, 37	2/139, −0.78, 0	3/267, −1.28, 0	2/129, −0.49, 0	2/127, −0.49, 0

[a]From the start, no structure finishes primary drying in 80 h. [b]Ice at the ramp, 0.02 %.

The third row ends case 9 at 19.97 h, 1.2 h before the controller, and the challenge at 19.15 h against 20.42 h of primary drying and 22.93 h against 24.13 h stoppered, with margins of 0.47 and 0.28 K against 1.26 and 0.85 K. The first grid of that rule stopped the pressure at 25 Pa and picked 20.24 h. Extended to the control's full range without changing the rule, it picks 30 Pa. With its box moved to 20 ± 5 Pa the controller ends at 20.37 h, within 0.13 h of the endpoint-aware policy at 25 Pa, so the pressure box sets the time. Over margins from 0 to 2 K on case 9 the controller ends 0.60 to 1.08 h before the endpoint-aware policy at 10 Pa, which loses the pressure for 15.5 to 73.3 min at every margin while the controller never does. When the condenser's refrigeration falls to 50 % at 3 h of case 9, the endpoint-aware policy loses the pressure for 849 min and ends primary drying at 34.28 h, while the controller loses none and ends at 31.94 h. One case is lost on margin. With the challenge's suggested settings and the estimator-and-gauges endpoint, the controller ends primary drying at 21.35 h against 22.05 h but brings the center to 0.13 K under its limit against 0.22 K, because after the hot copies enter their tails it keeps the colder copies at the limit.

The worst cases were run under every structure on case 4's batch (shelf target −22 °C for the recipe) and on case 9's (−24 °C), 18 conditions each, to stoppering or 80 h, as shown in Table 5.10. In the 26 runs of conditions the estimator knows or measures (none, the bleed stuck at 1.5 or 3 h, the condenser trip, the shelf fluid held, contact 70, 85 and 110 %, resistance 60, 145 and 160 %, probe bias ±3 K, on both loads), the controller collapses nothing, loses the pressure for at most 0.3 min and keeps every group at least 0.78 K under. In the same runs the endpoint-aware policy collapses nothing (closest 0.07 K) but loses the pressure for more than 10 min in 14 (12.8 to 849 min), the third row goes 0.19 to 0.39 K over at 145 and 160 % resistance, probe-limited feedback collapses a group in 16 and the recipe in 5. The condenser's ice never exceeds 11.9 kg of its 15 kg, and every run that finishes ends at or below 1.41 % moisture.

The controller's first build was less safe than the shipped policy in one condition. With the duct 20 % narrower than declared, the estimator turns the duct's shortfall into a learned resistance that falls to its 0.50 clamp, every copy reads cold, and the first build collapsed all three groups on both loads (641 and 588 K min). The endpoint-aware policy collapses three and one group (91 and 16 K min) and is held back on four shelves only by its bleed rule while it loses the pressure for 371 min, and given 15 Pa it collapses all three as well (580 and 429 K min). The revised plan judges the vials with $\max(\hat{r}, 1)$, the platform's declared value chosen before any run, so the learning can only make it more careful. After the change nothing collapses on either load (closest 0.49 and 0.62 K under, the pressure held, primary drying 22.95 h on four shelves), every comparison case moved by at most 0.01 h (case 9 from 21.153 to 21.156 h), and the tuning rule re-run keeps the same weights. The price falls where $\hat{r} < 1$ is genuine. At 60 % resistance primary drying lengthens from 17.81 to 19.67 h on four shelves and from 15.78 to 17.58 h on three, and with the center contact at 110 % from 19.59 to 20.39 h. Given to the endpoint-aware policy, the same floor also saves the narrow duct (no collapse,

24.62 h, the pressure lost for 125 min).

Two worst cases remain against the controller. A collapse temperature 1.75 K lower than the controls are told collapses groups under every structure, and the controller collapses as many as the endpoint-aware policy with more exposure (57 against 22 K min on three shelves, 127 against 79 K min on four). The cause is the pressure, because the endpoint-aware policy at 15 Pa gives 63 and 139 K min, and only the learner's margin can cover a limit no control is told. With the condenser at 30 % from the start no structure finishes, and on four shelves the controller goes 0.05 K over for 11 min (0.3 K min, not collapsed) where the endpoint-aware policy stays 0.43 K under. Its model holds the coil temperature constant over the horizon, and at 1.8 to 2.1 h the vapor stands 3.4 to 6.5 Pa above the pressure the plan set.

5.7 Technology in Practice

What Measures It

Thermocouples are preferred for the product because they measure at a point, best at the center of the vial bottom, while resistance temperature detectors read fixed points such as the shelf fluid at its inlet manifold [17]. Putting a probe into an open vial is a manual step, so production probes often sit only in the front row, a location the edge effect makes unrepresentative, and the probed vials may dry faster than the rest of the batch [17]. The simulator gives its probed vials 80 % of a normal vial's resistance, and cases 6 and 7 show what such probes do to a cycle.

> **TECHNOLOGY Sensor: A Pirani Gauge Beside a Capacitance Manometer**
>
> A capacitance manometer has a sealed, evacuated reference side and a metal diaphragm that flexes with the process pressure and changes a capacitance, so it reads the true pressure whatever the gas. A Pirani gauge holds a heated filament at constant temperature and reports the current that takes, which depends on the gas: the free molecular conductivity of water vapor is about 60 % higher than that of nitrogen. Calibrated with nitrogen, the Pirani reads high while ice sublimes and falls to the manometer's reading as the ice runs out. The manometer is accurate to about 0.25 % of reading, a Pirani gauge to 5 to 25 %. Pirani filaments can fail under repeated steam sterilization, and a vial fallen to the floor can fool the comparison [17].

What Moves It

Two final elements act on the batch. The fluid circulated through the shelves heats and cools the product, and a gas bleed, the dryer's vacuum control, sets the chamber pressure to improve the heat transfer to the product [13]. In a dryer controlled on the Pirani gauge, the nitrogen flow rises as the water vapor falls, which raises the true pressure, the heat transfer and the product temperature near the end of primary drying [17]. PIC-102 trims the bleed from the capacitance gauge, and because a bleed can only add gas, the valve shuts and the pressure is lost when the vapor alone exceeds the setpoint, as in case 3.

> **TECHNOLOGY Final Element: Shelves Fed by a Heat Transfer Fluid**
>
> The shelves of a freeze-dryer are heat exchangers with a serpentine liquid path, and a heat transfer fluid such as silicone oil circulates through them for heating and cooling the product [13]. The reading called shelf temperature is the fluid's inlet temperature, taken just upstream of the manifold that feeds every shelf, and the difference between inlet and outlet is proportional to the thermal load on the shelves [17]. One laboratory dryer's fluid system ran between −40 °C and 50 °C, cooling at 0.8 °C/min and heating at 1 °C/min at most [1]. Shelf fluid leaking into the chamber is cause to reject the batch [13]. TIC-101 stands for this loop with a thermal-unit duty between −2.5 and 3.0 kW.

Where the Controller Runs

One laboratory dryer retrofitted for model-based control ran its condenser, shelf and pressure loops as on-off controllers in programmable logic controllers. A remote supervisory computer drove the

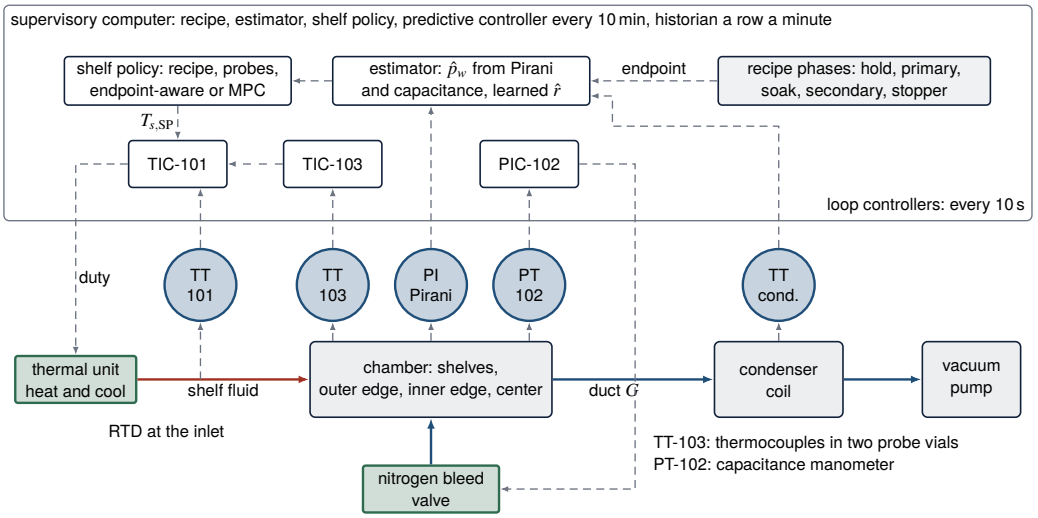


Figure 5.6 The chapter's loops on plant hardware. TIC-101 reads the RTD TT-101 at the shelf fluid inlet and sets the thermal unit's duty. PIC-102 reads the capacitance manometer PT-102 and moves the nitrogen bleed valve. TIC-103, in the probe-limited policy only (dashed), lowers the shelf target from the two probe thermocouples. The estimator reads the vapor from the Pirani and capacitance gauges and the condenser thermometer, and the shelf policy or the predictive controller sets the shelf target. Notice that no loop reads the product temperature of the unprobed vials.

dryer through its sequential steps, from freezing through primary and secondary drying to venting and stoppering, and the model-based strategy for primary drying sat on top of the low-level loops [1]. The batch control committee of ISA describes a functional model that relates recipe management, scheduling, sequential control, regulatory control and safety interlock systems [22]. Read in those terms, the vacuum hold, primary drying, the soak and secondary drying are the phases of one recipe, the endpoint rules are its transitions, and TIC-101, PIC-102 and TIC-103 are the regulatory control. Such loops run with cycle times from milliseconds to minutes [23], and the simulator's loops execute every 10 s, the predictive controller plans every 10 min, and the historian logs a row a minute, the rate best practice asks for to catch transient events [17]. Figure 5.6 draws the loops on plant hardware.

> **IN PRACTICE** **Commissioning the Pressure Pair**
>
> Best practice controls the chamber pressure on the capacitance manometer, with a temperature-controlled sensing head, and watches a Pirani gauge beside it. A capacitance manometer is calibrated against another one as a transfer standard, never against a thermal or mechanical gauge, and one on a steam-sterilized dryer is recalibrated every 3 months. Calibrating the Pirani matters less, because only its changes are read. Before the shelf is raised for secondary drying, the Pirani reading should nearly reach the manometer's: a difference of 5 to 10 millitorr works while the primary-drying pressure is above about 40 millitorr [17]. The simulator's ratio of 1.07 for 15 min plays that part.

What Is Modernizing It

Tunable diode laser absorption spectroscopy measures the concentration and the flow of water vapor in the duct between chamber and condenser with two lasers and two detectors, a non-invasive, in-line reading of the batch sublimation rate from which the heat and mass transfer coefficients are calculated [24]. It is the measurement the chapter's estimator reconstructs from two gauges and a declared duct. Wireless product sensors read probed vials without lead wires [17]. Model-based supervision that propagates the uncertainty of its heat transfer and resistance fits has run primary

drying at an accepted risk of failure [1], the strategy on which this simulator's primary drying is calibrated.

5.8 Limitations

Primary drying is calibrated on one laboratory study of one formulation in one vial, and its agreement with that study is consistency, not validation. Each group is one vial with no scatter, the front is flat, freezing is a checkpoint, and the chamber, duct, condenser, pump and thermal unit are teaching values for a pilot dryer, with no choked flow. Secondary drying borrows its kinetics and glass-transition law from another formulation, starts from a teaching bound water, and carries most of the measured advantage of the endpoint-aware policy. The collapse score is a convention that lets brief excursions pass, and a collapsed group carries no potency loss. The estimator inverts exactly the Pirani law and duct the plant obeys, so its learned resistance is optimistic here, and a 6 % error in the gauge factor or a 20 % narrower duct collapses every group. One measured flow cannot separate resistance from heat transfer, so a contact loss is read as resistance, and the moisture estimate misses it. The predictive controller's model inherits those blind spots and adds its own. It drops the vial's sensible heat (carried only as a decaying correction), desorption in primary drying, the thermal unit's saturation, the coil warming over the horizon and the condenser's narrowing near its ice capacity, its pressure lag omits the PI's feedthrough, and the floor on the learned resistance costs up to 1.86 h where a batch truly resists less. Energy is thermal duty, and the electricity figure rests on declared efficiencies. The comparison is on one batch design with six seeds whose noise barely moves the result (cycle ranges under 0.05 h), so seed spread is not an uncertainty estimate for a real dryer.

> **OPEN CHALLENGE** **The Vials Nobody Measures**
>
> The product temperature decides the cake, yet it cannot be controlled directly [24]. A thermocouple measures one point in one vial, inserting it by hand compromises sterility assurance, and production probes therefore often sit in the front row, where the edge effect makes them unrepresentative and the probed vials may dry faster than the batch. Wireless sensors avoid the lead wires, but in 2017 there was little published experience with them and best practices for them were called premature [17]. Progress would be a reading of the unprobed center vials that works with automatic loading.

5.9 Conclusion

A shelf of identical vials dries as a population, and this simulator keeps each group's ice, front and moisture beside what two thermocouples, two gauges, an assay and an estimator report. At 10 Pa the outer edge takes 32 % more heat than the center, dries 3.2 h sooner under a fixed shelf and comes 1.3 K closer to its limit, so the hottest group sets the shelf and the slowest sets the endpoint. A fourth shelf makes more vapor than the declared duct passes, and no setting of 115 meets an 18 h target. Probes in the center protect themselves while every group collapses, and front-row probes ramp with 18.6 % of the center's ice left. On the challenge batch an endpoint-aware policy stoppers 1.38 h before a recipe tuned for that batch and within 0.28 h of a bound that reads the truth, and it keeps its margin from 85 to 145 % resistance, but 1.07 h of its gain lies in secondary drying and its learned resistance fails on a 6 % gauge error. A predictive controller on the estimator's model holds the pressure on four shelves where that policy loses it for 36.7 min, with 31 K of shelf travel against 482 K, while the same policy retuned to 30 Pa ends primary drying 1.2 h sooner at a smaller margin. The model is lumped, calibrated on one study and optimistic about its own instruments. Further work would add vial scatter within groups, a nonlinear Pirani response with drift, a calibrated secondary-drying model for the same formulation, and an estimator that tests its gauge law before it learns from it. The simulator shows that a freeze-drying cycle is only as good as its knowledge of the vials it does not measure.

KEY POINTS

- A shelf of identical vials dries as a population: the edge vials take more heat and dry sooner, and the center vials dry last.
- At 10 Pa the outer edge takes 32 % more heat than the center, so the hottest group sets the shelf and the slowest the endpoint.
- Front-row probes ramp with 18.6 % of the center's ice left.
- An endpoint-aware policy stoppers 1.38 h before a tuned recipe, but 1.07 h of the gain lies in secondary drying.
- A predictive controller's measured gain is the chamber pressure it may set, not cycle time.

Where else this applies. The glass-transition law that limits the cake in secondary drying also sets the sticky point of milk powder in Volume 5, Chapter 10, where moisture and a glass transition are likewise inferred from slow measurements.

References

[1] B. Vanbillemont, N. Nicolaï, L. Leys, and T. De Beer, "Model-based optimisation and control strategy for the primary drying phase of a lyophilisation process," *Pharmaceutics*, vol. 12, no. 2, p. 181, 2020.

[2] M. J. Pikal, M. L. Roy, and S. Shah, "Mass and heat transfer in vial freeze-drying of pharmaceuticals: Role of the vial," *Journal of Pharmaceutical Sciences*, vol. 73, no. 9, pp. 1224–1237, 1984.

[3] S. Rambhatla and M. J. Pikal, "Heat and mass transfer scale-up issues during freeze-drying, I: Atypical radiation and the edge vial effect," *AAPS PharmSciTech*, vol. 4, no. 2, pp. 22–31, 2003.

[4] S. M. Patel, T. Doen, and M. J. Pikal, "Determination of end point of primary drying in freeze-drying process control," *AAPS PharmSciTech*, vol. 11, no. 1, pp. 73–84, 2010.

[5] D. Fissore, R. Pisano, and A. A. Barresi, "Process analytical technology for monitoring pharmaceuticals freeze-drying – a comprehensive review," *Drying Technology*, vol. 36, no. 15, pp. 1839–1865, 2018.

[6] R. Pisano, D. Fissore, S. A. Velardi, and A. A. Barresi, "In-line optimization and control of an industrial freeze-drying process for pharmaceuticals," *Journal of Pharmaceutical Sciences*, vol. 99, no. 11, pp. 4691–4709, 2010.

[7] N. Daraoui, P. Dufour, H. Hammouri, and A. Hottot, "Model predictive control during the primary drying stage of lyophilisation," *Control Engineering Practice*, vol. 18, no. 5, pp. 483–494, 2010.

[8] A. Chia, É. Poulin, J. Bouchard, P.-P. Lapointe-Garant, B. Van Meervenne, and F. Taveirne, "Experimental validation of multi-vial control for primary drying in a pilot-scale unit," *Chemical Engineering Research and Design*, vol. 193, pp. 281–293, 2023.

[9] L. F. Shackell, "An improved method of desiccation, with some applications to biological problems," *American Journal of Physiology-Legacy Content*, vol. 24, no. 3, pp. 325–340, 1909.

[10] E. W. Flosdorf and S. Mudd, "Procedure and apparatus for preservation in "lyophile" form of serum and other biological substances," *The Journal of Immunology*, vol. 29, no. 5, pp. 389–425, 1935.

[11] M. J. Pikal and S. Shah, "The collapse temperature in freeze drying: Dependence on measurement methodology and rate of water removal from the glassy phase," *International Journal of Pharmaceutics*, vol. 62, no. 2–3, pp. 165–186, 1990.

[12] B. C. Hancock and G. Zografi, "The relationship between the glass transition temperature and the water content of amorphous pharmaceutical solids," *Pharmaceutical Research*, vol. 11, no. 4, pp. 471–477, 1994.

[13] U.S. Food and Drug Administration, "Guide to inspections of lyophilization of parenterals," Inspection guide, July 1993, page content current as of 11/11/2014, 1993. [Online]. Available: https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/lyophilization-parenteral-793

[14] X. C. Tang, S. L. Nail, and M. J. Pikal, "Freeze-drying process design by manometric temperature measurement: Design of a smart freeze-dryer," *Pharmaceutical Research*, vol. 22, no. 4, pp. 685–700, 2005.

[15] V. R. Koganti, E. Y. Shalaev, M. R. Berry, T. Osterberg, M. Youssef, D. N. Hiebert, F. A. Kanka, M. Nolan, R. Barrett, G. Scalzo, G. Fitzpatrick, N. Fitzgibbon, S. Luthra, and L. Zhang, "Investigation of design space for freeze-drying: Use of modeling for primary drying segment of a freeze-drying cycle," *AAPS PharmSciTech*, vol. 12, no. 3, pp. 854–861, 2011.

[16] R. Pisano, D. Fissore, and A. A. Barresi, "Quality by design in the secondary drying step of a freeze-drying process," *Drying Technology*, vol. 30, no. 11–12, pp. 1307–1316, 2012.

[17] S. Nail, S. Tchessalov, E. Shalaev, A. Ganguly, E. Renzi, F. Dimarco, L. Wegiel, S. Ferris, W. Kessler, M. Pikal, G. Sacha, A. Alexeenko, T. N. Thompson, C. Reiter, J. Searles, and P. Coiteux, "Recommended best practices for process monitoring instrumentation in pharmaceutical freeze drying—2017," *AAPS PharmSciTech*, vol. 18, no. 7, pp. 2379–2393, 2017.

[18] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[19] 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.

[20] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[21] P. O. M. Scokaert and J. B. Rawlings, "Feasibility issues in linear model predictive control," *AIChE Journal*, vol. 45, no. 8, pp. 1649–1659, 1999.

[22] International Society of Automation, "ISA88: Batch control, committee scope and purpose," www.isa.org, standards committees, 2026, accessed 22 November 2026.

[23] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[24] F. Jameel, A. Alexeenko, A. Bhambhani, G. Sacha, T. Zhu, S. Tchessalov, L. Kumar, P. Sharma, E. Moussa, L. Iyer, R. Fang, J. Srinivasan, T. Tharp, J. Azzarella, P. Kazarin, and M. Jalal, "Recommended best practices for lyophilization validation—2021 part I: Process design and modeling," *AAPS PharmSciTech*, vol. 22, no. 7, p. 221, 2021.

5.10 Exercises

5.1 ⌕ Concept check · 2 min

Case 4 dries the pilot batch under a fixed shelf at −22 °C, 10 Pa and the Pirani endpoint. Which statement describes the three vial groups?

(a) The outer edge comes closest to the collapse limit and dries first, while the center dries last, so the edge sets the shelf and the center sets the endpoint

(b) The center comes closest to the limit because it holds ice longest, while the outer edge dries first and so sets the endpoint

(c) The three groups dry within minutes of each other, so one representative vial can set both the shelf and the endpoint

(d) The outer edge dries last, because the radiation it receives from the walls warms its cake instead of its ice

5.2 ⌕ Concept check · 2 min

In case 3 four shelves at −18 °C lose the chamber pressure for 737 min although PIC-102 is in Auto at 10 Pa. Why can the loop not hold its setpoint?

(a) Its integral time of 120 s is too long for the vapor surge, and a faster tuning would hold the setpoint

(b) The bleed can only add nitrogen, so once the vapor alone exceeds the setpoint the valve shuts and the vapor path sets the pressure

(c) The condenser coil warms past its duty and stops condensing, so water vapor accumulates in the chamber

(d) The capacitance gauge reads only the nitrogen, so the loop sees 10 Pa while the chamber is above it

5.3 ⌕ Concept check · 2 min

The estimator infers the water vapor as $\hat{p}_w = (P_{Pir} - P_{cap})/0.6$, the linear Pirani law with a factor of 1.6. Suppose the plant's Pirani in fact reads water vapor with a factor of 1.7. What happens to the endpoint-aware policy on the challenge batch?

(a) The Pirani ratio falls below 1.07 sooner, so primary drying ends with ice left in the center
(b) The estimator sees less vapor than the batch makes, learns a more resistant cake and runs the shelf cold and slow
(c) The estimator sees about a sixth more vapor than the batch makes, learns a cake that resists about half as much, and the shelf runs hot enough to collapse every group
(d) Nothing changes, because the learned multiplier is clipped between 0.5 and 2 and the Pirani is used only for the endpoint

5.4 Theory and calculation · 5 min

The vial heat transfer coefficient is $K_v = \alpha + \beta P_c/(1+\gamma P_c)$ in W/(m^2 K) with P_c in pascals. The calibration fits are $\alpha = 3.46$, $\beta = 1.93$, $\gamma = 0.0292$ for a center vial and $\alpha = -1.14$, $\beta = 4.46$, $\gamma = 0.0757$ for an outer-edge vial.

(a) Compute both coefficients at 10 Pa and the edge's extra heat for the same shelf-to-bottom difference. *Checkpoint:* 18.40 and 24.24 W/(m^2 K), 32 %.

(b) The inner edge sits 35 % of the way from center to edge. Compute its coefficient at 10 Pa. *Checkpoint:* 20.44 W/(m^2 K).

(c) The predictive controller of case 9 raises the pressure to 15 Pa. Compute both coefficients there, the relative rise of each and the new edge-to-center ratio. *Checkpoint:* 23.59 and 30.19, rises of 28.2 % and 24.5 %, ratio 1.280.

5.5 Theory and calculation · 6 min

The vapor leaves the chamber through a duct of declared conductance $G = 2.5\,\mathrm{m^3/s}$ to a coil at T_{cd}, so the condenser takes $W_{cd} = \frac{M_w G}{R T_{gas}}\left(p_w - p_{ice}(T_{cd})\right)$ with $M_w = 18.015\,\mathrm{g/mol}$, $R = 8.314\,\mathrm{J/(mol\,K)}$ and $T_{gas} = 293.15\,\mathrm{K}$. The ice vapor pressure is the last line of Equation (5.1).

(a) Compute p_{ice} at a coil temperature of −73.2 °C. *Checkpoint:* 0.161 Pa.

(b) Compute the vapor the duct carries with 10 Pa of water vapor in the chamber, in kg/h. *Checkpoint:* 0.654 kg/h.

(c) Four loaded shelves hold 11.87 kg of ice and three hold 8.90 kg. How long would each take to sublime at that rate, and what does that say about an 18 h target for primary drying? *Checkpoint:* 18.1 h and 13.6 h.

5.6 Theory and calculation · 6 min

Secondary drying follows Equation (5.4) with $k_{d,0} = 277\,\mathrm{s^{-1}}$, $E_{a,d} = 37\,714\,\mathrm{J/mol}$ and the Gordon–Taylor constants $K = 0.2721$, $T_{g,w} = 135\,\mathrm{K}$, $T_{g,s} = 347\,\mathrm{K}$. The moisture is $w = X/(1+X)$ and the limit at stoppering is $w = 2\,\%$.

(a) Compute the cake's glass transition when the ice leaves, at $X_0 = 0.08\,\mathrm{kg/kg}$, and at the moisture limit. *Checkpoint:* 25.7 °C and 59.1 °C.

(b) Below which moisture may the cake meet a shelf at 40 °C without passing its glass transition? *Checkpoint:* $w = 4.9\,\%$.

(c) Compute the desorption time constant at 25 °C and at 40 °C, and the time a fully dry cake held at 40 °C needs to fall from X_0 to the limit. *Checkpoint:* 4.1 h and 2.0 h, then 2.7 h.

5.7 App challenge · 8 min

Load scenario 1, A conservative recipe. Keep Auto, the default seed and the shipped settings, and set the speed to 3600, which runs an hour of the batch in a second.

(a) Run until the batch is stoppered and the run pauses. From the exported CSV, read the hour at which each group's ice first reaches zero (columns `truth_ice_o_pct`, `truth_ice_i_pct` and

`truth_ice_c_pct` against `hours`). Read the tiles Cycle time, worst group and Hottest group · peak over its limit. *Checkpoint:* 21.2 h, 23.6 h and 25.1 h, stoppered at 50.0 h, outer edge −3.7 K.

(b) Press Reset, set End primary drying when to Pirani meets capacitance (Endpoint view) before pressing Run, and run to stoppering. Read the hour primary drying ended, the first row whose `phase` is not `primary`, and the cycle time. *Checkpoint:* 25.4 h, stoppered at 35.4 h.

How many hours of the recipe's primary drying dried nothing? Give hours to within 0.1 h.

5.8 App challenge · 8 min

Load scenario 7, A premature primary-drying endpoint: both product probes in front-row vials, the shelf at −22 °C and primary drying ended when both probes reach the shelf. Keep the default seed and set the speed to 3600.

(a) Run to stoppering. From the CSV read when the probe vials ran out of ice (`truth_ice_p1_pct`) and when primary drying ended (the first row whose `phase` is not `primary`). Read the tiles Slowest group and Hottest group. *Checkpoint:* Probes dry at 15.4 h, ramp at 16.0 h, center 18.6 % of its ice at the ramp, +7.2 K, 286 K min.

(b) Press Reset, set End primary drying when to Pirani meets capacitance before pressing Run, and run to stoppering. Read the same hour and the cycle time. *Checkpoint:* Ramp at 19.8 h, stoppered at 29.8 h, all groups pass.

Why did the probe rule end primary drying early, and what did that cost? Give hours to within 0.1 h, the ice share to within 0.2 points, the peak to within 0.1 K and the exposure to within 2 K min.

5.9 App challenge · 9 min

Load scenario 9, MPC: the product limit and the condenser together (four shelves, Shelf policy on MPC). Keep the default seed and set the speed to 3600. For each run, read the hour primary drying ended from the CSV (the first row whose `phase` is not `primary`) and the minutes on the tile Condenser · pressure held.

(a) Run the scenario as loaded to stoppering. *Checkpoint:* 21.16 h, 0 min lost.

(b) Press Reset, set Shelf policy, primary drying to Endpoint-aware (estimator) on the Recipe card before pressing Run, and run to stoppering. *Checkpoint:* 21.81 h, 37 min lost.

(c) Press Reset, keep the endpoint-aware policy, set Chamber pressure, primary drying to 15 Pa, and run to stoppering. *Checkpoint:* 21.11 h, 0 min lost.

What does the predictive controller add on this load? Give hours to within 0.03 h and minutes to within 1 min.

5.10 Application area · 4 min

A filling plant puts both product thermocouples in front-row vials, because those are the only vials an operator can reach without disturbing the loaded shelves under sterile handling, and it ends primary drying when both probes reach the shelf. What does the chapter predict, and what would you recommend?

5.11 Application area · 4 min

A manager reads that the endpoint-aware settings stopper the challenge batch 1.38 h sooner than the best tuned recipe and proposes changing a validated production cycle on that evidence. What does the chapter's evidence support, and what would you require first?

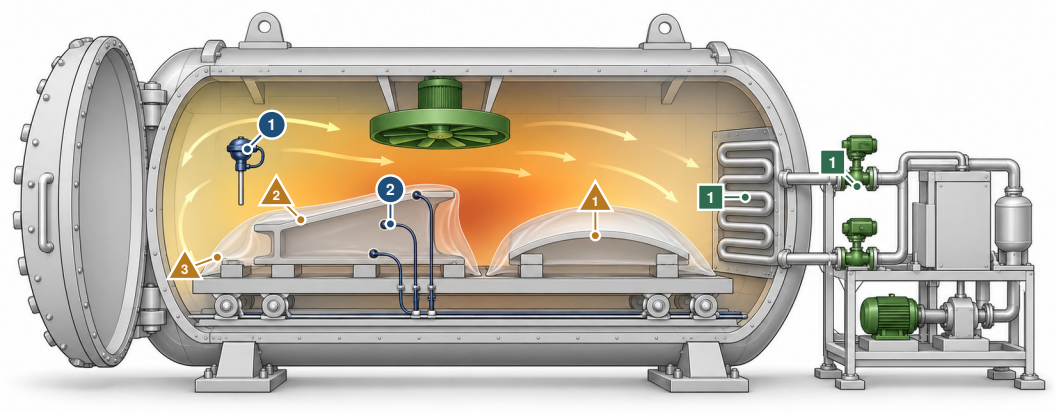
5.12 Review the agent · 5 min

An AI assistant runs the challenge with Reveal suggested tuning, then again with Estimator learns the batch's resistance from the chamber set to Off (Advanced). It reports: "With learning off the batch stoppers at 23.7 h instead of 24.7 h, every group passes and the challenge target is met. The learned

resistance only slows the cycle, so ship the settings with learning off.” Find the flaw and say how you would check it.

CHAPTER 6

Slowest Part, Hottest Core: Lead/Lag Control of a Composite Autoclave Cure



The controller moves 1 heater and cooling coil, driven by one split signal

The controller reads 1 air temperature 2 thermocouples at the bag side, middle and tool face of each part

What disturbs it 1 a part in the lee of another, in weaker air flow 2 a thicker part that heats itself as it cures 3 a torn vacuum bag

The thin parts follow the air while the thick part lags behind it and can overheat at its core, so the coldest thermocouple sets the pace and the hottest sets the limit. Illustration, not to scale.

IN PLAIN TERMS One Hot Oven, Many Different Parts

The main idea. Parts made of carbon fiber and resin are baked in a big pressurized oven until the resin sets hard. A thin part warms up in minutes, but a thick part on a heavy mold needs more than an hour.

The challenge. The schedule must wait for the slowest part, and the hottest part must not get too hot. The setting resin makes its own heat, so the center of a thick part can overheat before any sensor notices.

Why it matters. Too fast a cycle can spoil a thick part, and too slow a cycle wastes hours on every load.

THREE WAYS TO READ THIS CHAPTER

Explore	8 min	the first page, the history in Section 6.2, the first fact in Section 6.3, the second fact in Section 6.4, the try-it in Section 6.7, 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 6.7), Technology in Practice (Section 6.8), Limitations (Section 6.9) and the open challenge

The control question An autoclave cures parts of different thickness, on tools of different mass, with one actuator, the air temperature. How can the coldest of twelve thermocouples set the ramp while the hottest one, and a core that no thermocouple reaches, set the limit?

Learning objectives

- Compute the film, tool and conduction time constants of a part and name the one that sets its lag behind the air.
- Explain how a low selector with external reset runs the recipe on the lagging thermocouple while a max loop holds the hottest one at its limit.
- Explain why a limit loop on thermocouples cannot stop the exotherm of a thick part, and how a feedforward from a cure observer acts first.
- Relate the cure target to the glass transition and end a hold on the observer's estimate instead of a clock.
- Compare the cycle time and margins of lead/lag, the feedforward, an optimized recipe and a shrinking-horizon MPC.

Prerequisites Sections 1.4.2 (recipes and guards), 1.4.4 (path and endpoint constraints) and 1.6.2 (the PI of the air loop), and Chapter 5 for a batch whose hottest and slowest units differ. Section 1.6.3 is needed only for the MPC.

The app https://apmonitor.com/apps/autoclave-cure/

See also Volume 4, Chapter 3 for min and max selection over limiters whose integrators track the applied command, and Volume 4, Chapter 2 for a protective override that rests on its clamp until it is needed. Volume 9, Chapter 4 for offset-free MPC with an estimated disturbance.

6.1 Introduction

An autoclave is a heated pressure vessel in which composite parts, laid up from prepreg and sealed under vacuum bags on their tools, are taken through a recipe of ramps and holds. In the load simulated here a 3 mm skin on a thin aluminum plate, a 25 mm spar on a 450 kg invar tool, a 10 mm frame on steel and a 6 mm part in the lee of the spar share one vessel at 7 bar gauge. With the air stepped to 110 °C, the air reads 105 °C at 4.2 min and the coldest thermocouple of the skin gets there 10.1 min later, while the shadowed part needs 57.8 min and the spar 82.4 min. The standard recipe for the resin, run on the air's clock, cures every part but takes 438.3 min, and 37.5 min pass between the air reaching the hold band and the spar's tool face doing so.

A block diagram from air temperature to part temperature hides the two facts that decide this process. The first is that the controlled variable is not one measurement but the slowest and the fastest of many, so a recipe written for the part has to be run on the part that lags and limited by the part that leads. The second is that the product is a reaction. Thick thermoset sections release their heat of cure faster than conduction removes it, so the center overshoots the tool and the air, a behavior that through-thickness cure models describe [1, 2] and that cure cycles with cooling and reheating steps were later designed to suppress [3, 4]. In the simulator a 35 mm spar ramped at 3 °C/min passes 195 °C inside the laminate at 101.8 min, when its hottest thermocouple reads 187.3 °C, and peaks at 243.9 °C, although the limit loop on the hottest thermocouple has already pulled the air down. The misconception this chapter corrects is that a limit loop on the measured temperatures protects the part. It acts after the measurement moves, and a self-heating core moves first. The semi-batch reactor of Chapter 2 teaches the same lesson from the other side: there the protection is the dosing that limits the heat still to be released, not the temperature loop.

The class is broader than composites. Any batch whose many units share one heating medium, with a lagging unit that sets the schedule, a leading unit that sets the limit and a heat release the schedule can trigger, has the same shape, among them pharmaceutical shelf loads, kilns, heat-treatment furnaces and fed reactors. In the freeze-dryer of Chapter 5 the hottest vials likewise limit the shelf while the slowest decide the endpoint. Selector and override structures for such constraints, and the

windup that follows when a loop is passed over, are a subject in control theory [5], and model-based cure control of the autoclave has its own literature [6–10].

This chapter works with a four-part autoclave load rebuilt as a browser simulation, which runs the same mixed load on the air's clock and on the lagging part's clock, shows a core passing the limit between its thermocouples, and lets the recipe be shortened under the same constraints. The true temperature and degree of cure through every part are logged beside the twelve thermocouples a controller may read. On this load the chapter measures what a lead/lag supervisor, an exotherm feedforward from a cure observer and an offline recipe search do to cycle time and margins. The kinetics, the glass-transition relation and the viscosity law are published forms [11–13], the parameters come from a public report [14], lead/lag and override control are established practice, and model-based and predictive cure control have been reported [7–9]. The measured results belong to this model. Among them, the feedforward buys margin rather than minutes, and a slab bound that it acts on overestimates the core's rise while the reaction runs and underestimates it after. On the implementation side, the model's energy ledgers close to within 10^{-11} of the heater's energy over a cycle, its operator-split integrator is checked against an independent explicit solution, and the equations and constants given here suffice to rebuild it. A model predictive controller (MPC) that plans the air over the remaining recipe, a shrinking horizon like that of a 1994 miniature-autoclave study [8], is compared with the shipped, retuned and prediction-fed loops, with the safety defect of its first build and the fix's price.

Section 6.2 traces the domain from the cure kinetics to model-based cure control. Section 6.3 sets out the load as modeled, Section 6.4 explains what sets the lags of the parts and the exotherm limit, and Section 6.5 describes the control layers, with the MPC in Section 6.5.6. Section 6.6 separates verification of the code from agreement with the source of the resin data. Section 6.7 measures the nine cases and the challenge, each with a completion rule (Section 1.6.4), and the MPC in Section 6.7.6. Section 6.8 describes the hardware of a real autoclave, Section 6.9 states the limitations and Section 6.10 concludes.

6.2 Autoclaves, Cure Models and Cure Control

The thermoset chemistry came first. In 1973 an autocatalytic rate law was proposed from isothermal and scanning calorimetry of a polyester, in good agreement with measured rates and integral heats [11]. Its form, a rate proportional to $(k_1 + k_2\alpha^m)(1-\alpha)^n$, is the one extended to two terms in the parameter source used here [14]. A 1982 model of mold filling and curing for reaction injection molding predicted the pressure rise, extent of reaction and temperature in instrumented molds [13], and the viscosity law with a divergence at the gel point used here carries the names of its authors. In 1990 the DiBenedetto equation was re-evaluated as a convenient constitutive relation between the glass transition and the extent of reaction of a thermoset [12].

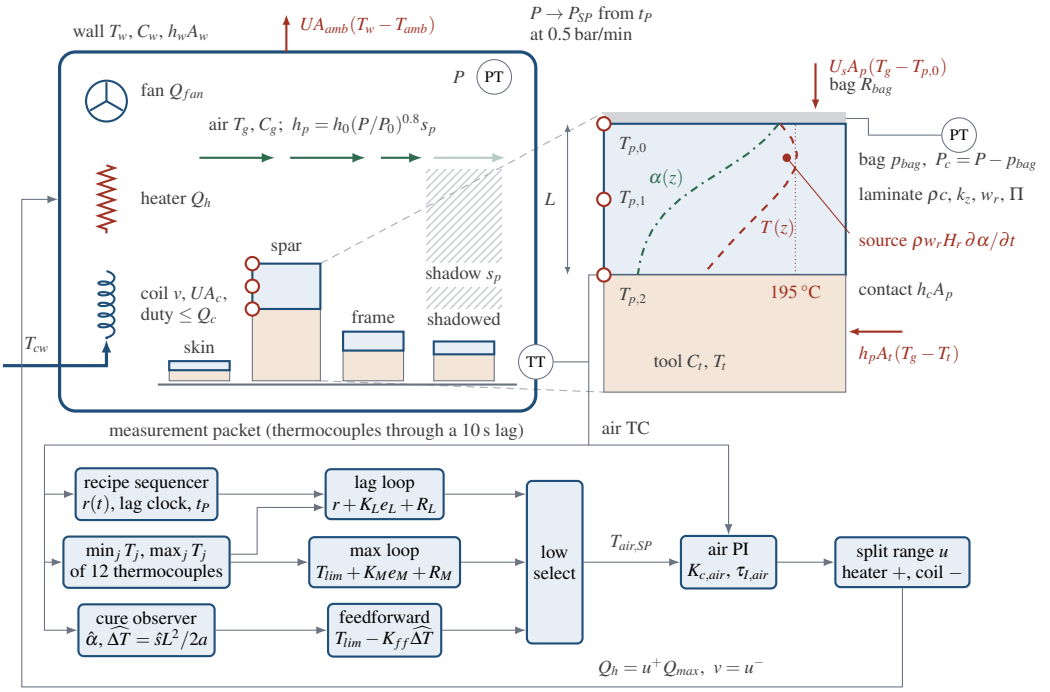
The process model followed for prepregs. In 1983 a computer code for flat-plate laminates gave, for a specified cure cycle, the temperature distribution, the degree of cure, the viscosity, the voids and the residual stress through the part [1]. Thick sections came next. A two-dimensional simulation of 1991 coupled transient anisotropic conduction to the cure kinetics with a heat generation term, predicted complex gradients in temperature and degree of cure, and showed the influence of the tool [2]. The consequence for recipes was drawn in 1997, when a cycle with cooling and reheating steps, timed by the cure rate and temperature at the midpoint, reduced the overshoot of 15 and 30 mm laminates [3], and in 2002, when measurements on a 20 mm laminate showed substantial lag and overshoot at its center and the conclusion that cycles recommended by prepreg manufacturers for thin laminates should be modified for thick ones [4].

Control of the vessel moved from the clock toward the part. A rule-based expert system for real-time control of autoclave temperature and pressure was published as a book in 1990 [6]. In 1994 model-based optimal temperature profiles were carried out through a knowledge-based system [7], and shrinking-horizon model predictive control was applied to the curing of laminates in a miniature

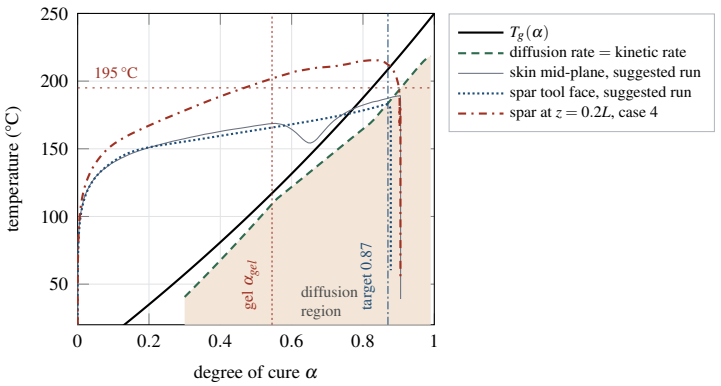
Table 6.1 Variables by Role, With Their Ranges and Initial Values

Variable	Role	Initial	Range
$T_{lag} = \min_j T_{p,j}$	CV of the lag loop (TC)	20 °C	—
$T_{max} = \max_j T_{p,j}$	CV of the max loop (TC)	20 °C	limit $\leq$ 190 °C
$T_{g,m}$	CV of the air loop (air TC)	20 °C	—
$T_{air,SP}$	MV of the supervisor	20 °C	20–225 °C
u	MV, split range	0	−100 to 100 %
P_{SP}	MV, pressure setpoint	7 bar g	0–7 bar g
s_p	DV, shadow factor	0.45	0.1–1
L_{spar}	DV, spar thickness	25 mm	10–50 mm
α_0	DV, resin lot	0	0 or 0.12 (aged)
leak, heater	DV, events	—	bag torn, half heater
T_g, T_w	states, air and wall	20 °C	—
$T_{p,i}$, $T_{t,p}$	states, 55 nodes, 5 tools	20 °C	—
$\alpha_{p,i}$, $\Pi_{p,i}$	states, cure and voids	0, 0	0–1, %
P, p_{bag}	states, vessel and bag	1.013 bar, 0.005 bar	absolute
five ledgers	states, energies	0	—

autoclave, treated as a batch reactor [8]. Time-optimal cure cycles computed by sequential quadratic programming improved on the manufacturers' recommended cycles in 1997 [15], and a 1999 study posed the shortest operating time as a constrained optimization and tracked the result with internal model control [9]. Cure monitoring in commercial carbon/epoxy composites during the autoclave cycle [16] and dynamic optimization for cure control [10] continued the line. The heat path to the parts was studied as well, and computational fluid dynamics of an autoclave run agreed best with measured part temperatures only when thermal radiation and a measured inlet velocity profile were included [17]. The lead/lag arrangement taught here is not the subject of these papers, and the simulator builds it from override control with external reset [5] on the thermocouples the parts carry.



(a)



(b)

Figure 6.1 The autoclave load as modeled and its cure map. (a) The vessel with heater, coil, fan and wall around the air node, four parts on their tools with the shadow behind the spar, and the spar enlarged through its thickness L with its three thermocouples $T_{p,j}$, the bag, the tool, the reaction source and its true temperature $T(z)$ (dashed) and degree of cure $\alpha(z)$ (dash-dotted) at 101.8 min of case 4 (teaching truth), with the 195 °C limit dotted. Below are the control layers as implemented, all fed by the measurement packet. Process model only, not to scale. (b) Temperature against degree of cure from the application's kinetics, with the glass transition (solid black), the line where the diffusion rate equals the kinetic rate (dashed, bounding the shaded diffusion region), gelation and the cure target (thin vertical lines), and three true trajectories: the skin's mid-plane (thin gray), the spar's tool face in the suggested run (dotted) and the spar at $z = 0.2L$ in case 4 (dash-dotted).

> ⧖ HISTORY **1985: The Cure Cycle Becomes a Design Variable**
>
> In 1985 a progress report to the Langley Research Center of NASA took up the optimal design of autoclave cure cycles for thick laminates. It coupled heat conduction through the laminate with the chemical kinetics of the prepreg resin, solved the two together by finite elements, and treated the surface temperature, the one the autoclave can control, as the design variable. Its goal was a temperature uniform across the thickness. A simulation of a 192-ply laminate agreed closely with the center temperature of an earlier report, but one complete cycle of 275 minutes in the autoclave took 17,056 seconds of computer time, because the equations were stiff [18].

6.3 The Load as Modeled

The vessel is 3 m long and 2 m across, as shown in Figure 6.1a. A 300 kW heater bank, a cooling coil on 15 °C water and a fan act on one well-mixed air node, a wall of 3 MJ/K exchanges heat with the air and loses it through the insulation to a 20 °C room, and four parts sit on a cart. Each part is a carbon/epoxy laminate resolved through its thickness L_p on 11 nodes, with its bag side ($z = 0$) facing the gas through the bag and breather and its far side on a lumped tool. Three thermocouples per part sit at the bag side, mid-thickness and the tool face, and an air thermocouple, a pressure transmitter and a vacuum transducer complete the instruments. An optional fifth part, a 15 mm fitting on 200 kg of steel, can be loaded before a run. The roles of the variables are listed in Table 6.1.

6.3.1 Air, Wall and Utilities

The air and the wall exchange heat with the utilities, the parts and the room, as shown in Equation (6.1).

$$\begin{aligned} C_g \frac{\mathrm{d}T_g}{\mathrm{d}t} &= Q_h - Q_c + Q_{fan} - h_w A_w (T_g - T_w) \\ &\quad - \textstyle\sum_p U_{s,p} A_p (T_g - T_{p,0}) \\ &\quad - \textstyle\sum_p h_p A_{t,p} (T_g - T_{t,p}) \\ C_w \frac{\mathrm{d}T_w}{\mathrm{d}t} &= h_w A_w (T_g - T_w) - U A_{amb} (T_w - T_{amb}) \\ Q_h &= f_h \, u^+ Q_{max}, \quad v = u^-, \quad Q_{fan} = Q_{fan,0} \, P / P_0 \\ Q_c &= v \, G_c \, (T_g - T_{cw}) \\ G_c &= \min \left(U A_c, \; Q_{c,max} / \max(1\,\mathrm{K}, \, T_g - T_{cw}) \right) \end{aligned} \tag{6.1}$$

where $u^+ = \max(0, u)$ and $u^- = \max(0, -u)$ split one signal between the heater and the coil's valve, f_h is 1 or 0.5 after a heater-bank failure, P is the absolute vessel pressure and $P_0 = 1.013\,25$ bar. The coil conductance G_c is fixed at its value at the start of each step, so the coil passes $U A_c (T_g - T_{cw})$ up to its rated duty of 300 kW. Without the cap the 6 kW/K coil removes 990 kW at 180 °C, three times the heater's authority, and the air loop reversed between heating and cooling 38 times in the cool-down of the suggested challenge run, against twice with it. Every part's coefficient follows the gas density at a fixed fan speed, as shown in Equation (6.2).

$$\begin{aligned} h_p &= h_0 \left(\frac{P}{P_0} \right)^{0.8} s_p, \qquad h_w = \frac{h_p}{s_p} \\ U_{s,p} &= \frac{1}{1/h_p + R_{bag}} \end{aligned} \tag{6.2}$$

At 7 bar gauge h is 73.2 W/(m^2 K), 5.23 times its value at one atmosphere, so a vessel that is not yet pressurized also starves the parts of heat. The shadowed part's factor of 0.45 leaves it 32.9 W/(m^2 K).

6.3.2 Parts Through Their Thickness

Each laminate conducts in z only and carries the heat of its own cure, as shown in Equation (6.3).

$$\begin{aligned}
\rho c\,\frac{\partial T}{\partial t} &= k_z\frac{\partial^2 T}{\partial z^2} + \rho\, w_r H_r \frac{\partial \alpha}{\partial t}, \qquad 0 < z < L_p \\
-k_z \frac{\partial T}{\partial z}\Big|_0 &= U_{s,p}\big(T_g - T(0)\big) \\
-k_z \frac{\partial T}{\partial z}\Big|_{L_p} &= h_c\big(T(L_p) - T_{t,p}\big) \\
C_{t,p}\frac{\mathrm{d}T_{t,p}}{\mathrm{d}t} &= h_c A_p\big(T(L_p) - T_{t,p}\big) \\
&\quad + h_p A_{t,p}(T_g - T_{t,p})
\end{aligned} \tag{6.3}$$

where $T(z)$ and $\alpha(z)$ belong to part p, discretized on 11 nodes with half cells at both faces, and $T_{p,j}$ in Equation (6.1) is the node of thermocouple j (nodes 0, 5 and 10). The source term heats a fully cured laminate adiabatically by $w_r H_r/c = 185.5\,\mathrm{K}$. Three time scales follow from the declared values. The bag-side film needs $\rho c L/U_s$, 1.5 min for the skin and 12.8 min for the spar. The tool needs $C_t/(h_p A_t)$, 29.3 min for the spar's invar and 27.3 min for the shadowed part's aluminum in its lee. Conduction across the laminate needs L^2/a with $a = k_z/\rho c = 3.45\times10^{-7}\,\mathrm{m^2/s}$, 0.4 min for the skin, 30.2 min for the 25 mm spar and 59.1 min at 35 mm. The thin parts therefore follow their surface and the spar follows its tool and its own conduction, both about half an hour.

6.3.3 Resin Kinetics, Glass Transition and Viscosity

The cure rate is the sum of two autocatalytic terms of the Kamal and Sourour type [11] placed in series with a diffusion rate, the form fitted in the public material model of the resin [14], as shown in Equation (6.4).

$$\begin{aligned}
\dot\alpha_k &= A_1 \mathrm{e}^{-E_1/RT}(1-\alpha)^{n_1}(\alpha+\alpha_{01})^{m_1} \\
&\quad + A_2 \mathrm{e}^{-E_2/RT}(1-\alpha)^{n_2}\alpha^{m_2} \\
\dot\alpha_d &= K_d\,\mathrm{e}^{-B/f}, \qquad f = a_d\big(T - T_g(\alpha)\big) + b(T_g) \\
\frac{\partial\alpha}{\partial t} &= \Big(\frac{1}{\dot\alpha_k} + \frac{1}{\dot\alpha_d}\Big)^{-1}
\end{aligned} \tag{6.4}$$

where T is in kelvin, b rises linearly from 0.021 to 0.031 as T_g goes from 120 to 195 °C and is constant outside, and f is floored at 10^{-30}. The glass transition follows the DiBenedetto relation in the form re-evaluated by Pascault and Williams [12], and the viscosity a Castro–Macosko form [13] with two Arrhenius terms, as shown in Equation (6.5).

$$\begin{aligned}
\frac{T_g - T_{g0}}{T_{g\infty} - T_{g0}} &= \frac{\lambda\alpha}{1-(1-\lambda)\alpha} \\
\eta &= \eta_{01}\mathrm{e}^{E_{\eta1}/RT} \\
&\quad + \eta_{02}\mathrm{e}^{E_{\eta2}/RT}\Big(\frac{\alpha_{gel}}{\alpha_{gel}-\alpha}\Big)^{A_\eta}, \quad \alpha < \alpha_{gel}
\end{aligned} \tag{6.5}$$

The diffusion term is what makes this resin unlike a plain autocatalytic model, as shown in Figure 6.1b. Held at 180 °C from $\alpha = 0$ the resin reaches 0.80 in 32.42 min and 0.87 in 56.50 min, the diffusion rate equals the kinetic rate at $\alpha = 0.858$, where T_g is 205.0 °C, and at 0.90 the series rate is $2.90\times10^{-7}\,\mathrm{s^{-1}}$ against a kinetic $3.38\times10^{-5}\,\mathrm{s^{-1}}$. After 6 h the cure stands at 0.8992. A target of $\alpha \geq 0.95$ could never be met at that hold, so the application's cure target is $\alpha \geq 0.87$, which DiBenedetto's relation maps to $T_g \geq 208.7\,^\circ\mathrm{C}$, and the demold rule $T_g \geq 208\,^\circ\mathrm{C}$ is the same constraint seen through it.

🕮 **Pause.** Try Exercise 6.3 before reading on.

6.3.4 Pressure, Vacuum and the Porosity Index

The compressor moves the vessel toward its setpoint at no more than 0.5 bar/min from the pressure step's time t_P and vents at demold, and the bag's gas relaxes toward a leak-weighted share of the vessel's, as shown in Equation (6.6).

$$\begin{aligned} P_{k+1} &= P_k + \mathrm{clip}\big(P_0 + P_{cmd} - P_k,\ \pm r_P \Delta t\big) \\ p_{bag,k+1} &= \bar{p} + (p_{bag,k} - \bar{p})\,\mathrm{e}^{-\Delta t/\tau_{bag}}, \quad \bar{p} = \frac{\lambda_b P_{k+1}}{1+\lambda_b} \\ \frac{\mathrm{d}\Pi}{\mathrm{d}t} &= \begin{cases} r_v(1 - P_c/P_{min}) & \eta < \eta_{flow},\ P_c < P_{min} \\ -\Pi/\tau_v & \eta < \eta_{flow},\ P_c \geq P_{min} \\ 0 & \text{otherwise, or } \alpha \geq \alpha_{gel} \end{cases} \end{aligned} \tag{6.6}$$

where P_{cmd} is zero before t_P and at demold and P_{SP} between, $P_c = P - p_{bag}$ is the consolidation pressure, λ_b is 0.005 for a sound bag and 3 after a tear, and $\tau_{bag} = 60$ s. The porosity index Π of each node grows while the resin can flow and the consolidation is under 4 bar, collapses in 10 min once it is over, and freezes at gel. It is a teaching index that rewards pressure before gelation and nothing else, not a model of void growth.

6.3.5 Instruments

The controller reads a measurement packet and nothing else. Each part thermocouple reads its node through an exact first-order lag of 10 s with colored noise of 0.2 °C standard deviation and a 30 s correlation time, the air thermocouple likewise, and the packet also carries the coldest and hottest of the part thermocouples with their positions, the vessel pressure in bar gauge and the bag pressure. Every noise draw comes from one stream seeded by the case, one normal per instrument per step, so Reset replays a run exactly (Section 1.6.1). Truth columns in the comma-separated (CSV) export carry a `truth_` prefix. A grep-level test fails whenever the controller, the sequencer or the observer reads a truth field.

6.3.6 Parameters and Their Evidence

Table 6.2 lists every parameter. The 26 resin constants are those of the public material model [14], compared one by one with the report by the application's technical review and found identical. The report characterizes the resin sold as Hexcel 8552, whose name appears here only as the public name of that source. The laminate, the vessel, the parts and the limits are declared teaching values for a 3 m vessel and a typical carbon/epoxy prepreg. The standard recipe, 1 °C/min to 110 °C for 60 min and 1 °C/min to 180 °C for 120 min, cooled at 2 °C/min to 60 °C, follows the shape of the cycle the report quotes from the manufacturer, 30 to 60 min at 225 °F and 120 min at 350 °F, rounded to 110 and 180 °C. Its ramps of 1 °C/min are slower than the quoted 3 to 5 °F/min, a conservative choice of the simulated shop.

6.3.7 Integration, Initialization and Ledgers

The thin nodes of the skin, 0.3 mm apart, are far too stiff for an explicit method at a useful step, so the application integrates with its own operator split, one of the apps' own steps that Section 1.6.1 mentions. Each physics step of $\Delta t = 5$ s takes half a step of cure at every node, advancing α and its adiabatic heating together by fourth-order Runge-Kutta, then the linear heat network over the whole step by the L-stable second-order TR-BDF2 method, whose two stages share one matrix solved exactly by a tridiagonal sweep per part and a Schur complement on the air node, then the second half step of cure (a Strang split). The pressure and the bag move first, the porosity index last. Five ledgers integrate the heater, fan, coil, room and reaction energies with the same quadrature as the temperatures, so the stored heat closes against them to round-off. The controller runs every 10 s, and a log row is written at the same interval. Every case starts with the whole load, the air and the wall at

Table 6.2 Model Parameters and Their Evidence. S: Read in the Material Model Report [14], T: Teaching Value, D: Derived. A Marks Values Adjustable in the Parameters Panel

Symbol	Meaning	Value	Unit	Ev.
Resin and porosity				
A_1, E_1	first kinetic term	153 900.5, 64 929.5	s^{-1}, J/mol	S, A(E_1)
n_1, m_1, α_{01}	exponents, offset	2.347, 1.413, 0.1594	—	S
A_2, E_2	second kinetic term	3.963×10^{11}, 133 168.3	s^{-1}, J/mol	S
n_2, m_2	exponents	1.029, 5.586	—	S
K_d, B, a_d	diffusion term	4.0, 0.21, 4.8×10^{-4}	s^{-1}, –, K^{-1}	S
b	diffusion offset, T_g 120–195 °C	0.021–0.031	—	S
H_r	heat of reaction	600	J/g resin	S, A
T_{g0}, $T_{g\infty}$, λ	DiBenedetto	−7, 250, 0.78	°C, °C, –	S, A($T_{g\infty}$)
η_{01}, $E_{\eta 1}$	first viscosity term	7.5×10^{-11}, 81 908	Pa s, J/mol	S
η_{02}, $E_{\eta 2}$	second viscosity term	0.0481, 13 228	Pa s, J/mol	S
α_{gel}, A_η	gel point, exponent	0.545, 2.466	—	S, A(α_{gel})
q	adiabatic rise, $w_r H_r / c$	185.5	K	D
η_{flow}, P_{min}	porosity flow, pressure	1000, 4	Pa s, bar	T
r_v, τ_v	void growth, collapse	0.10, 10	%/min, min	T
Vessel, laminate, parts, instruments and limits				
C_g, C_w	air with heaters and fan, wall	0.25, 3.0	MJ/K	T, A(C_w)
A_w, UA_{amb}, T_{amb}	wall area, loss, room	25, 150, 20	m^2, W/K, °C	T
Q_{max}, $Q_{fan,0}$	heater rating, fan heat at 1 atm	300, 3	kW	T, A(Q_{max})
UA_c, $Q_{c,max}$, T_{cw}	coil, rated duty, water	6, 300, 15	kW/K, kW, °C	T
h_0, R_{bag}	coefficient at 1 atm, bag	14, 0.004	W/(m^2 K), m^2 K/W	T, A
h_c	part to tool contact	400	W/(m^2 K)	T
ρ, c, k_z	laminate density, heat, conduction	1580, 1100, 0.6	SI	T, A(c, k_z)
w_r	resin mass fraction	0.34	—	T, A
L, A_p	skin, spar, frame, shadowed	3, 25, 10, 6; 2.0, 0.6, 0.8, 0.8	mm; m^2	T
C_t	tools, same order	29.2, 231.8, 58.8, 54.0	kJ/K	T
A_t, s_p	tool areas; shadow	2.0, 1.8, 1.2, 1.0; 0.45	m^2; –	T
r_P, τ_{bag}, λ_b	compressor, bag lag, leak ratio	0.5, 60, 0.005	bar/min, s, –	T
τ_{TC}, σ_{TC}	thermocouple lag, noise	10, 0.2	s, °C	T
limits	peak, cure, voids, band	195, 0.87, 1, ±5	°C, –, %, °C	T

20 °C and the resin at its lot's initial cure. The state has 179 entries, and a whole challenge cycle of 438 min costs 228 ms of computing headless and about 22 s at the top speed of 1200 times real time in a browser.

6.3.8 Assumptions

Each part is one-dimensional, so edges, ply drops and corners, which run hotter in a thick part, are absent, and the tool is one lumped mass in perfect contact through h_c. The air is one well-mixed node and each part's heat transfer is a constant times a pressure power and a shadow factor, while a real vessel has a flow field whose survey finds its slow positions and whose radiation matters [17]. The laminate's properties are constant, although its specific heat rises with temperature. Resin flow, bleed, compaction, volatiles and residual stress are outside the model, so a dwell earns nothing here that it earns in a real cure. The porosity index is a teaching law. The 195 °C exotherm limit, the cure target, the 1 % porosity threshold and the ±5 °C hold band are teaching values, and the real ones come from the resin's specification and the part's process specification. A simulated pass is not a qualification of any part or process.

Table 6.3 Spar Peak Temperature (°C, Teaching Truth) on Air Control and the Standard Recipe, Against Thickness and Ramp

Ramp	10 mm	20 mm	25 mm	30 mm	35 mm	40 mm
1 °C/min	180.1	181.9	188.5	199.2	213.6	231.1
2 °C/min	180.0	193.8	210.9	233.3	260.4	279.3
3 °C/min	180.0	195.7	215.4	241.6	269.8	288.5

✦ DID YOU KNOW? Windward and Leeward in a Full-Size Autoclave

One full-size industrial autoclave has a cylindrical cavity 3.7 m across and 9.6 m long, with air or nitrogen driven by a fan. Fifteen thermocouples on a large steel mold read hottest on the windward side near the door and coldest on the leeward side [19]. The shadow factor of this chapter stands for that leeward lag.

6.4 What Sets the Lags and the Limit

The recipe asks every part for the same temperature, and the load cannot deliver it at once. On the standard recipe run on the air's clock the skin tracks the air within 1 °C (a peak of 181.4 °C against the air's 180.8 °C), while the spar's tool face reaches the hold band 37.5 min after the air, because both of its half-hour time constants lie in series. A recipe on the air's clock either cuts the spar's hold short or pads every hold for it. The standard recipe pads, the least-cured point still reaches $\alpha = 0.887$, and the padding is cycle time.

The heat release scales with thickness and ramp, as shown in Table 6.3. On air control and the standard recipe a 10 mm spar never exceeds 180.1 °C at any ramp, a 25 mm spar peaks at 188.5 °C at 1 °C/min and 215.4 °C at 3 °C/min, and a 40 mm spar reaches 231.1 °C even at 1 °C/min. The mechanism is a ratio of rates. When the air ramps faster than the core can follow, the core arrives late at the temperature where the cure accelerates, with most of its resin unreacted, and its heat must leave through half the laminate, L^2/a of half an hour or more, while the second kinetic term of Equation (6.4) doubles its rate every 8.8 K near 177 °C. Lowering the air then removes heat only from the faces.

> **DID YOU KNOW? Thickness Beats Ramp**
>
> From the simulator: on air control and the standard recipe, a 30 mm spar ramped at the slowest rate, 1 °C/min, peaks at 199.2 °C, hotter than a 20 mm spar ramped three times as fast, which peaks at 195.7 °C (Table 6.3).

Thermocouples lie on the wrong side of this mechanism. At 101.8 min of case 4 the spar's node at $z = 0.2L$ is at 195.8 °C while its thermocouples read 187.4, 182.6 and 148.8 °C, as drawn in Figure 6.1a, and the hottest of them passes 195 °C 1.5 min after the core does.

Pause. Try Exercise 6.1 before reading on.

6.5 The Control Layers as Implemented

Every 10 s the controller runs the cure observer, the recipe sequencer, the supervisor that writes the air setpoint and the air loop, in that order, on the measurement packet. The application offers four structures, Air only, Lead/lag, Lead/lag with the feedforward (FF) and the MPC of Section 6.5.6.

6.5.1 The Air Loop

The air loop is the velocity-form PI of Section 1.6.2, written for this loop in Equation (6.7).

$$\begin{aligned} e_k &= T_{air,SP,k} - T_{g,m,k}, \quad s_k = s_{k-1} + K_c(e_k - e_{k-1}) \\ \hat{u}_k &= s_k + I_{k-1}, \qquad u_k = \mathrm{slew}\big(\mathrm{clip}(\hat{u}_k, \pm 1),\ 1/(180\,\mathrm{s})\big) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - \hat{u}_k}{\tau_I}\right]\Delta t_c \end{aligned} \tag{6.7}$$

with $K_c = 10\,\%$ of output per °C, $\tau_I = 10$ min, no derivative and $\Delta t_c = 10$ s. In the symbols of Equations (1.15) to (1.17), s_k is S_k, $\hat{u}_k$ is v_k with $u_b = 0$, and $T_i = \tau_I$. The output moves from zero to full in no less than 3 min, so the air overshoots a large step. In case 1 the air reaches 145.7 °C at 5.8 min on a step to 110 °C and stays within 2 °C of it from 13.6 min.

6.5.2 The Recipe Sequencer

The sequencer advances a reference r through ramp, dwell, ramp, hold, cool and demold for one clock thermocouple, the air on Air only and the lagging thermocouple T_{lag} on the lead/lag structures. On the lag clock the ramp is guaranteed. The reference advances at the ramp rate only while it is less than 10 °C ahead of T_{lag}, and a dwell or hold counts its time only while the clock thermocouple is within 5 °C of it. The hold ends on its time or, with the observer rule, once the observer's lowest $\hat{\alpha}$ is at least 0.875 and its $\hat{T}_g$ at least 210 °C, a guard on an estimate in the sense of Section 1.4.2. In the cool-down the reference falls at 2 °C/min, and demold comes when the hottest thermocouple reads 60 °C. The pressure is commanded on at t_P. The framework's shared stage sequencer was not used, because it switches on/off stages against a demand and has no ramp target, hold clock or choice of clock thermocouple.

6.5.3 Lead/Lag With a Maximum Selector

On the lead/lag structures a minimum selector across the twelve part thermocouples gives the lagging temperature T_{lag} and a maximum selector gives the hottest, T_{max} (Table 6.1). Two PI loops write candidate air setpoints from them, the lag loop on T_{lag} and the max loop on T_{max}, and a low selector

takes the smaller, as shown in Equation (6.8).

$$\begin{aligned} e_L &= r - T_{lag}, \qquad e_M = T_{lim} - T_{max} \\ \ell &= \mathrm{clip}(K_L e_L + R_L,\ -20,\ \Delta_{lead}) \\ o_M &= \mathrm{clip}(K_M e_M + R_M,\ -60,\ 40) \\ T_{lim} &= \min\left(190\,^\circ\mathrm{C},\ T_{seg} + 5\,^\circ\mathrm{C}\right) \\ T_{air,SP} &= \mathrm{clip}\big(\min(r+\ell,\ T_{lim}+o_M,\ T_F), \\ &\qquad 20\,^\circ\mathrm{C},\ 225\,^\circ\mathrm{C}\big) \\ R_X &\leftarrow R_X + \left(\bar{R}_X - R_X\right)\min(1,\ \Delta t_c/\tau_{I,X}) \end{aligned} \tag{6.8}$$

where T_{seg} is the dwell temperature during the first ramp and the dwell and the hold temperature after, T_F is the feedforward candidate (absent on Lead/lag), and $\bar{R}_L = T_{air,SP} - r$ and $\bar{R}_M = T_{air,SP} - T_{lim}$ express the applied setpoint in each loop's own terms. This is the external-reset form of override control [5], the integrator tracking that the limiters of the turbofan in Volume 4, Chapter 3 also use under their selector. For the selected loop $\bar{R} - R$ equals its own proportional term, so it acts as a PI, and the loop passed over sits $K_X e_X$ above the applied setpoint and takes over without a bump. The app's switch named Maximum selector turns the max loop's candidate off, and the Lead/lag card reads "max selector" while that candidate holds the air. The text calls it the max loop. The defaults are $K_L = 1$ with $\tau_{I,L} = 20\,\mathrm{min}$, a lead of at most $\Delta_{lead} = 40\,^\circ\mathrm{C}$, and $K_M = 2$ with $\tau_{I,M} = 10\,\mathrm{min}$. In Manual, on the first Auto period after Manual and after a change of structure, each reset term is set so that its loop's output equals the setpoint last applied. Measured on case 3, every transfer from Manual or from Air only into Lead/lag keeps the first period's setpoint within 0.01 °C of the one last applied, which the technical review added after finding steps of up to 7.3 °C. The selector cannot wind up the loop it passes over. In case 4 the lead sat at its 40 °C clip for 49.2 min with no windup, and the largest setpoint change in one period at a change of selection was 0.75 °C against 4.32 °C elsewhere in the run.

6.5.4 The Cure Observer and the Feedforward

The observer integrates Equation (6.4) at each of the twelve part thermocouples on its measured temperature, by fourth-order Runge-Kutta over each control period at the mean of the period's two readings, from the initial cure the load sheet states. It reports the lowest $\hat{\alpha}$ and its $\hat{T}_g$ for the demold rule, and for each part a predicted self-heating rate and excess, as shown in Equation (6.9).

$$\begin{aligned} \hat{s}_p &= \frac{w_r H_r}{c}\,\dot{\alpha}\left(T_{p,1}, \hat{\alpha}_{p,1}\right), \qquad \widehat{\Delta T}_p = \frac{\hat{s}_p L_p^2}{2a} \\ T_F &= T_{lim} - \min\left(80\,\mathrm{K},\ K_{ff}\max_p \widehat{\Delta T}_p\right) \end{aligned} \tag{6.9}$$

where $T_{p,1}$ is the mid-thickness reading. The excess is the quasi-steady rise of a slab of thickness L_p heating itself at the present rate and cooled through one face. The feedforward candidate T_F enters the low selector only on the Lead/lag + FF structure, in Auto, with the observer on and during the ramps, the dwell and the hold, with $K_{ff} = 0.4$. A feedforward subtracted after the selector made the lag loop's tracking ratchet during the build, which is why it is a candidate instead. The bound $\widehat{\Delta T}$ is loose and one-sided. On case 4 on Lead/lag it reads 656.8 K at 106.0 min while the true rise of the spar's core over its hotter face peaks at 62.4 K, and in the suggested challenge run it reads 101.1 K against a true 9.7 K, hence the small gain. With the feedforward acting it held in all 1011 control periods of case 4 in which it exceeded 1 K. Without the feedforward, once the reaction has burned out, the heat it left in the core exceeds the collapsed rate's estimate by up to 36.6 K at 115.3 min, because a rate that has stopped no longer predicts the heat already released. For the shipped structures no control model is identified, so this copy of the kinetics, with the load sheet's thicknesses and the declared

laminate and resin, is their only internal model, and the MPC's own model is given in Section 6.5.6. It knows the declared resin exactly but sees no node between the thermocouples and believes the lot's stated initial cure. A heat of reaction or a laminate changed in the Model dialog changes the plant and leaves the observer on its declared values. On case 4 with H_r raised to 700 J/g the spar then peaks at 215.2 °C on Lead/lag + FF.

6.5.5 The Recipe Optimizer

The optimizer is a grid search run offline on the application's own engine. For one structure it runs every recipe with hold temperatures of 175, 180, 185 and 190 °C and hold times of 10, 20, 30, 45, 60, 90 and 120 min on the lagging clock, without a dwell, to demold on the case's seed, and keeps the shortest feasible recipe judged on the truth with margins of 0.005 on the cure and 2 °C on the exotherm. Its result is shipped as the Optimized recipe, 185 °C for 10 min with the feedforward and 180 °C for 30 min on Lead/lag alone, and the application's tests rerun the search and confirm the same answers. The optimizer is not a receding-horizon controller and acts on no running cycle.

6.5.6 A Shrinking-Horizon Predictive Controller

The fourth structure replaces the lag loop with an MPC, a receding-horizon solution of a constrained optimal control problem [20] (Section 1.6.3), that writes the candidate u for the air setpoint. The air loop of Equation (6.7) stays beneath it, and three shipped protections of Lead/lag + FF stay above it in a low select, $T_{air,SP} = \min(u,\ r+\Delta_{lead},\ T_F,\ T_{lim}+o_M)$, so the lead clamp, the feedforward candidate and the max loop of Equations (6.8) and (6.9) can each lower the air under the plan, in the external-reset form already described. Industrial MPC is layered the same way, above the regulatory loops and beneath the plant's limit and safety logic [21]. While the MPC is selected, the lag loop's reset follows the applied setpoint, so its gain and integral time no longer act.

The controller's model is the heat network of Equation (6.3) written again from the declared constants of Table 6.2, never from the run's parameters, on 3 nodes through the load sheet's thickness of the skin, frame and shadowed part and 5 for the spar and the fitting, whose cores the thermocouples miss. Each part keeps its bag film $U_{s,p}A_p$, its contact h_cA_p and a lumped tool, with h_p from Equation (6.2) at the measured vessel pressure rounded to 0.5 bar and rebuilt when that rounding moves. The closed air loop is one lag, $\tau_a\dot{\hat{T}}_g = u - \hat{T}_g$ with $\tau_a = 44.6$ s, the sum of the time constant and dead time of a first-order-plus-dead-time fit ($K = 1.003$, 9.2 s, 35.4 s, residual 2.8 % of the response) to a ±10 °C step test on the air setpoint at 150 °C. One integrating air offset d_p per part enters its bag film and its tool's exposure. The reaction enters as a source along the plan, as Equation (6.10) states.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{b}\,u_k + \Gamma\bar{\mathbf{s}}_k, \qquad \hat{\mathbf{y}}_k = \mathbf{C}\mathbf{x}_k \\ \bar{s}_{k,i} &= \frac{w_r H_r}{c}\,\frac{\tilde{\alpha}_i(t_{k+1}) - \tilde{\alpha}_i(t_k)}{\Delta t_m} \end{aligned} \tag{6.10}$$

Here $\mathbf{x}$ holds the air, every node, every tool and every d_p (23 states, 30 with the fifth part), $\mathbf{A}$, $\mathbf{b}$ and Γ are the exact discretization at the decision period $\Delta t_m = 300$ s, and $\hat{\mathbf{y}}$ holds the 12 thermocouple positions and the two interior nodes of the spar (14 outputs, 19 with the fitting's three and two). The cure $\tilde{\alpha}_i$ of every model node is integrated with the declared kinetics of Equation (6.4) along the plan's own predicted temperatures, by fourth-order Runge-Kutta in 60 s steps, from the observer's $\hat{\alpha}$ at the thermocouples, linear between them. Each decision simulates this nominal path, solves, integrates the kinetics again along the new plan and solves again, with the source relaxed half and half between passes, up to six passes or until the plan moves less than 0.5 °C. Run open loop on the setpoints and cure increments of whole cycles, the model stays within 0.55 to 1.06 °C root mean square of each thermocouple on the standard and optimized recipes and with the fifth part (2.6 °C at worst), 0.78 to 1.49 °C on case 4, and 20.5 °C on case 6's shadowed part, whose shadow of 0.1 the model takes as 0.45. An independent 10 °C air step at 80 °C, with a twin run subtracted, keeps every thermocouple within 0.42 °C of the model over 3 h.

A steady-state Kalman filter, its gain from the discrete Riccati equation at 7 bar gauge, estimates **x** from the air and the twelve thermocouples with a measurement deviation of 0.25 °C and random walks per decision of 1.0 °C for the air, 0.3 °C for the nodes and tools and 1.5 °C for the offsets. Between decisions the estimate is propagated every control period with the setpoint actually applied and the observer's own cure increments, and it is corrected at each decision. The offsets make the plan offset free [22], as the fuel bias of the boiler-turbine MPC in Volume 9, Chapter 4 does, and absorb a wrong film, a heavier tool, a shadow and the reaction heat the model misses, all as air that a part does not receive.

Every 30 control periods, on the clock rounded to the control period and first on the priming call at $t = 0$, the plan covers what is left of the recipe, $N \le 72$ steps of 5 min (6 h), in 12 blocks re-spread over N at each decision, one step each while $N \le 12$ and from 1 to 11 steps at $N = 72$. The horizon ends at the end of the hold on its time or, with the observer's end rule, where the plan's predicted cure at every thermocouple reaches $\alpha \ge 0.875$ and $T_g \ge 210\,^{\circ}\mathrm{C}$. With less than one step left the last setpoint is held. The recipe itself is the preview: the reference r_k is the sequencer's on the lagging clock, extended ahead with its ramps, dwell and hold, which the lag loop also follows, so there is no preview to switch off. The plan solves Equation (6.11).

$$\begin{aligned} \min_{u,\varsigma} \; & \sum_{k=1}^{N} \sum_{j \in \mathscr{L}_k} (\hat{y}_{j,k} - r_k)^2 + S \sum_{\ell} \Delta u_\ell^2 + \sum_i \left(\rho_2 \varsigma_i^2 + \rho_1 \varsigma_i\right) \\ \text{s.t. } & \hat{y}_{j,k} \le T_{lim,k} - 0.5 - \tfrac{1}{6}\Delta T_{rxn,j,k} + \varsigma_j \\ & \hat{y}_{o,k} \le 193 - \tfrac{1}{6}\Delta T_{rxn,o,k} + \varsigma_o, \qquad \varsigma \ge 0 \\ & 20 \le u_k \le \min\left(225,\; r_k + \Delta_{lead},\; T_{lim,k} + 40\right) \end{aligned} \tag{6.11}$$

Temperatures are in °C. The set $\mathscr{L}_k$ holds the coldest predicted thermocouple position, tracked both ways as the lag loop tracks it, and every other position predicted below r_k, with unit weight, and $S = 0.1$ prices the block moves Δu_ℓ. The first ceiling holds the twelve thermocouple positions j 0.5 °C under the selector's limit $T_{lim,k}$, the lowest it takes in step k, and the second the spar's and fitting's interior nodes o 2 °C under the 195 °C exotherm limit. $\Delta T_{rxn,i,k}$ is the plan's own predicted rise of output i at step k from the reaction heat still to come (the response to the source term alone from zero), so the plan keeps room for a heat of reaction up to a sixth above the declared 600 J/g, the 700 J/g at the end of the parameter's range, and ς_i is the slack of output i. Both ceilings are soft. Each output carries a single slack for the whole horizon, priced at $\rho_2 = 10^3$ per °C^2 and $\rho_1 = 3 \times 10^4$ per °C, and the linear price is large enough that a ceiling the plan can meet is met exactly [23]. The bounds on u are hard: the setpoint's range, the lead clamp at the reference at the block's start, and with the selector on its loop's output clip. The first move may differ from the applied setpoint by 15 °C and each later block boundary by 15 °C per step between block centers, a ramp rather than one cap per boundary. The plan has at most 12 move variables and 14 to 19 slacks. Per-step slack cost 10 to 50 times the solve (case 4 115 ms against 6 ms) and moved no limit consistently, and no plan of the first build's failing runs predicted a violation (Section 6.7.7), so it was not adopted.

A plan is accepted when the solver reports optimal, when the first move, the planned inputs and outputs, the objective and the slacks are all finite with the expected dimensions, the slacks non-negative, the first move equal to the first planned move and every planned setpoint inside 20 °C to 225 °C. Otherwise Lead/lag + FF takes the air, whose lag loop has tracked the applied setpoint, at no more than 0.5 °C a control period until it meets its own request, the historian is marked, and the next accepted plan resumes. Faults injected at 60, 135 and 150 min handed over at 0.18, 0.26 and 0.09 °C a period with the noise off and resumed with steps of 5.19, 3.14 and 0.05 °C. No fallback occurred in the 80 runs of the MPC over the comparison, the worst cases and the parameter extremes. The plan reads only the packet, the load sheet, $\hat{\alpha}$, the recipe and its own state. A replay of a changed plant's packets into a fresh controller reproduced all 36 decisions, and the truth test catches 18 of 18 doctored copies.

Table 6.4 Code Verification and Agreement With the Source's Tests

Check	Result
Code verification	
Energy ledgers over the challenge cycle, air and lead/lag	6.6×10^{-12}, 6.2×10^{-12} of the heater energy
Two disturbed runs (fifth part, aged lot, heater loss, leak)	3.5×10^{-12}, 3.1×10^{-12}
Reaction ledger against $\rho w_r H_r \sum V \Delta\alpha$	equal to 4 decimals (9.9578 MJ)
Case 4 peak, app scheme at 0.1 s, explicit RK4 method of lines at 0.05 s	within 0.071 and 0.072 °C of 243.87, $\lvert\Delta\alpha\rvert \leq 3.33 \times 10^{-4}$
11, 21, 41 nodes: shipped, suggested, case 4 peaks	188.48/188.49/188.51, 189.30/189.31/189.31, 243.87/244.47/244.43, verdicts kept
Step 5 to 2.5 s, peaks	shipped 0.057, case 4 0.023 °C
Each case loaded first, from the list and after Reset	identical, 10 of 10
Agreement with the source's tests [14]	
26 resin constants against the report	identical
α after 120 °C, 480 min	0.636 against 0.636
α after 150 °C, 360 min	0.786 against 0.788
α after 180 °C, about 200 min	0.894 against 0.898 and 0.904
Peak rate, 120/150/180 °C	5.25×10^{-5}, 2.15×10^{-4}, 7.28×10^{-4} s^{-1} at 105, 24.9, 6.7 min, about the source's
T_g at α 0.52, 0.70, 0.90	110.7, 158.9, 218.0 against 108.2, 165.9, 218.1 and 226.7 °C
Lowest viscosity on 1, 2, 4 °C/min ramps	5.6, 4.1, 3.2 Pa s against about 6, 4, 3

A rule written before any search, judged on case 9 alone (verdict and margins, then the shortest cycle), set the weights from a flat 81-setting grid (280.3 to 284.8 min). A decision costs 5.3 ms on average in Node on case 9 (17.0 ms at most). In Chrome, in the page's own frame loop at 1200 times real time under a load average of 9.3 from parallel work, the logged decision averaged 9.9 ms (41.7 ms at most), 2.5 % of a core, with 59.9 frames per second. Because a decision costs more than 2 ms, a frame guard skips decisions, never frames, while the page lags real time.

6.6 Code Verification and Agreement With the Source

Code verification, agreement with the published fits and educational use are kept apart in Table 6.4. The first block checks that the code integrates the equations as written. The second compares the kinetics with the tests from which they were fitted, so it measures the fit and the transcription, not an independent validation. The source has two isothermal tests at 180 °C, which end at 0.898 and 0.904, and over its eleven isothermal tests the glass-transition relation is within 13.0 °C of the measured values with a root mean square of 5.3 °C, which is the scatter of the fit itself. No part or autoclave was measured for the model, and its use in teaching has been tried only in review runs. No learning study has been made.

Case 4's peak also depends on the noise draw. Over six seeds it spans 243.8 to 246.0 °C, and without noise it is 244.58, 244.04, 243.37 and 243.27 °C at steps of 5, 2.5, 1 and 0.5 s, so every run lies far past the 240 °C the case's prediction offers.

6.7 Nine Cases and a Challenge, Measured

The cases appear in Table 6.5, and Section 1.6.4 describes how a case is loaded and judged. Each starts from the standard recipe on Air only, pressure from $t = 0$, the default tunings and the default seed, and departs only as its row says. Each value below comes from the headless engine, advanced step by step as in the browser, with each change named in the text made before Run. Table 6.6 gathers the outcomes from the application's scorecard, judged on the truth at demold.

6.7.1 The Lag, by Hand and on the Air's Clock

Case 1 holds the air by hand. With the setpoint stepped to 110 °C at 1 min, the coldest thermocouples of the skin, the frame, the shadowed part and the spar reach 105 °C 10.1, 29.8, 57.8 and 82.4 min after the air does. At 30 min the spar reads 95, 78 and 71 °C from the bag side to the tool face, so the invar tool, not the laminate, holds it back. In case 2 the air enters the hold band at 214.9 min, the hold

Table 6.5 The Cases

#	Case and settings	Teaching point
1	Manual, air setpoint by hand	each part's lag
2	Standard recipe on Air only	the spar lags into the hold
3	The same, to be run on Lead/lag	slowest ramps, hottest limits
4	35 mm spar, Lead/lag, 3 °C/min, no dwell	a runaway the selector cannot stop
5	Pressure on at 240 min	voids frozen at gel
6	Shadow 0.1, Lead/lag	a cycle as long as its worst part
7	Lead/lag, no dwell, 180 °C for 10 min	the observer and demold
8	Aged lot (α_0 0.12), pressure at 170 min	gelation comes sooner
9	Optimized recipe, MPC	one plan against three rules
10	Challenge: the standard recipe on Air only	minimum cycle under limits

Table 6.6 Measured Outcomes at Demold, Default Seed. Cycle (min), True Peak (°C), Least-Cured Point and Minutes Outside the Hold Band

#	Configuration	Cycle	Peak	α_{min}	Band
2	as shipped	438.3	188.5	0.887	78.4
3	Lead/lag	463.2	186.1	0.896	4.2
4	as shipped (Lead/lag)	392.5	243.9	0.895	0.25
	Lead/lag + FF	392.2	184.8	0.895	—
	Lead/lag, 0.5 °C/min	541.3	184.7	—	—
5	as shipped, voids 12.49 %	438.2	200.5	0.875	142.2
6	as shipped (Lead/lag)	766.8	185.0	0.895	8.7
	Air only	534.3	—	0.865	179.7
7	as shipped	280.7	192.6	0.858	0.1
	hold ends on the observer	304.0	—	0.881	—
8	as shipped, voids 1.19 %	438.3	192.2	0.886	100.0

— not part of the case's measured record.

begins counting at 220.0 min, and the spar's tool face enters the band 37.5 min after the air. Over the cycle the part thermocouples spend 78.4 min outside the band, the worst 22.6 °C below it, and demold comes at 438.3 min.

6.7.2 Lead/Lag: the Slowest Ramps, the Hottest Limits

Case 3 runs the same recipe on Lead/lag. With the Maximum selector switch off, so with no max loop, the lag loop leads the air by up to 40 °C for the spar and the skin follows the air, to 139.1 °C in the 110 °C dwell and 202.7 °C three minutes into the hold, with 124.2 min outside the band. With the switch on, the max loop first takes the air at 63.6 min as the skin nears its 115 °C limit and holds it for 134 min of the cycle in all, the skin stops at 114.9 °C in the dwell and 185.1 °C in the hold, and only 4.2 min are spent outside the band. While the max loop held the air, the hottest thermocouple stayed within 5.30 °C under its limit. The cycle lengthens to 463.2 min, 24.9 min more than on the air's clock, because every hold now waits for the spar. Lead/lag buys the spar its recipe, not time.

🕮 **Pause.** Try Exercise 6.2 before reading on.

6.7.3 The Exotherm a Selector Cannot Stop

Case 4 thickens the spar to 35 mm and ramps it at 3 °C/min with no dwell, as shown in Figure 6.2. On Lead/lag the max loop has held the air since well before the runaway, and the air setpoint falls from 172.9 °C at 100 min to its floor of 125 °C (the limit less the 60 °C clip of the max loop) by 104.5 min. The core still passes 195 °C at 101.8 min, while the hottest thermocouple reads 187.3 °C, and peaks at 243.9 °C at 109.1 min. Retuning does not help. A limit of 180 °C still gives 231.0 °C, a max loop at $K_M = 6$ and 5 min gives 243.9 °C, and without the max loop the core reaches 301.5 °C. The screenshot of Figure 6.4 is taken at the moment the core crosses the limit.

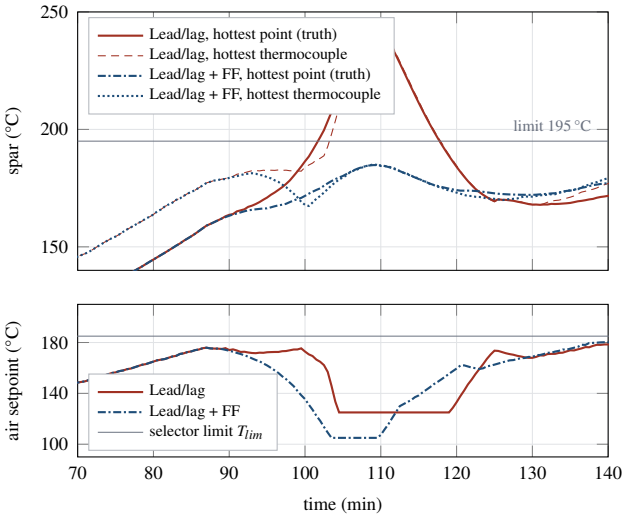


Figure 6.2 The exotherm of case 4, a 35 mm spar ramped at 3 °C/min. Top, the spar's hottest point (teaching truth) and its hottest thermocouple on Lead/lag (solid and dashed) and on Lead/lag + FF (dash-dotted and dotted), with the 195 °C limit as a thin line. Bottom, the applied air setpoint on Lead/lag (solid) and on Lead/lag + FF (dash-dotted) and the selector's limit (thin line). On Lead/lag the max loop cuts the air to its floor after the core is already running away.

> **TRY IT IN THE APP** **4 · The Exotherm: A Thick Spar at 3 °C/min**
>
> Load the scenario and press Run. On the Autoclave view, watch the spar's three thermocouples and the true peak badge. The core passes 195 °C at 101.8 min while the hottest thermocouple still reads 187.3 °C, and it peaks at 243.9 °C. Set Control structure to Lead/lag + FF, press Reset and run again: the spar now peaks at 184.8 °C.

The feedforward acts on the reaction instead of its temperature. Its candidate is first selected at 89.5 min, it reaches its 80 °C cap at 103.5 min, and at 100 min it has the air setpoint at 135.5 °C where the max loop had 172.9 °C. The core peaks at 184.8 °C in a cycle of 392.2 min against 392.5 min without it. The effect belongs to the observer's prediction and not to a static cap, because $K_{ff} = 0$ reproduces Lead/lag exactly (243.9 °C), and gains of 0.1, 0.2 and 0.4 give 239.2, 215.2 and 184.8 °C. A slower recipe also keeps the spar under the limit, though not at 1 °C/min (225.3 °C), and at 0.5 °C/min it peaks at 184.7 °C in a cycle 149 min longer, 541.3 min.

6.7.4 Pressure, Shadow, the Observer and an Aged Lot

Case 5 pressurizes the vessel at 240 min. The skin gels at 212.3 min and the frame at 238.0 min under vacuum alone, the voids formed while their resin flowed are frozen in, and the porosity index ends at 12.49 %. With the pressure on at 0, 165, 175 and 180 min the index ends at 0.00, 0.30, 0.81 and 1.30 %, so the step must come tens of minutes before the first gel. The late pressure also starves the parts of heat, the spar lags into the hold 87.7 min instead of 37.5 min, and when the pressure arrives its exotherm reaches 200.5 °C.

Case 6 pushes the shadowed part to a factor of 0.1, about 7.3 W/(m^2 K) at pressure. On Lead/lag it carries the lagging thermocouple for 94 % of the dwell and hold, the recipe waits for it, and the cycle takes 766.8 min against 463.2 min at 0.45. On Air only the cycle is 534.3 min, but the shadowed part reaches the band 126.6 min after the air and is left at $\alpha = 0.865$, under target. A lead/lag cycle is as long as its worst part, and on the air's clock the same part is under-cured.

Table 6.7 The Challenge, Default Seed. Cycle (min), True Peak (°C), Least-Cured Point, Minutes Outside the Band

Policy	Cycle	Peak	α_{min}	Band	Verdict
standard recipe, Air only (shipped)	438.3	188.5	0.887	78.4	fails
standard recipe, Lead/lag	463.2	186.1	0.896	4.2	fails
standard recipe, Lead/lag + FF	472.5	184.5	0.895	0.1	fails
standard hold, no dwell, Lead/lag	390.7	192.6	0.896	0.1	fails
185 °C, 10 min, Air only	275.5	203.4	0.808	9.8	fails
185 °C, 10 min, Lead/lag	285.8	195.3	0.878	0.2	fails
180 °C, 30 min, Lead/lag (its optimum)	301.2	192.6	0.879	0.1	passes
182 °C, 15 min, Lead/lag	288.0	193.8	0.872	—	passes
185 °C, 10 min, Lead/lag + FF (suggested)	289.8	189.3	0.878	0.5	passes

Case 7 holds 180 °C for 10 min on Lead/lag with no dwell. At demold (280.7 min) the least-cured point is at $\alpha = 0.858$ with T_g 205.0 °C, and the observer reports 0.858 and 205.1 °C. Ending the hold on the observer lengthens it from 21.7 to 45.0 min of clock time, because the hold counts only while the lagging part is in the band, and demolds at 304.0 min with $\alpha = 0.881$ and T_g 212.0 °C. Over both runs the observer stays within 0.014 of the true degree of cure at every thermocouple.

Case 8 starts an aged lot at $\alpha_0 = 0.12$ with the pressure step at 170 min, a time that suits a fresh lot. The skin gels at 200.8 min instead of 208.3 min, and the porosity index ends at 1.19 % against 0.50 % for a fresh lot. On six seeds it spans 1.178 to 1.188 %, all over the 1 % threshold. With the pressure at 150 min it ends at 0.16 %.

6.7.5 The Challenge

The challenge asks for demold within 310 min with no point of any part above 195 °C, every point at $\alpha \geq 0.87$, a porosity index of at most 1 % and no more than 10 min outside the hold band, on the load and resin as shipped. Table 6.7 lists the policies measured. The shipped standard recipe on the air's clock fails on cycle and band. The suggested configuration, Lead/lag + FF on the optimized 185 °C hold of 10 min, passes in 289.8 min with a peak of 189.3 °C and $\alpha = 0.878$ (T_g 211.1 °C), and on five other seeds of the noise in 289.7 to 291.2 min, all passing. The 310 min limit lies between the optimizer's best on Lead/lag alone with its margins, 301.2 min, and the standard recipe on any structure, 438 min or more.

The feedforward buys margin, not minutes. With the optimizer's margins it is what lets the hold run at 185 °C, since the same recipe on Lead/lag alone peaks at 195.3 °C and fails, and Lead/lag's own margined optimum takes 11.4 min longer. Without the margins Lead/lag alone keeps up, and 182 °C for 15 min passes in 288.0 min (287.7 to 289.8 min on six seeds, all passing) with its peak 1.2 °C under the limit and $\alpha = 0.872$, where the suggested run keeps 5.7 °C and 0.008 in hand. The optimum is set by the cure margin and not by the edge of the grid. Extending the grid to holds of 0 and 5 min and 2.5 °C steps keeps both optima, the shorter holds miss the cure margin at every temperature (the best, 187.5 °C for 5 min, takes 294.0 min), and a 1 min grid would save 2.3 min with 185 °C for 8 min. A 50 mm spar is infeasible within the limit. None of 38 settings, over three structures, feedforward gains of 0.4 to 2, four recipes and further dwells and ramps, passed, and the shortest that kept it under 195 °C took 440.8 min. Figure 6.3 draws the policies of Table 6.7 as bars against the two limits.

6.7.6 Case 9: One Plan Against Three Rules

Case 9 runs the challenge's suggested recipe, 185 °C for 10 min on the lagging clock with no dwell, under the MPC. Every row below shares the seed, instruments and noise, runs to demold and keeps case 9's MPC settings. The MPC demolds at 291.3 min with a peak of 189.3 °C and $\alpha = 0.879$ (T_g 212 °C) in 37 solves, against Lead/lag + FF's 289.8 min, 189.3 °C and 0.878, a tie within the spread of the noise. On six seeds it takes 290.0 to 291.3 min against 289.7 to 291.2 min, all passing. The MPC keeps the lagging thermocouple 1.47 °C behind the reference on the second ramp on average,

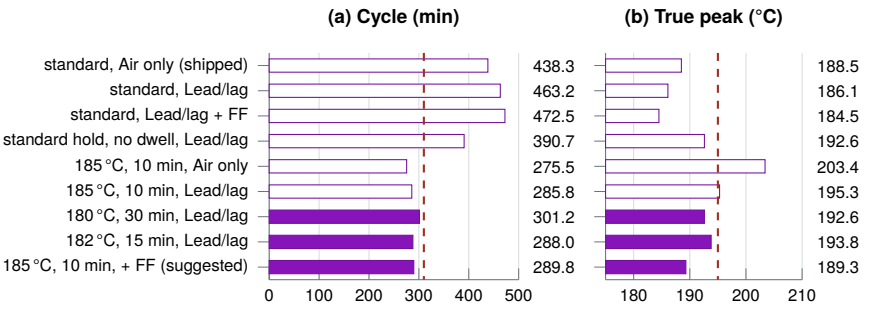


Figure 6.3 The policies of Table 6.7 on the challenge. (a) Cycle time with the 310 min limit dashed, (b) the true peak with the 195 °C limit dashed, dark bars for the policies that pass and white bars for those that fail. Notice that the three passing policies all sit just under 310 min, while the two fastest bars fail on the peak or the cure, and that the suggested run passes with the most room under the peak limit.

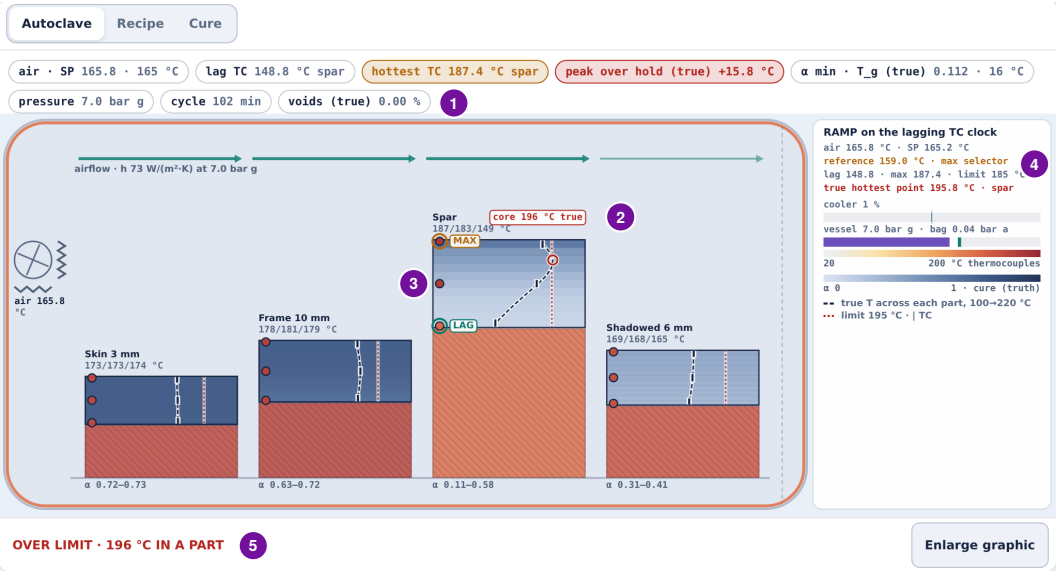


Figure 6.4 The core of a thick spar past the limit between its thermocouples, case 4 at $t = 101.8$ min on the Autoclave view, Lead/lag in Auto on the second ramp, cropped to the plant view. (1) The badges give the hottest thermocouple and the true peak over the hold. (2) The spar's dashed true profile crosses the red dotted 195 °C line between its thermocouple ticks, ringed as "core 196 °C true" against readings of 187, 183 and 149 °C, (3) with MAX on its bag side and LAG on its tool face, while the skin, frame and shadowed part read 165 to 181 °C. (4) The side panel gives the air at 165.8 °C against a setpoint of 165.2 °C chosen by the max loop (the panel reads "max selector"), the reference at 159.0 °C, the limit at 185 °C, the true hottest point at 195.8 °C and the vessel at 7.0 bar gauge with the bag at 0.04 bar absolute, and (5) the status strip reads OVER LIMIT.

against 3.96 °C, with 318 °C of setpoint travel against 362 °C. The feedforward candidate above it then holds the air for 62.5 min of the recipe (56.3 min under Lead/lag + FF), the last thermocouple enters the hold band at 179.6 min against 179.1 min, and both cool-downs take 101 min.

The three rows required beside the headline are in Table 6.8. Lead/lag + FF, retuned a single time under the rule that tuned the MPC and over all six of its settings ($K_L = 10$, $\tau_{I,L} = 40$ min,

Table 6.8 Cycle (min) and True Peak (°C) Under Lead/Lag + FF as Shipped, Retuned Once by the MPC's Rule, Fed the MPC's Prediction, and Under the MPC. Default Seed, Each Run to Demold

Case	FF	retuned	FF + pred.	ret. + pred.	MPC
9, optimized	289.8/189.3	281.3/190.0	291.5/189.2	281.5/190.0	291.3/189.3
standard recipe	472.5/184.5	452.7/185.4	474.7/184.4	468.5/185.0	456.8/184.5
4, 35 mm spar	392.2/184.8	367.5/240.0	401.7/184.0	384.7/194.6	414.8/184.0
50 mm, optimized	365.8/231.7	335.0/279.7	380.5/190.4	348.7/225.7	429.8/187.6
4, H_r 700 J/g	402.7/215.2	368.2/317.6	409.5/185.4	397.8/237.3	413.5/184.2
4, k_z 0.4	420.3/222.2	384.2/283.1	431.2/197.1	398.2/218.0	426.5/184.5
6, shadow 0.1	772.0/184.9	726.8/185.3	774.7/185.0	798.0/185.0	752.2/184.8

$K_M = 4$, $\tau_{I,M} = 2$ min, $K_{ff} = 0.1$, $\Delta_{lead} = 40$ °C) beats the MPC on case 9, 281.3 min at 190.0 °C, with 771 °C of travel, but the same tuning runs case 4's spar to 240.0 °C with 15.8 min above 195 °C. The strongest classical form of what the MPC knows feeds the max loop the MPC's own prediction, the hottest node over the next 30 min with the air held, from its estimator and its model with the kinetics. That row ties case 9 (291.5 min) and is faster than the MPC on the thick spars (case 4 in 401.7 against 414.8 min, the 50 mm spar in 380.5 against 429.8 min) at peaks up to 2.8 °C higher, but it loses case 4 with $k_z = 0.4$ (197.1 °C, 4.9 min above 195 °C) and the standard recipe (474.7 min).

The MPC's measurable gain is on the long recipes, as shown in Figure 6.5. On the standard recipe it demolds at 456.8 min against 472.5 min. The lagging thermocouple runs on average 0.6 °C behind the reference on the first ramp against 7.1 °C, and 2.3 °C against 7.3 °C on the second, so the hold begins at 225.9 min against 240.8 min, and the feedforward holds the air for 70.0 min under the MPC against 107.8 min. The retuned row is faster there (452.7 min) but leaves 13.9 min outside the band against 0.4 min. The same gain appears with the aged lot (460.8 against 475.2 min), the fifth part (455.8 against 473.2 min) and the shadow of case 6 (752.2 against 772.0 min, with no thermocouple over the selector's limit against 2.9 min). The aged lot reaches both structures through the same observer. With $\alpha_0 = 0.12$ and the load sheet at 0 both lose 7.0 min on the optimized recipe (MPC 295.5 against 288.5 min with the sheet right, Lead/lag + FF 293.3 against 286.3 min), and the fifth part leaves them at 289.5 and 289.7 min.

6.7.7 A Safety Defect in the First Build and Its Price

The first build kept only the max loop above the plan and set its ceilings without the reaction margin. Over 46 worst cases, every parameter of Table 6.2 that the application lets a reader change at both ends of its range plus the shadow of 0.1, a doubled spar tool, the aged lot and the fifth part, each on the optimized recipe and on case 4, the technical review found it worse than Lead/lag + FF on the selector's limit in 12. With $H_r = 700\,\mathrm{J/g}$ on the optimized recipe it peaked at 194.3 °C with a thermocouple more than 0.5 °C over the limit for 4.8 min (3.15 °C at most), where Lead/lag + FF kept 189.4 °C and none. The slack was not the cause. Every one of the 36 plans of that run had zero slack, so the plan believed its hottest thermocouple sat on its ceiling. At 125 min it predicted 178.4 °C five minutes ahead, the thermocouple read 184.7 °C at 130 min and 193.2 °C at 135 min, and the spar's estimated air offset grew from about 0.5 °C before 100 min to 7.7 °C at 130 min, because the filter read heat released in the core as air the part was not receiving. The same MPC told the true heat of reaction peaks at 190.4 °C with no minute over the limit, in 283.3 min, so the error was the model's heat of reaction. With $h_0 = 5$ on case 4 the first build also ran the air 114.5 °C above the reference, where the lead clamp allows 40 °C.

The fix was chosen by a rule written before any candidate ran. A row failed when the MPC spent more than 0.1 min longer than Lead/lag + FF either with a thermocouple 0.5 °C over the limit or with any point above 195 °C, and among the fewest failures the shortest case 9 won, then the fewest minutes above 195 °C. The rule first compared A, ceilings lowered by a sixth of the plan's predicted reaction rise, B, the shipped feedforward candidate above the MPC, and both together. Its result was AB, with 2 rows failing (case 4 with $h_0 = 5$ for 8.2 min at up to 3.9 °C, and $E_1 = 55\,\mathrm{kJ/mol}$

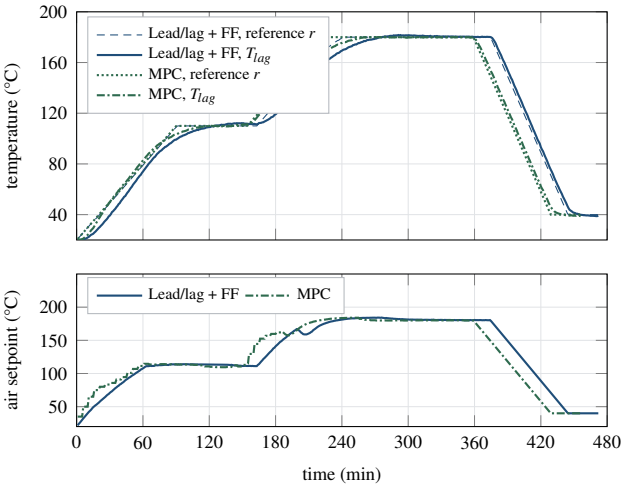


Figure 6.5 The standard recipe under Lead/lag + FF and under the MPC. Top, the recipe's reference on the lagging clock (dashed under Lead/lag + FF, dotted under the MPC) and the lagging thermocouple (solid and dash-dotted). Bottom, the applied air setpoint, the MPC's after the protections above it (solid under Lead/lag + FF, dash-dotted under the MPC). The MPC keeps T_{lag} closer to r on both ramps, so its hold begins 14.9 min sooner.

Table 6.9 Candidates for the Behavior Fix: Rows of 46 Worse Than Lead/Lag + FF, Case 9 and Case 4 Cycles (min), and Minutes Above 195 °C Summed Over the 46 Rows

Candidate	Rows	Case 9	Case 4	Σ min
first build	12	281.0	384.5	47.3
A, reaction margin	6	281.8	399.5	36.9
B, feedforward above	3	291.3	398.3	26.8
AB (original rule)	2	291.3	409.8	15.5
D, lead clamp above	7	281.0	389.0	43.0
AD	4	282.0	403.7	36.4
BD	0	291.3	402.0	16.1
ABD (kept)	0	291.3	414.8	8.8

by a margin) and case 9 at 291.3 min. The rule was then amended once, after these results, to add D, the lag loop's lead clamp kept above the MPC and modeled in its plan, a bound the shipped controller puts on its own command, with its combinations, as listed in Table 6.9. ABD was kept. Its price is case 9 from 281.0 to 291.3 min and case 4 from 384.5 to 414.8 min, against Lead/lag + FF's 392.2 min. The cheaper BD, the feedforward and the clamp without the margin, also fails no row and is 12.8 min faster on case 4, but spends 16.1 min above 195 °C summed over the rows where every structure goes over, against 8.8 min. The first build's late-pressure runs had also left porosity indices of 0.30 % (case 5) and 0.57 % (case 8) against none under Lead/lag + FF, from air run far ahead of the reference under vacuum alone. With the lead clamp above it both end at 0.00 %. Porosity is not in the plan, since no shipped protection acts on it.

6.7.8 Worst Cases Under Every Structure

Table 6.10 runs the worst cases the application offers under every structure, the first build and the MPC as kept, with the two limits that the challenge and the selector judge. After the fix the MPC never puts a thermocouple more than 0.5 °C over the selector's limit and no point above 195 °C in any of these rows, and it has the lowest or an equal peak wherever another structure passes 195 °C. Over the 46 parameter extremes it goes above 195 °C only where every structure does, with $k_z = 0.3$

Table 6.10 Worst Cases Under Every Structure: Cycle (min) / True Peak (°C) / Minutes Any Point Is Above 195 °C / Minutes a Thermocouple Is More Than 0.5 °C Over the Selector's Limit, the First Two Values Above the Last Two in Each Cell. Default Seed, Each Run to Demold. "First MPC" Is the Build Before the Fix of Section 6.7.7

Case	Air only	Lead/lag	Lead/lag + FF	FF retuned	first MPC	MPC
4, 35 mm spar, 3 °C/min	274.8/279.9 26.1/51.7	392.5/243.9 16.1/18.8	392.2/184.8 0/0	367.5/240.0 15.8/21.5	384.5/185.2 0/0	414.8/184.0 0/0
4 at 4 °C/min	261.0/279.3 26.0/55.4	391.3/244.3 16.1/18.9	391.3/184.8 0/0	366.0/239.6 15.8/21.2	373.7/184.7 0/0	405.2/183.9 0/0
50 mm spar, standard	452.7/265.5 39.5/55.1	583.7/209.5 15.8/53.7	558.8/183.0 0/0	536.3/208.9 15.1/43.3	513.0/186.4 0/0	547.3/182.9 0/0
50 mm spar, optimized	291.7/305.0 39.8/17.6	378.7/292.7 26.6/29.3	365.8/231.7 20.7/22.7	335.0/279.7 25.7/26.2	371.3/191.4 0/0	429.8/187.6 0/0
5, pressure at 240 min	438.2/200.5 10.0/19.3	633.0/184.7 0/0.3	641.8/184.5 0/0	626.7/185.1 0/19.8	618.7/184.8 0/0	634.7/184.4 0/0
8, aged, pressure at 170 min	438.3/192.2 0/14.8	574.2/184.8 0/0.1	589.7/184.4 0/0	570.0/185.2 0/3.8	557.5/184.7 0/0.8	582.7/184.4 0/0
6, shadow 0.1	534.3/188.5 0/8.5	766.8/185.0 0/6.2	772.0/184.9 0/2.9	726.8/185.3 0/5.0	723.8/184.8 0/0	752.2/184.8 0/0
optimized, H_r 700 J/g	275.5/219.8 15.8/17.2	307.0/200.2 5.2/8.7	289.5/189.4 0/0	282.7/192.8 0/3.2	284.8/194.3 0/4.8	288.5/189.4 0/0
4, H_r 700 J/g	274.8/363.7 33.2/52.6	394.8/311.7 17.8/20.6	402.7/215.2 11.8/15.7	368.2/317.6 17.4/25.7	389.0/192.9 0/9.4	413.5/184.2 0/0
4, k_z 0.4 W/(m K)	276.3/311.8 37.4/63.3	413.2/284.4 20.9/23.7	420.3/222.2 14.8/18.8	384.2/283.1 20.7/27.7	414.2/200.6 8.0/14.5	426.5/184.5 0/0
optimized, h_0 8	308.5/205.1 12.4/7.6	375.7/194.4 0/11.5	372.0/189.6 0/0	349.0/197.6 4.5/7.3	345.3/190.5 0/0	362.3/189.5 0/0
optimized, spar tool 2×	316.5/192.2 0/2.8	410.3/186.2 0/0	413.7/189.7 0/0	382.8/190.0 0/0	380.2/189.8 0/0	402.7/189.7 0/0
optimized, selector off	275.5/203.4 12.6/15.3	289.8/201.8 30.0/61.0	289.7/189.3 0/0	281.0/190.3 0/1.9	281.0/190.8 0/0	291.3/189.3 0/0
optimized, heater fails at 60 min	275.5/203.5 12.7/15.5	301.2/193.0 0/7.7	289.7/189.4 0/0	281.3/190.0 0/0	281.0/190.4 0/0	290.0/189.4 0/0

on the optimized recipe (197.5 °C for 4.8 min against Lead/lag + FF's 202.1 °C for 8.0 min) and with $E_1 = 55\,\mathrm{kJ/mol}$ on case 4 (210.8 °C for 4.1 min against 217.0 °C for 5.2 min), and with $w_r = 0.45$ on case 4 it peaks at 190.2 °C against 302.0 °C. Its heater saturates for at most 1.2 min (case 4 with a 100 kW heater), and every structure saturates 10.3 to 11.0 min with a 8 MJ/K wall. The vacuum leak leaves a porosity index of 4.30 % under the MPC and 4.23 to 5.41 % under the others, because the pressure is not the MPC's handle.

The MPC loses time wherever the protections above it bind. Case 4 takes 414.8 min against Lead/lag + FF's 392.2 min, and the 50 mm spar on the optimized recipe 429.8 min against 365.8 min, although there the feedforward lets the core reach 231.7 °C with 20.7 min above 195 °C and the MPC holds it at 187.6 °C. Two pre-existing defects were fixed in the same round. The log row at $t = 0$ carried the run's peak as $-\infty$ under every structure, because the referee's peaks rose only at the first step. They now start at the opening temperature, 20 °C, which moved one cell of the challenge's log and no other run. The observer and the feedforward read the run's H_r, w_r, c and k_z instead of the declared values, so a changed resin reached Lead/lag + FF. On case 4 its peak with $k_z = 0.4$ was 196.6 °C before the fix and is 222.2 °C now, and with $H_r = 700\,\mathrm{J/g}$ 202.6 °C and 215.2 °C, while the MPC, whose model was already declared, peaks at 184.5 °C and 184.2 °C.

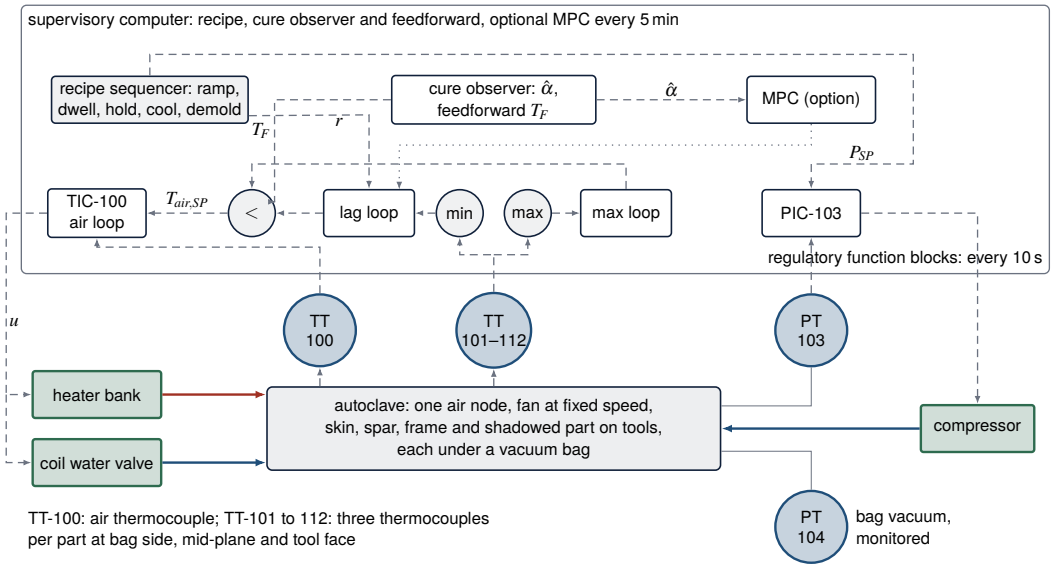


Figure 6.6 The chapter's loops on plant hardware. TIC-100 drives the heater and the coil's water valve on one split-range signal u. Minimum and maximum selectors feed the lag and max loops, and a low selector ($<$) passes the smallest candidate, with the observer's feedforward T_F, as the air setpoint. The sequencer sets the reference r and the pressure step of PIC-103. The optional MPC (dotted) replaces the lag loop. Notice that no loop reads the core of the spar.

6.8 Technology in Practice

What Measures It

The controlled variables of this chapter are thermocouple readings: the coldest keeps the recipe's clock and the hottest meets the limit. A process specification for cure panels states that all its temperatures are part temperatures and asks for thermocouples that monitor and record representative panels [24]. In one survey of a full-size autoclave, fifteen thermocouples on a large mold were recorded at intervals of 1 min, and one of them failed and showed an abnormal fluctuation, so its point was left out of the analysis [19]. The model's twelve part thermocouples, each with a 10 s lag, sit at the bag side, the mid-plane and the tool face, the places a real layup can reach. None sits where a thick core runs hottest.

> **TECHNOLOGY** **Sensor: Thermocouples Laid Up With the Part**
>
> A thermocouple is a junction of two dissimilar metals that makes its own voltage, and the instrument compensates for the second junction where the wires end. It is less accurate than a resistance thermometer but rugged, with a wide span [25]. On a cured panel the junctions sit at the laminate's mid-plane near an edge that is trimmed off after cure, or between the part and its caul plate [24]. The most common failure is an open circuit, or burnout, after which electrical noise can read as a wildly varying temperature, so the input is set to fail high or low, whichever is safer [25]. Behind the maximum selector of this chapter a reading failed high would cut the air, and behind the minimum selector one failed low would stall the recipe. The model's 10 s lag and noise stand for these sensors.

What Moves It

Two kinds of final element act on the load. One split-range signal drives the heater bank and the valve of the cooling-water coil, and a compressor raises the vessel pressure, while the vacuum bag holds the laminate against its tool. A process specification applies the pressure at a stated rate in pounds per

square inch per minute and keeps a minimum vacuum under the bag through the rest of the cure [24]. The model's compressor moves the vessel at no more than 0.5 bar/min, and a reader can halve the heater or tear the bag to see what each final element buys.

> **TECHNOLOGY** **Final Element: Heater and Cooler on One Signal**
>
> An autoclave heats its gas, commonly air or nitrogen, and a fan drives it over the tools and parts for forced convection [19]. When one controller output drives two final elements, they are split-ranged. In exclusive sequencing both are shut at half output, one opens as the output rises above it and the other as the output falls below it, so the two never act together [25]. The air loop of this chapter sequences its heater bank and its cooling-water coil in that way. The gas is a design choice too. In one simulation of a large framed mold, helium in place of air cut the mold's temperature difference by 68.27 % and raised its heating rate by 32.76 % [19].

Where the Controller Runs

Programmable logic controllers (PLCs) hold user programs for logic, timing, counting and PID control, and they serve as local controllers within a supervisory scheme [26]. Supervisory and lower-level loops run with cycle times from milliseconds to minutes [26]. The simulator's control layers run every 10 s, and its predictive controller decides every 5 min. Read as plant hardware, the air loop, the two selectors and the lag and max loops are regulatory function blocks, the recipe sequencer is the sequential layer that switches from ramp to dwell, hold, cool and demold, and the cure observer, the feedforward and the predictive controller belong to a supervisory computer. Figure 6.6 draws these layers on the vessel's hardware.

> **IN PRACTICE** **Running the Cure on the Lagging Thermocouple**
>
> A public process specification for curing the qualification panels of one aerospace prepreg shows how practice writes such a recipe. All temperatures are part temperatures, and the ramps and holds are based on the lagging thermocouple: a hold at 225 °F starts when that thermocouple reaches 215 °F. Before the cure the bag is leak checked, losing no more than 1 inch of mercury over 2 min, the vacuum is held, and the pressure is applied at a stated rate before the heating starts [24]. The chapter's sequencer, with its hold band and its pressure step, follows the same pattern.

What Is Modernizing It

Two developments aim at what the thermocouples miss. Embedded optical fibers have taken real-time spectra of a curing resin and composite, and Bragg gratings written into the same fibers respond to strain [27]. An ultrasonic scanner placed under the tool plate has been designed to watch defects form during cure, first in an oven as a step toward the autoclave, and to validate physics-based models that predict porosity and fiber waviness [28].

6.9 Limitations

The resin is one published fit, and its agreement with the tests it was fitted to is consistency, not validation. Everything else about the load and the vessel is a teaching value, and each part is one-dimensional on a lumped tool in one well-mixed gas, so corners, ply drops, gaps and the flow field are absent. The porosity index rewards pressure before gel and knows no resin flow, so the dwell earns nothing and the optimizer drops it. The observer shares the plant's kinetics exactly and sees no node between thermocouples, and its slab bound is loose while the reaction runs and fails after burn-out. The supervisor's gains were tuned on case 4 and the challenge, the optimizer searched one load on a 28-cell grid, and its optimum is optimal only for that load and seed. The MPC's model has 3 or 5 nodes per part where the plant has 11, takes the declared film, shadow, tool and resin whatever the reader sets, omits the air's overshoot on a large step, the heater's rating, the porosity, the gel point and a pressure step ahead, and linearizes the reaction along its own plan with a margin sized for one

parameter, the heat of reaction, up to a sixth above its declared value. Its tuning rule was judged on case 9 alone, and the gain it shows on the long recipes was not tuned for them.

OPEN CHALLENGE **Measuring the Cure Instead of the Temperature**

Indirect cure sensors such as fluorescence, dielectric loss and ultrasound are often hard to correlate with the chemistry that sets the material's properties. In one fiber-optic Raman record, a temperature program then used in commercial autoclaves ran about two hours past full cure [29]. Fibers stripped to touch the resin are fragile and unlikely to survive the heat and pressure of processing [27]. Progress would be a rugged reading of the cure inside the part, so that a hold could end on the state of the resin, which this chapter's observer can only estimate from thermocouples.

6.10 Conclusion

A mixed autoclave load is controlled through its slowest and its hottest thermocouple, and this simulator logs both beside the true temperature and cure through every part. On the air's clock the spar's tool face enters the hold band 37.5 min after the air and the standard cycle takes 438.3 min. Lead/lag gives the spar its recipe at 463.2 min, with the leading skin held at its limit by the max loop through a low selector. A 35 mm spar ramped at 3 °C/min passes 195 °C while its hottest thermocouple reads 187.3 °C and peaks at 243.9 °C under that max loop, while the observer's feedforward keeps it at 184.8 °C at no cost in cycle time. On the challenge an optimized hold of 10 min at 185 °C demolds in 289.8 min, but Lead/lag alone passes in 288.0 min with less margin, so the feedforward buys margin rather than time. The model is one-dimensional, its resin one fit and its limits teaching values. A fuller study would add a two-dimensional part with its corners, a flow field with radiation, resin flow and consolidation that give a dwell its purpose, an observer whose kinetics differ from the plant's, and a predictive controller whose margin covers the laminate and the film as well as the resin. With the shipped protections above it, the MPC ties the feedforward on the optimized recipe and shortens the standard recipe by 15.7 min, while its first build, planning to the limit on its own model, was less safe in 12 of 46 worst cases. The cases show that a cure recipe has to be run on the part, and that the hottest point of a thick part is one no thermocouple reads.

KEY POINTS

- A mixed load is cured through its slowest and its hottest thermocouple, not through the air.
- On the air's clock the spar enters the hold band 37.5 min after the air, and Lead/lag gives it its recipe at 463.2 min.
- A limit loop on thermocouples acts after the measurement moves, so a 35 mm spar peaks at 243.9 °C under the max loop.
- The observer's feedforward keeps that spar at 184.8 °C, and on the challenge it buys margin rather than time.
- With the shipped protections above it, the MPC ties the feedforward on the optimized recipe and shortens the standard recipe by 15.7 min.

Where else this applies. The freeze-dryer of Chapter 5 has the same shape: its hottest vials limit the shelf while the slowest decide the endpoint.

References

[1] A. C. Loos and G. S. Springer, "Curing of epoxy matrix composites," *Journal of Composite Materials*, vol. 17, no. 2, pp. 135–169, 1983.

[2] T. A. Bogetti and J. W. Gillespie, "Two-dimensional cure simulation of thick thermosetting composites," *Journal of Composite Materials*, vol. 25, no. 3, pp. 239–273, 1991.

[3] J. S. Kim and D. G. Lee, "Development of an autoclave cure cycle with cooling and reheating steps for thick thermoset composite laminates," *Journal of Composite Materials*, vol. 31, no. 22, pp. 2264–2282, 1997.

[4] J. H. Oh and D. G. Lee, "Cure cycle for thick glass/epoxy composite laminates," *Journal of Composite Materials*, vol. 36, no. 1, pp. 19–45, 2002.

[5] A. H. Glattfelder and W. Schaufelberger, *Control Systems with Input and Output Constraints*, ser. Advanced Textbooks in Control and Signal Processing. London: Springer, 2003.

[6] P. R. Ciriscioli and G. S. Springer, *Smart Autoclave Cure of Composites*. Lancaster, PA: Technomic Pub. Co., 1990.

[7] V. K. Pillai, A. N. Beris, and P. S. Dhurjati, "Implementation of model-based optimal temperature profiles for autoclave curing of composites using a knowledge-based system," *Industrial & Engineering Chemistry Research*, vol. 33, no. 10, pp. 2443–2452, 1994.

[8] M. M. Thomas, J. L. Kardos, and B. Joseph, "Shrinking horizon model predictive control applied to autoclave curing of composite laminate materials," in *Proceedings of 1994 American Control Conference*, vol. 1. IEEE, 1994, pp. 505–509.

[9] P. Dufour, Y. Toure, D. J. Michaud, and P. S. Dhurjati, "Optimal trajectory determination and tracking of an autoclave curing process: A model based approach," in *1999 European Control Conference (ECC)*. IEEE, 1999, pp. 2048–2053.

[10] N. G. Pantelelis, "Towards the dynamic optimisation for the cure control of thermoset-matrix composite materials," *Composites Science and Technology*, vol. 65, no. 7-8, pp. 1254–1263, 2005.

[11] M. R. Kamal and S. Sourour, "Kinetics and thermal characterization of thermoset cure," *Polymer Engineering & Science*, vol. 13, no. 1, pp. 59–64, 1973.

[12] J. P. Pascault and R. J. J. Williams, "Glass transition temperature versus conversion relationships for thermosetting polymers," *Journal of Polymer Science Part B: Polymer Physics*, vol. 28, no. 1, pp. 85–95, 1990.

[13] J. M. Castro and C. W. Macosko, "Studies of mold filling and curing in the reaction injection molding process," *AIChE Journal*, vol. 28, no. 2, pp. 250–260, 1982.

[14] A. Shahkarami, "Material characterization for processing: Hexcel 8552, material model development, final project wrap-up," Version 1.0, reviewed by D. Van Ee and approved by A. Poursartip. Public document of the NCAMP 8552 program, National Institute for Aviation Research, Wichita State University, Nov. 2009. [Online]. Available: https://www.wichita.edu/industry_and_defense/NIAR/Research/hexcel-8552/Additional-Documents-3.pdf

[15] N. Rai and R. Pitchumani, "Optimal cure cycles for the fabrication of thermosetting-matrix composites," *Polymer Composites*, vol. 18, no. 4, pp. 566–581, 1997.

[16] G. M. Maistros and I. K. Partridge, "Monitoring autoclave cure in commercial carbon fibre/epoxy composites," *Composites Part B: Engineering*, vol. 29, no. 3, pp. 245–250, 1998.

[17] J. N. E. Kluge, T. S. Lundström, L. G. Westerberg, and T. Nyman, "Modelling heat transfer inside an autoclave: Effect of radiation," *Journal of Reinforced Plastics and Composites*, vol. 35, no. 14, pp. 1126–1142, 2016.

[18] J. W. Hou, "Optimal cure cycle design for autoclave processing of thick composites laminates: A feasibility study," Old Dominion University, for NASA Langley Research Center, Norfolk, VA, Progress Report, NASA Contractor Report NASA-CR-181261, Nov. 1985, NASA Technical Reports Server 19870018301.

[19] G. Zhang, B. Zhang, L. Luo, T. Lin, and X. Xue, "Influence of mold and heat transfer fluid materials on the temperature distribution of large framed molds in autoclave process," *Materials*, vol. 14, no. 15, p. 4311, 2021.

[20] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[21] 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.

[22] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[23] P. O. M. Scokaert and J. B. Rawlings, "Feasibility issues in linear model predictive control," *AIChE Journal*, vol. 45, no. 8, pp. 1649–1659, 1999.

[24] National Center for Advanced Materials Performance, "NCAMP process specification: Fabrication of NMS 380 qualification, equivalency, and acceptance test panels, Toray TC380 prepregs," Wichita State University, National Institute for Aviation Research, Wichita, KS, Process Specification NPS 83800, Revision B, Oct. 2024, distribution Statement A, approved for public release.

[25] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[26] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[27] K. H. Wood, T. L. Brown, M. C. Wu, and C. B. Gause, "Fiber optic sensors for cure/health monitoring of composite materials," NASA Langley Research Center, Hampton, VA, Tech. Rep., 2004, NASA Technical Reports Server 20040086070.

[28] T. B. Hudson, F. L. Palmieri, T. B. Abbott, J. P. Seebo, and E. R. Burke, "Design of an automated ultrasonic scanning system for in-situ composite cure monitoring and defect detection," in *SAMPE 2019 Conference*, Charlotte, NC, May 2019, NASA Technical Reports Server 20200002665, report NF1676L-31834.

[29] J. B. Cooper, K. L. Wise, and B. J. Jensen, "Methods used for monitoring cure reactions in real-time in an autoclave: Summary of research, NASA langley research center training grant NGT-1-52124," Old Dominion University, Norfolk, VA, Tech. Rep. ODURF-162661, Oct. 2000, NASA Technical Reports Server 20000108909.

6.11 Exercises

6.1 Concept check · 2 min

In case 4 a 35 mm spar ramped at 3 °C/min on Lead/lag peaks at 243.9 °C, although the max loop has already cut the air setpoint toward its floor. Why does the limit loop fail?

(a) Its gain is too low, and a gain of 6 with an integral time of 5 min would hold the core under the limit

(b) The low selector passes over the max loop while the lag loop still asks for more heat on the ramp

(c) The thermocouple noise hides the rise of the hottest reading until the core has passed the limit

(d) The core heats itself before any thermocouple moves, and a lower air cools only the faces of the spar

6.2 Concept check · 1 min

The standard recipe takes 438.3 min on Air only and 463.2 min on Lead/lag. What makes the Lead/lag cycle longer?

(a) Every dwell and hold counts its time only while the lagging spar thermocouple is in the band

(b) The max loop holds the air below the recipe for the whole cycle, so every ramp runs slower

(c) The lead clamp of 40 °C slows the ramp of the thin skin, which then reaches the hold late

(d) The lag loop's integral winds up while the max loop is selected and overshoots each hold

6.3 Concept check · 2 min

The application's cure target is $\alpha \geq 0.87$, which maps to $T_g \geq 208.7\,^{\circ}\mathrm{C}$, and not the 0.95 often quoted for a fully cured part. Why?

(a) The DiBenedetto relation is not defined above $\alpha = 0.9$, so a higher target has no glass transition

(b) At a 180 °C hold the diffusion term stalls the cure near 0.90, so 0.95 is never reached

(c) The porosity index freezes at gel, and this resin gels at $\alpha = 0.87$

(d) The observer's estimate drifts above 0.87, so a higher target could not be checked during the cure

6.4 Theory and calculation · 6 min

The spar is a 25 mm laminate with $\rho = 1580\,\mathrm{kg/m^3}$, $c = 1100\,\mathrm{J/(kg\,K)}$ and $k_z = 0.6\,\mathrm{W/(m\,K)}$, on an invar tool of $C_t = 231.8\,\mathrm{kJ/K}$ with $A_t = 1.8\,\mathrm{m^2}$. The heat transfer coefficient is $h_0 = 14\,\mathrm{W/(m^2\,K)}$ at one atmosphere, $P_0 = 1.013\,25\,\mathrm{bar}$, the bag adds $R_{bag} = 0.004\,\mathrm{m^2\,K/W}$, and the spar is not shadowed.

(a) Use Equation (6.2) to find h_p and U_s at 7 bar gauge, and the factor by which pressure raises h_p. *Checkpoint:* $h_p = 73.2\,\mathrm{W/(m^2\,K)}$, factor 5.23, $U_s = 56.6\,\mathrm{W/(m^2\,K)}$.

(b) Find the three time scales of the spar: the bag-side film $\rho cL/U_s$, the tool $C_t/(h_p A_t)$ and conduction L^2/a with $a = k_z/\rho c$. *Checkpoint:* 12.8 min, 29.3 min and 30.2 min.

(c) Repeat the film and conduction scales for a 35 mm spar and say which grows faster with thickness. *Checkpoint:* 17.9 min and 59.1 min. Conduction grows with L^2.

6.5 Theory and calculation · 7 min

The laminate has a resin mass fraction $w_r = 0.34$, a heat of reaction $H_r = 600\,\mathrm{J/g}$ of resin and a specific heat $c = 1100\,\mathrm{J/(kg\,K)}$, with $a = 3.45 \times 10^{-7}\,\mathrm{m^2/s}$.

(a) Find the adiabatic rise $w_r H_r/c$ of a laminate that cures fully, and the rise with $H_r = 700\,\mathrm{J/g}$, the top of the parameter's range. State the second as a fraction above the first. *Checkpoint:* 185.5 K and 216.4 K, one sixth higher.

(b) In case 4 on Lead/lag the observer's bound $\widehat{\Delta T}$ of Equation (6.9) for the 35 mm spar reads 656.8 K at 106.0 min. Find the self-heating rate $\hat{s}$ and the cure rate $\dot{\alpha}$ it implies. *Checkpoint:* $\hat{s} = 0.370\,\mathrm{K/s}$, $\dot{\alpha} = 2.0 \times 10^{-3}\,\mathrm{s^{-1}}$.

(c) With $K_{ff} = 0.4$ and $T_{lim} = 185\,^\circ\mathrm{C}$, find the feedforward candidate T_F at that moment, and compare the bound with the true rise of 62.4 K. *Checkpoint:* $T_F = 105\,^\circ\mathrm{C}$ at the 80 K cap. The bound is about ten times the true rise.

6.6 Theory and calculation · 5 min

The first kinetic term of the resin has an activation energy of 64 929.5 J/mol, E_1, and the second one of 133 168.3 J/mol, E_2. Its glass transition follows Equation (6.5) with $T_{g0} = -7\,^\circ\mathrm{C}$, $T_{g\infty} = 250\,^\circ\mathrm{C}$ and $\lambda = 0.78$.

(a) Near 450 K an Arrhenius factor doubles over about $\Delta T = \ln 2\, RT^2/E$. Find ΔT for each term. *Checkpoint:* 18.0 K and 8.8 K.

(b) Find T_g at $\alpha = 0.87$ and at 0.875, and relate them to the demold rule and to the observer's rule for ending the hold. *Checkpoint:* 208.7 °C and 210.2 °C.

6.7 App challenge · 9 min

Load scenario 3, Lead/lag: ramp for the slowest part, and set Control structure to Lead/lag. Run at the default seed and a speed of 1200, which plays 20 min of the cycle each second, so a cycle takes about 23 s. The run pauses itself at DEMOLD and writes a note with the cycle time, the peak, the least-cured point and the minutes outside the hold band.

(a) Run to DEMOLD and read the note. *Checkpoint:* 463 min, 186.1 °C, α 0.896, 4 min.

(b) Set the Maximum selector switch Off, press Reset and run to DEMOLD again. *Checkpoint:* 439 min, 202.2 °C, α 0.906, 124 min.

(c) Compare both runs with the standard recipe on Air only, which takes 438.3 min with 78.4 min outside the band, and say which part sets the ramp and which sets the limit. *Checkpoint:* The spar sets the ramp, the skin the limit.

Accept the cycle and the band within 1 min and the peak within 0.2 °C.

6.8 App challenge · 9 min

Load scenario 4, The exotherm: a thick spar at 3 °C/min, which opens on Lead/lag. Use the default seed and a speed of 1200, and run each case to DEMOLD, where the run pauses itself.

(a) Run as loaded and export the CSV file. Find the first row in which `truth_peak_c` exceeds 195 °C, its time `time_s` and the hottest thermocouple `max_c` in that row, and read the peak in the DEMOLD note. *Checkpoint:* 6110 s (101.8 min), 187.4 °C, peak 243.9 °C.

(b) Set Control structure to Lead/lag + FF, press Reset and run again. Read the peak and the cycle. *Checkpoint:* 184.8 °C in 392 min.

(c) Set the Exotherm feedforward gain to 0.1, press Reset and run again. *Checkpoint:* 239.2 °C.

Accept a temperature within 0.2 °C, a time within 10 s and a cycle within 1 min. Say what the three runs show about acting on the measurement and acting on the reaction.

6.9 App challenge · 8 min

Load the challenge, Cure the load in minimum time. Use the default seed and a speed of 1200, and run each case to DEMOLD, where the run pauses with the verdict.

(a) Run the shipped settings and read the verdict and the cycle. *Checkpoint:* Fails on cycle time and hold band, 438 min with 78 min outside the band.

(b) Press Reveal suggested tuning, then Reset, and run again. *Checkpoint:* Passes in 290 min, peak 189.3 °C, α 0.878.

(c) Set Control structure to Lead/lag, which leaves the Recipe at Optimized and writes Lead/lag's own optimum, then Reset and run again. Read the hold temperature and time the recipe now shows, the verdict and the cycle. *Checkpoint:* 180 °C for 30 min, passes in 301 min, peak 192.6 °C.

Accept a cycle within 1 min and a peak within 0.2 °C. Explain why Lead/lag + FF can run the hotter, shorter hold.

6.10 Application area · 3 min

A shop proposes to switch the autoclave pressure on later in the cycle to save compressor time. Use the measured results of case 5 and case 8, both on the standard recipe and Air only, with the 1 % porosity threshold.

(a) How late may the pressure come on for a fresh lot, and why must it come well before the first part gels? *Checkpoint:* About 175 min (0.81 %), tens of minutes before the first gel.

(b) What does an aged lot ($\alpha_0 = 0.12$) do to that answer? *Checkpoint:* At 170 min the index is 1.19 %. At 150 min it is 0.16 %.

6.11 Application area · 3 min

The suggested configuration of the challenge demolds in 289.8 min against 438.3 min for the standard recipe on the air's clock.

(a) Find the minutes saved per cycle, the fraction saved, and the cycles that fit in 24 h of vessel time with no loading time. *Checkpoint:* 148.5 min, 33.9 %, 3.29 against 4.97 cycles.

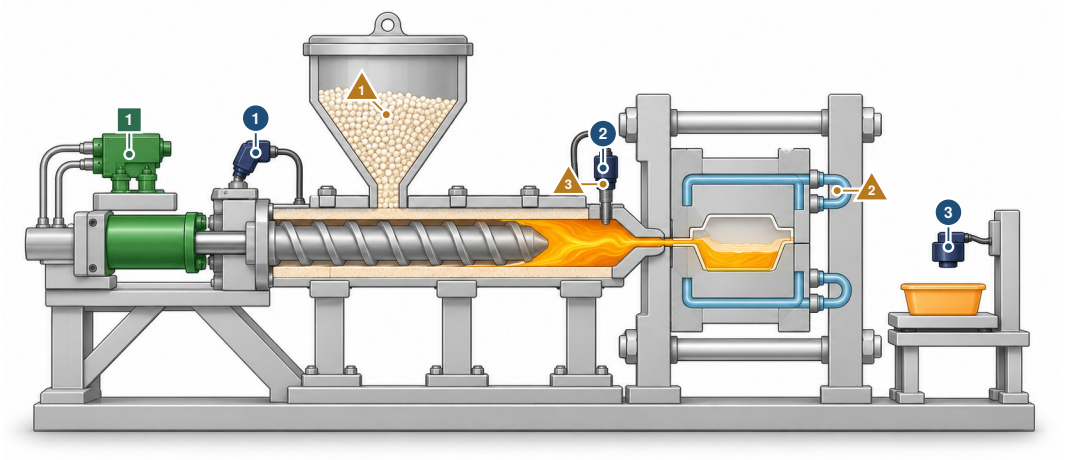
(b) Name two features of a real load that the model leaves out and that a thermocouple survey must check before the shorter recipe is adopted. *Checkpoint:* For example corners and ply drops, and the flow field.

6.12 Review the agent · 5 min

An AI assistant reviews the case 9 results and writes: "Lead/lag + FF retuned once ($K_L = 10$, $\tau_{I,L} = 40$ min, $K_M = 4$, $\tau_{I,M} = 2$ min, $K_{ff} = 0.1$, lead 40 °C) demolds the optimized recipe in 281.3 min at 190.0 °C, which is 10 min faster than the MPC and 8.5 min faster than the shipped tuning, with every limit met. Ship this tuning as the default for every load." Find the flaw and say how you would check it.

CHAPTER 7

Injection Molding Simulator for Bumpless Switchover and Shot-to-Shot Mass Correction



The controller moves 1 servo valve that drives the ram, for both fill and pack

The controller reads 1 screw position and speed 2 melt pressure ahead of the screw 3 weight of each finished part

What disturbs it 1 new resin lot or recycled feed 2 less coolant flow 3 drifting pressure reading

One valve fills the cavity at a set speed and then packs it at a set pressure, and only the scale at the right sees what each part weighs. Illustration, not to scale.

IN PLAIN TERMS One Valve, Two Jobs and a Scale

The main idea. A molding machine pushes melted plastic into a cold steel mold, first at a set speed, then at a set pressure while it shrinks. One valve does both jobs, so the controller hands it from speed to pressure in every shot without a jolt.

The challenge. A part's weight is fixed when the narrow entrance to the mold freezes, and no sensor sees that moment. A thicker batch of plastic can make every part light while speed and pressure stay on target.

Why it matters. Weighing each part and nudging the pressure for the next shot keeps parts right, if the nudges may grow big enough.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history and the first fact in Section 7.2, the second fact and the two try-its in Section 7.5, and the key points
Study	45 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 7.5), Technology in Practice (Section 7.6), Limitations (Section 7.7) and the open challenge

The control question Two loops take turns on one valve inside every shot, and neither of them measures the part. How is the valve handed from the speed loop to the pressure loop without a bump, and how much authority does a correction from the scale need to hold part weight?

Learning objectives

- Explain how tracking the applied valve makes the switchover bumpless.
- Relate the switchover position to a short shot or to flash.
- Derive the loop gain, the stability bound and the variance factor of the shot-to-shot correction.
- Predict part length from ejection temperature, and explain why weight alone does not certify a part.
- Choose the gain and the limits of the correction for a drifting campaign.

Prerequisites Sections 1.4.2 (phases and the handover between controllers), 1.4.5 (run-to-run control, stability and delay), 1.6.2 (velocity form and tracking) and 1.6.4 (scenarios and data).

The app https://apmonitor.com/apps/injection-molding/

See also Volume 2, Chapter 2 for velocity-form PID with bumpless transfer between manual and automatic, Volume 2, Chapter 3 for a loop held at an actuator bound by back-calculation, and Volume 2, Chapter 12 for a drive whose mechanical resonance caps the speed loop.

7.1 Introduction

The loops met in a first control course keep their actuator for as long as the plant runs, and a loop that holds its setpoint is taken to be doing its job. Cyclic manufacturing breaks both habits. The actuator passes from one controller to another at events inside every cycle, and the quality of the product is decided by states that none of the loops measures. Transferring the actuator without a bump then matters as much as tuning either controller [1, 2], and so does correcting the recipe between cycles from a measurement of the product [3].

Injection molding is one of the most versatile processes for mass-producing plastic parts in net shape [4], and it breaks both habits in a cycle of a few seconds. A reciprocating screw pushes melt into a closed, cooled mold under speed control until a switchover point, after which a pressure loop holds the melt while the part shrinks and the gate freezes [5, 6]. Both loops drive the same hydraulic valve. The part weighs what the cavity holds when the gate freezes, and that depends on the pressure reaching the cavity through a narrowing, cooling gate and on the temperature of the melt at that moment [7, 8]. A more viscous resin lot, a drifting pressure transducer or a warmer mold therefore change the part while both loops sit on their setpoints. Recycled resin brings its own variation into the hopper, and the recycled share has room to grow, since about 9 % of the 6300 Mt of plastic waste generated by 2015 had been recycled [9].

The simulator of this chapter tests the belief that a process whose loops all hold their setpoints is making good product. In it the fill speed and the pack pressure can stay on their setpoints while the part comes out 1.4 % light, and the parts can weigh the right amount while their length walks out of tolerance. The broader class has two halves, and both are general. Any system that changes controllers at an event, such as a start-up handing over to production or a normal loop yielding to an override, needs the idle controller to follow the applied actuator so that the transfer produces no bump [1, 2], the handover at a phase change of Section 1.4.2. Any process that repeats a cycle, such as a batch reactor, a semiconductor process step or an additive build, can correct its recipe between cycles from a measurement of the product, and run-to-run, batch-to-batch and iterative learning control are one family of methods for doing so [3] (Sections 1.4.5 and 1.4.6). The additive build of Chapter 10 learns a correction along each layer from the layers before it.

This chapter works with an injection molding simulator in which the switchover point, the handover between the two loops, the hold time and a bounded shot-to-shot correction can each be changed and their consequence read off a scale. Velocity-to-pressure switchover, bumpless transfer by tracking and run-to-run correction are all established practice [4, 10]. The model of one shot has seven states.

It conserves mass exactly, lets a frozen skin close the gate on its own clock, and keeps the true states apart from what the encoder, the two pressure sensors and the scale report. Every equation and parameter is given, so that the model can be rebuilt. Eight scenarios and a thirty-part campaign separate what the inner loops control from what decides the part. Section 7.2 traces the control of injection molding. Section 7.3 sets out the model of one shot, its instruments and its disturbances. Section 7.4 describes the cycle sequencer, the two loops that share one valve and the shot-to-shot correction with its stability bound. Section 7.5 reports the runs of every scenario and of the campaign, Section 7.6 describes the transducers, valves and controllers of real machines, Section 7.7 states the limitations and Section 7.8 sums up. The exercises in Section 7.9 close the chapter.

7.2 Control of the Molding Cycle

Quantitative modeling of the cycle began with the cavity. A 1972 model wrote the continuity, momentum and energy equations for the polymer in the cavity through filling, packing and cooling, with a non-Newtonian melt whose density and viscosity depend on temperature, and it needed the pressure history at the cavity entrance as an input [7]. A finite-element and finite-difference simulation of the filling stage followed in 1980 [11]. For control the cavity had to be joined to the machine that feeds it. A physically based model of the whole cycle divided it into the hydraulic system, the ram and screw, the barrel and the delivery system. It had to account for the elasticity of the melt and its non-Newtonian flow, it included the growing solid skin in the sprue, runner and gate, and it agreed with experiments on high density polyethylene [8]. The simulator of this chapter keeps the same subsystems in lumped form.

By 1987 a review of injection molding process control could sort the controlled variables into those regulated in every phase, those triggered during a specific phase, and cycle-to-cycle control, in which data from earlier cycles are used to predict trends and take corrective action. The same review set the conventional proportional-integral and PID loops beside self-tuning, optimal and statistical process control [10]. The phase-dependent category contains the two loops of the simulator, and the cycle-to-cycle category contains its correction. Cycle-to-cycle and within-cycle adaptive control of the nozzle pressure during packing and holding followed at the end of the 1990s, with the nozzle pressure preferred to the cavity pressure because one transducer then serves different molds, and with dynamics found to be nonlinear and time varying [12]. Learning control was applied to the electrohydraulic injection unit [13], adaptive control to the filling velocity [14], and robust iterative learning control to the molding process [15].

> ⧖ HISTORY **1986 to 1994: Correcting the Next Shot**
>
> A doctoral thesis of 1996 on cavity pressure control reviewed how learning from one shot for the next had entered molding. A study published in 1986 and 1987 joined a hydraulic servo valve to a microprocessor to reduce the variation of the peak cavity pressure, acting from cycle to cycle rather than within the cycle. A 1991 proposal applied learning control to the cavity or nozzle pressure during filling, so that the error at each instant of one cycle corrected the same instant of the next. It worked in simulation but was not shown on a machine. A 1992 auto-adaptive controller concluded that resins of different viscosity could make parts of constant weight. A 1994 study moved the cavity pressure during cooling by the coolant over successive cycles, because the mold's large thermal mass rules out acting within one [16].

The switchover itself has been compared method by method. Seven triggers, from screw position and injection time through machine, nozzle, runner and cavity pressures to a cavity temperature at the end of flow, were set to give similar part weights and then subjected to perturbations of six standard deviations of the standard process. Conventional triggers such as screw position gave lower short-term variation, while the others rejected long-term variation better [5]. A 2005 review organized the field into process setup, machine control, process control and quality control, and it found that online quality control without human intervention had yet to be realized, primarily for

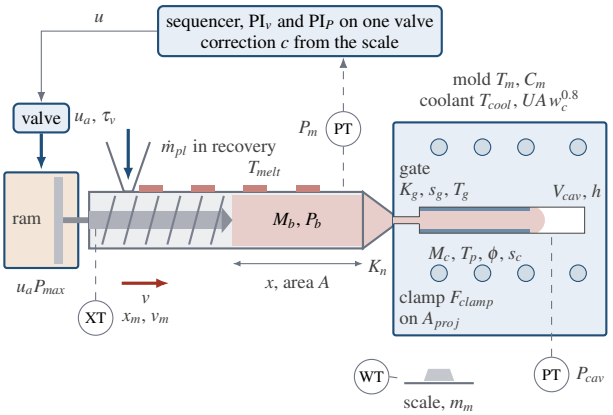


Figure 7.1 One shot of the simulated injection molding machine. The servo valve opening u_a follows its command u through a lag τ_v and sets the ram force, expressed as the melt pressure $u_a P_{max}$, which drives the screw at speed v against the melt pressure P_b of the mass M_b ahead of it, at position x from the front. The melt passes the nozzle K_n and the gate, of open fraction s_g and temperature T_g, into the cavity of volume V_{cav} and wall h, where it fills a fraction ϕ under a frozen skin (dark lines) with mass M_c and temperature T_p. The mold at T_m is cooled through $UA\,w_c^{0.8}$ by coolant at T_{cool}, and the clamp F_{clamp} over A_{proj} sets the flash limit. The encoder XT, the melt pressure transducer PT, the cavity sensor at 80 % of the flow path and the scale WT are all the controller reads. During recovery the rotating screw adds melt at $\dot{m}_{pl}$.

lack of transducers that measure quality in real time and of robust models linking the process to quality [4]. Weighing each part and correcting the next shot answers that gap directly. It belongs to the run-to-run and batch-to-batch family that surveys connect with iterative learning and repetitive control [3], whose exponentially weighted moving average (EWMA) form has a well-characterized stability and sensitivity [17]. The process control, monitoring and optimization of injection molding are now collected in a monograph [6].

Table 7.1 Variables of One Shot at the Nominal Recipe

Symbol	Role	Nominal	Range
v	CV in fill, screw speed	40 mm/s	SP 10–100
P_b	CV in pack, melt pressure	60 MPa	SP 20–120
m	CV per part, part mass	24.10 g	±1 %
u	MV, valve command (both loops)	54, 30 %	0–90 %
c	MV per part, pack correction	0 MPa	$\pm\Delta_{total}$
x	state, screw position	48 to 14.5 mm	0–48
u_a	state, applied valve opening	as u	0–1
M_b	state, melt ahead of the screw	35.6 g	—
M_c, M_cT_p	states, cavity melt, enthalpy	0 to 24.1 g	—
T_g	state, gate temperature	230 °C	—
T_m	state, mold temperature	68 °C	—
$\bar{\lambda}$	DV, resin lot viscosity factor	1	0.6–2.0
σ_r	DV, regrind variation (1σ)	0	0–25 %
w_c	DV, coolant flow	100 %	5–100 %
b, $\dot{b}$	DV, transducer bias, drift	0, 0	±30, 0–5
L	output, length of a 100 mm feature	99.54 mm	±0.12

✦ DID YOU KNOW? Where the Energy of Plastics Goes

An estimated 272 TBtu of energy was used in 2010 to make plastics and rubber products in the United States. Of the savings that the best technologies already available could bring, injection molding holds the largest share, 25.6 TBtu a year or 30 percent [18].

7.3 Model of One Shot

The simulated machine has a 30 mm screw and fills one cavity of 24 cm^3 with a 2 mm wall and 100 cm^2 of projected area, held shut by an 800 kN clamp. The resin is a generic amorphous thermoplastic close to ABS, melted at 230 °C, and the coolant enters the mold at 60 °C. The nominal recipe fills at 40 mm/s, switches to a 60 MPa pack pressure at a screw position of 14.5 mm, holds it for 3.5 s and cools for 4.0 s more, and it makes a 24.107 g part every 10.86 s. The machine, the mold and the instruments are shown in Figure 7.1.

Table 7.1 classifies the variables, with the bias in MPa and the drift in MPa per minute. Seven states are integrated: the screw position x, the applied valve opening u_a, the melt mass ahead of the screw M_b, the melt mass in the cavity M_c and its enthalpy content M_cT_p, the gate temperature T_g and the mold temperature T_m. The screw speed, the pressures, the fill fraction and the flow are algebraic in these states.

7.3.1 Injection Unit

The servo valve sets the force on the ram, expressed as the melt pressure u_aP_{max} it could hold with the screw at rest, and the valve opening follows its command u through a first-order lag. The screw and ram are pushed with tens of kilonewtons (60 MPa on the 30 mm screw is 42 kN), so their inertia is neglected and the screw speed follows the linear force-speed line of the drive, as shown in Equation (7.1).

$$\tau_v \dot{u}_a = u - u_a, \quad v = v_{max}\left(u_a - \frac{P_b}{P_{max}}\right), \quad \dot{x} = -v \tag{7.1}$$

where x is measured from the front stop, a positive speed moves the screw forward, and v is set to zero when the screw rests on its stop. The drive runs at full speed with no load and stands still at its force limit, and the same line pushes the screw backward whenever the melt pressure exceeds the drive force, which is how the screw recovers against the back pressure. All of the drive lag sits in the valve.

The melt ahead of the screw is compressible. Its mass is the state, so mass is conserved exactly, and its pressure follows from a linear pressure-volume-temperature (PVT) law, Equation (7.2).

$$P_b = K\left[1 - \frac{V_d + Ax}{M_b\, v_0(T_{melt})}\right], \quad \dot{M}_b = \dot{m}_{pl}\,\delta_{rec} - \dot{m} \tag{7.2}$$

where $A = \pi D^2/4$ is the screw area, V_d the melt held in the nozzle and screw tip, K the bulk modulus, $\dot{m}$ the mass flow into the mold and δ_{rec} one during recovery and zero otherwise. The specific volume at pressure P is $v_0(T)(1 - P/K)$, and the zero-pressure value v_0 is made of two straight lines, melt and glass, joined smoothly at the glass transition T_{gl}, Equation (7.3).

$$\begin{aligned} v_0(T) = v_{gl} + a_S(T - T_{gl}) \\ + (a_M - a_S)\, w \ln\left[1 + e^{(T - T_{gl})/w}\right] \end{aligned} \tag{7.3}$$

with $w = 3\,\mathrm{K}$. The melt at 230 °C has $v_0 = 1.040\,\mathrm{cm^3/g}$ and a density of $965\,\mathrm{kg/m^3}$ at the 6 MPa back pressure, so the $36.9\,\mathrm{cm^3}$ ahead of the screw at the shot size holds 35.6 g, and the solid at 23 °C has $v_0 = 0.950\,\mathrm{cm^3/g}$.

7.3.2 Flow Path and Frozen Skin

Melt passes from the barrel through the hot nozzle, the gate and the filled length of the cavity. Each section is treated as a power-law slit of index n, and the pressure drops of the three add, Equation (7.4).

$$P_b - P_c = \lambda\left[K_n a_n + \frac{K_g a_g}{s_g^{2n+1}} + \phi_f\,\frac{K_c a_c}{s_c^{2n+1}}\right]|Q|^{n-1}Q \tag{7.4}$$

where Q is the volumetric flow, P_c the compression pressure of a full cavity, λ the lot factor of the resin in this shot, $\phi_f = \min(1, \phi)$ the filled share of the flow length, and s_g and s_c the open fractions of the gate and of the filled cavity. The viscosity shift is Arrhenius about 230 °C, $a(T) = \exp[(E/R)(1/(T + 273.15) - 1/503.15)]$, taken at the melt temperature for the nozzle and at a core temperature $(T_{melt} + T)/2$ for the gate and the cavity, so a cooling section also grows stiffer.

Melt touching the steel freezes into a skin that grows inward. If the skin temperature runs linearly from the mold temperature to the no-flow temperature around a core near the melt temperature, a section of lumped temperature T keeps the open fraction of Equation (7.5).

$$s(T) = \mathrm{clamp}\left(1 - \frac{T_{melt} - T}{T_{melt} - T_s},\, 0,\, 1\right) \tag{7.5}$$

where $T_s = (T_m + T_{nf})/2$ is the mean skin temperature, and $s = 0$ once $T \le T_s$. The skin grows slowly while hot melt keeps arriving and quickly once it stops. No flow is computed while the nozzle is shut, while s_g or s_c is below 10^{-3}, or once the gate has been declared frozen. The mass flow is taken at the density of the melt it leaves, $\dot{m} = QM_b/V_b$ for $Q \ge 0$ with $V_b = V_d + Ax$, and $\dot{m} = QM_c/(\phi_f V_{cav})$ for backflow. The skin is what lets an early switchover leave a part short, because without it a slowing front would still reach the end of the cavity.

7.3.3 Cavity, Gate and Mold

The fill fraction is the zero-pressure volume of the cavity melt over the cavity volume, and past one the melt is compressed, Equation (7.6).

$$\phi = \frac{M_c\, v_0(T_p)}{V_{cav}}, \quad P_c = \begin{cases} K(1 - 1/\phi) & \phi > 1 \\ 0 & \phi \le 1 \end{cases} \tag{7.6}$$

The pressure falls linearly from the gate to the flow front. With ΔP_f the drop across the filled cavity, the last term of Equation (7.4), the mean cavity pressure is $\bar{P} = P_c + \Delta P_f/2$, and a sensor at 80 % of

the flow path reads $P_s = P_c + \Delta P_f(1 - 0.8/\phi_f)$ once the front has passed it and zero before. The three thermal states follow Equations (7.7) to (7.9).

$$\dot{M}_c = \dot{m}, \qquad \frac{\mathrm{d}(M_c T_p)}{\mathrm{d}t} = \dot{m} T_{in} - \frac{M_c(T_p - T_m)}{\tau_p} \tag{7.7}$$

$$\dot{T}_g = -\frac{T_g - T_m}{\tau_g} + \frac{|Q|}{V_{gh}}(T_{in} - T_g) \tag{7.8}$$

$$\begin{aligned} C_m \dot{T}_m = {} & \frac{c_p M_c (T_p - T_m)}{\tau_p} - UA\, w_c^{0.8}(T_m - T_{cool}) \\ & - k_r(T_m - T_r) \end{aligned} \tag{7.9}$$

where T_{in} is T_{melt} for inflow and T_p for backflow, and $\tau_p = h^2/(\pi^2 \alpha) = 4.05\,\mathrm{s}$ is the slowest conduction mode of a slab of thickness h. The gate cools toward the mold with time constant τ_g and is renewed by the melt passing it, V_{gh} being the volume whose passage replaces it. The coolant flow fraction w_c, bounded to 0.02 to 2, scales the heat transfer as a turbulent film does, and $k_r = 3\,\mathrm{W/K}$ is a loss to the room at $T_r = 23\,^\circ\mathrm{C}$. The mold time constant $C_m/(UA + k_r)$ is 47.6 s at full coolant flow and 240 s at a tenth of it. The gate is reset to the melt temperature when the next shot begins, because the frozen gate leaves with the part.

7.3.4 Part Mass, Length and Defects

The gate is declared frozen once it would pass less than $0.01\,\mathrm{cm^3/s}$ under a 50 MPa drop, that is once $\lambda K_g a_g s_g^{-(2n+1)}$ exceeds $50 \times 10^6/(10^{-8})^n = 3.2 \times 10^{10}$ in the units of K_g. The test runs between the switchover and the ejection, and the latch holds until the next shot. Nothing enters or leaves the cavity afterward, so the mass weighed at ejection is the mass at gate freeze. In the nominal recipe the gate freezes 3.42 s after injection starts, 2.56 s after the switchover. The length of a nominally 100 mm feature is an illustrative proxy, Equation (7.10).

$$\begin{aligned} L &= 100\,\mathrm{mm}\left[1 - f s_V - k_{ej}(T_{ej} - T_{ej,0})/100\right] \\ s_V &= 1 - m\, v_0(T_r)/V_{cav} \end{aligned} \tag{7.10}$$

where m is the part mass, s_V its volume shrinkage from the cavity to the room, f the share of that shrinkage showing in the length while the mold restrains the part, and T_{ej} the part temperature at ejection, so that a part ejected warm finishes shrinking free of the mold. The nominal part shrinks 4.5 % in volume and measures 99.545 mm against $99.54 \pm 0.12\,\mathrm{mm}$.

A part is short if its fill fraction never reaches 0.999, light or heavy if its weighed mass is outside ± 1 % of the target, and labeled SIZE if its length is out of tolerance. It is flashed if the mean cavity pressure times the projected area ever exceeds the clamp force, $\bar{P} > F_{clamp}/A_{proj} = 80\,\mathrm{MPa}$, a flag that adds no mass. A shot whose valve command sits at its limit for more than 20 ms counts, like a flash, as a pressure excursion. The energy of a part is the drive work divided by an efficiency of 0.6, plus 720 kJ/kg for the melt plasticized and a 1200 W base load over the cycle, 32.8 kJ in the nominal recipe.

7.3.5 Instruments and Disturbances

The controller never reads a true state. The encoder rounds the screw position to 1 µm, and the measured speed v_m is its first difference over the 2 ms control period passed through an exact first-order filter of 8 ms. The transducer reading P_m follows the melt pressure through a 5 ms first-order lag and adds a bias b, a drift accumulated at the rate $\dot{b}$ from the start of the run, and white noise of 0.2 MPa. The cavity sensor adds 0.1 MPa of noise to P_s, and the scale adds 0.01 g to each part (1σ). The lot factor of every shot after the first is $\lambda_k = \bar{\lambda}\exp(\sigma_r z_k)$, log-normal and uncorrelated from shot to shot, where $\bar{\lambda}$ is the lot setting and σ_r the regrind variation. The deviates z_k and the scale errors are

Table 7.2 Machine and Material Parameters. A Marks a Value Adjustable in the Model Dialog and H a Hidden One

Symbol	Meaning	Value	Class
P_{max}	drive force as melt pressure	200 MPa	A
P_{lim}	injection pressure limit	180 MPa	A
v_{max}	screw speed at full drive, no load	400 mm/s	A
τ_v	valve and drive lag	20 ms	A
D	screw diameter	30 mm	A
V_d	melt in nozzle and screw tip	3 cm^3	H
x_{shot}	shot size, screw position	48 mm	A
$\dot{m}_{pl}$	plasticizing rate	8 g/s	A
P_{back}	back pressure	6 MPa	A
t_{open}	mold open, eject and close	2.5 s	A
T_{melt}	melt temperature	230 °C	A
T_{nf}	no-flow temperature	150 °C	A
T_{gl}, v_{gl}	glass transition, its volume	105 °C, 0.970 cm^3/g	H
a_M, a_S	melt, solid expansion	0.56, 0.24 mm^3/(g K)	H
K	melt bulk modulus	1.6 GPa	A
n	power-law index	0.35	A
E/R	viscosity activation temperature	6500 K	A
c_p	specific heat	2000 J/(kg K)	H
α	thermal diffusivity	0.10 mm^2/s	A

Table 7.3 Mold, Instrument, Quality and Energy Parameters

Symbol	Meaning	Value	Class
V_{cav}, h	cavity volume, wall thickness	24 cm^3, 2.0 mm	A
τ_g	gate thermal time constant	3.0 s	A
V_{gh}	melt volume that renews the gate	2 cm^3	H
K_n, K_g, K_c	consistencies, Pa $(s/m^3)^n$	$(3.4, 8, 15)10^8$	H
F_{clamp}, A_{proj}	clamp force, projected area	800 kN, 100 cm^2	A
C_m	heat capacity of the mold inserts	3.0 kJ/K	A
UA	coolant heat transfer at full flow	60 W/K	A
T_{cool}	coolant temperature	60 °C	A
$T_{m,0}$	mold temperature at the start	67.55 °C	A
—	transducer noise, lag	0.2 MPa, 5 ms	A
—	cavity sensor noise	0.1 MPa	A
—	encoder step, speed filter	1 µm, 8 ms	A
—	scale noise	0.01 g	A
—	mass tolerance	±1 %	A
L_{spec}	length specification	99.54 ± 0.12 mm	A
f	volume shrinkage seen in length	0.10	A
$T_{ej,0}$	ejection temperature reference	91.3 °C	A
k_{ej}	extra shrinkage of a warm part	0.015 %/K	A
—	drive efficiency	0.6	H
—	plasticizing energy	720 kJ/kg	A
—	base load	1200 W	A

drawn once from the scenario seed, so every configuration of a scenario molds the same material and reads the scale with the same errors.

The parameters are listed in Tables 7.2 and 7.3. Fixed in the code are the viscosity reference of 230 °C, the room at 23 °C, the 3 K join width, the sensor position, the room loss k_r and the gate-freeze threshold. Every scenario starts with the screw at the shot size, the melt ahead of it at the back

Table 7.4 Recipe and Controller Defaults, With Slider Ranges

Symbol	Meaning	Default	Range
v_{SP}	fill speed setpoint	40 mm/s	10–100
x_{sw}	switchover position	14.5 mm	5–40
P_{sw}, t_{sw}	switchover on P_{cav}, on time	15 MPa, 0.84 s	1–60, 0.2–3
t_{hold}, t_{cool}	hold, cooling time	3.5 s, 4.0 s	0.5–10, 1–20
P_{pack}	pack pressure setpoint	60 MPa	20–120
$K_{c,v}$, $\tau_{I,v}$	speed loop	1.0 %/(mm/s), 20 ms	0.05–4, 5–500
$K_{c,p}$, $\tau_{I,p}$	pressure loop	0.5 %/MPa, 50 ms	0.02–3, 5–1000
u_{fill}, u_{pack}	manual openings	40, 30 %	0–100
m_{SP}	part mass target	24.10 g	23–25.5
G	correction gain	40 MPa/g	0–200
Δ_{shot}, Δ_{total}	step, total limit	4, 15 MPa	0.5–20, 1–50
n_{start}	parts before correcting	3	0–10

pressure, an empty cavity, the gate at the melt temperature and the mold at $T_{m,0}$, which is the periodic operating temperature of the nominal recipe unless the scenario sets another.

7.3.6 Assumptions

The machine feeds one cavity straight from a hot nozzle, with no runner and a check ring that never leaks, so the shot size repeats exactly. Multi-cavity molds and runners add a balance problem that this model does not have.

The screw and ram have no inertia and the hydraulics no resonance. A machine with a soft hydraulic column would put a resonance inside the speed loop, as the coupling of the servo axis in Volume 2, Chapter 12 caps its speed gain.

Temperatures are lumped, one each for the part, the gate and the mold, and the skin of Equation (7.5) is a geometric construction on those temperatures rather than a solution of the conduction problem, which the cavity models of the literature solve [7]. The melt front has no temperature of its own, so a slow fill thickens the skin only through the mean part temperature. Real short shots also come from a front that stalls in a thin rib far from the gate.

The PVT law is linear in pressure and piecewise linear in temperature, and the viscosity is a power law with an Arrhenius shift. Near the glass transition a Tait equation and a WLF shift describe the melt better, while at the pressures and temperatures of this recipe the simple forms carry the trends.

Flash, the short-shot criterion and the length tolerance are illustrative thresholds, not measurements of a real mold, and the length proxy ignores anisotropic shrinkage and warpage. The scale weighs every part without delay, whereas many cells weigh a sample or report several shots late.

7.4 Sequence and Control as Implemented

The machine controller executes every 2 ms, and the seven states are integrated with a classical fourth-order Runge-Kutta step of 1 ms. Halving the step changes the part mass by 0.000 65 % and the peak mean cavity pressure by 0.08 %, inside the limits of 0.01 and 0.5 % that one of the application's own checks enforces. The historian logs at 50 Hz, and the playback speeds of 0.25 to 20 simulated seconds per wall second change only the display, as for every app of the framework (Section 1.6.1). The structure is drawn in Figure 7.2, and the recipe and controller defaults are listed in Table 7.4.

7.4.1 The Cycle Sequencer

A shot starts with the mold closed and the screw at the shot size. The fill ends at the switchover, triggered by the measured screw position reaching x_{sw}, by the cavity sensor reaching P_{sw}, or by the fill lasting t_{sw}, with a 3 s fill-time limit behind all three. The pack lasts the hold time. Recovery then rotates the screw, adding melt at 8 g/s, until the measured position is back at the shot size, while the drive holds $u = P_{back}/P_{max} = 3\,\%$. The cooling timer runs from the end of the hold, so recovery overlaps cooling. The mold then opens for t_{open}, the part drops onto the scale, the nozzle is shut, and

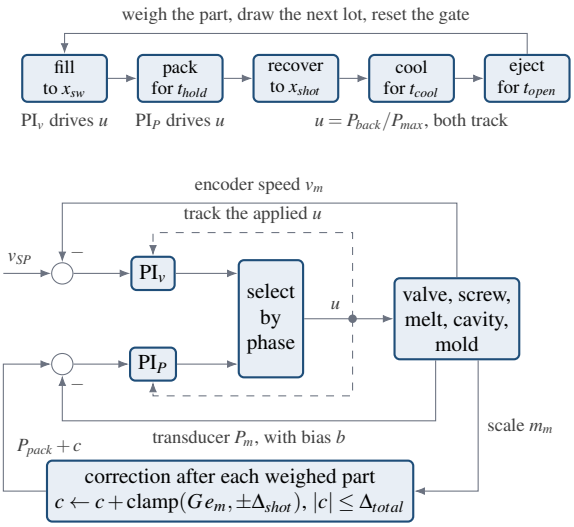


Figure 7.2 Sequencer and control structure. The fill ends at the switchover and the pack after the hold time, recovery refills the screw to the shot size, and the cooling timer runs from the end of the hold. The speed loop PI_v drives the valve during the fill and the pressure loop PI_P during the pack, while the loop not in charge tracks the applied valve command (dashed). Outside fill and pack the valve holds the back pressure. After each weighed part the correction moves the pack pressure setpoint by a bounded step computed from the mass error $e_m = m_{SP} - m_m$.

the next shot starts with a new lot factor and the gate at the melt temperature. The nominal cycle is 0.86 s of fill, 3.5 s of hold, 4.0 s of cooling and 2.5 s open, 10.86 s in all. In the terms of Section 1.4.2 the fill ends on a measured event, while the pack, the cooling and the mold opening end on a clock.

7.4.2 Two Loops on One Valve

Both loops use the shared PID of the framework, written in the velocity (incremental) form with derivative action on the measurement, Equation (7.11), which is Equations (1.14) to (1.17) of Section 1.6.2 with the action $a = +1$ and no output bias.

$$
\begin{aligned}
&e_k = r_k - y_k, \qquad P_k = K_c e_k, \qquad D_k = -K_c \tau_D \dot{y}_{f,k} \\
&S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\
&c_k = S_k + I_{k-1}, \qquad u_k = \text{clamp}\,(c_k,\ 0,\ u_{max}) \\
&I_k = I_{k-1} + \Delta t \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - c_k}{\tau_t} \right]
\end{aligned}
\tag{7.11}
$$

where r is the setpoint, y the measurement, y_f its optionally filtered value, c_k the output the algorithm demands (written v_k in Chapter 1) and $\tau_t = \tau_I$ the tracking time. The running sum S_k is never clamped, the clamp acts on the emitted output only and the back-calculation on the integral alone, as Section 1.6.2 explains. The integral used at period k was accumulated through period $k-1$, and both integrals start from the 3 % back-pressure opening. Neither loop uses derivative action ($\tau_D = 0$) or a rate limit, the sample time is $\Delta t = 2$ ms, and the output is limited to $u_{max} = 100 \min(1, P_{lim}/P_{max}) = 90$ %, the machine's 180 MPa injection pressure limit. Both loops are reverse acting in the ISA convention of Section 1.6.2, because opening the valve raises both the screw speed and the melt pressure.

The speed loop acts on the filtered encoder speed, and the pressure loop acts on the transducer reading toward the setpoint $P_{pack} + c$. During the fill the speed loop drives the valve and the pressure

loop runs in its tracking mode, on its own setpoint and measurement, with the integral chosen at every period so that its output equals the applied command, $I_k = u_k - S_k$. During the pack the roles reverse, with the idle speed loop tracking at zero error, and outside the two phases both loops track the 3 % back-pressure opening. At the switchover the pressure loop therefore starts from exactly the valve the speed loop left, and its first output differs from it only by its own proportional increment over one period, $-K_{c,p}(P_{m,k} - P_{m,k-1})$. In manual mode the operator sets one opening for the fill and one for the pack, each held below the 90 % limit, the sequencer runs as before, and both loops track, so a return to automatic is bumpless as well [1], by the same tracking as the manual and automatic transfer of the cruise control loop in Volume 2, Chapter 2. In automatic mode the two manual openings follow the last fill and pack commands, so a switch to manual mid-run holds the valve where the loops left it.

Pause. Try Exercise 7.2 before reading on.

7.4.3 Shot-to-Shot Correction

After each part is weighed, in automatic mode only and once more than n_{start} parts have been made, the pack pressure correction moves by a bounded step, Equation (7.12).

$$\begin{aligned} \Delta c_k &= \mathrm{clamp}\big(G(m_{SP} - m_{m,k}),\, \pm\Delta_{shot}\big) \\ c_{k+1} &= \mathrm{clamp}\big(c_k + \Delta c_k,\, \pm\Delta_{total}\big) \end{aligned} \tag{7.12}$$

where $m_{m,k}$ is the weighed mass of part k, and the setpoint in use is $\mathrm{clamp}(P_{pack} + c, 0, P_{lim})$. The correction is an integral action in the shot index with one shot of delay, and it reads only the scale. Near 60 MPa the part gains $\partial m/\partial P = 0.0137\,\mathrm{g/MPa}$ of pack pressure (23.974 g at 50 MPa, 24.247 g at 70 MPa), so about 18 MPa moves it by 1 %. Inside its limits, with $e_k = m_{m,k} - m_{SP}$ the weighed deviation, d_k the disturbance including the scale error, and z_k the mass change the correction has made, the loop is the recursion of Equation (7.13).

$$e_k = d_k + z_k, \qquad z_{k+1} = z_k - k\,e_k, \qquad k = G\,\frac{\partial m}{\partial P} \tag{7.13}$$

The error from a persistent disturbance shrinks by the factor $1 - k$ per part, so the correction is monotone for $k < 1$, alternates for $1 < k < 2$ and diverges beyond. The correction is the integral form of an EWMA controller whose weight divided by its model gain is G, and $0 < k < 2$ is the stability condition established for that controller [17]. In the notation of Section 1.4.5 the loop gain k is the product $\lambda\xi$ of Equation (1.8), and the scale reports each part before the next shot, the case without metrology delay. The default gain gives $k = 0.55$, and the bound $k = 2$ lies at 146 MPa/g, inside the 200 MPa/g reach of the gain slider. For a disturbance drawn afresh each shot with variance σ^2, z_k depends only on earlier draws, and the stationary variance of the error is $2\sigma^2/(2-k)$. An integral correction therefore widens the spread of an uncorrelated disturbance at any gain, by 8 % in standard deviation at $k = 0.27$ and by 17 % at $k = 0.55$.

Pause. Try Exercise 7.3 before reading on.

The scorecard counts the parts in tolerance after five start-up parts, the rejects by cause, the pressure excursions, the cycle time, the energy per good part, the length of the last part, the spread of the weighed masses and the correction in use. The challenge is met when at least 29 of parts 6 to 35 are within ± 1 % of 24.10 g with no flash and no shot at the pressure limit. It is judged on the case as given, against the fixed 24.10 g specification, so changing the mass target, a model parameter or a disturbance setting voids the attempt.

7.5 Results

The application ships nine scenarios, each with a briefing and a completion rule (Section 1.6.4): 1, Nominal recipe (one cycle and its handover); 2, A thicker resin lot (a change the speed loop hides); 3, Recycled feed (noise that a correction amplifies); 4, Early switchover (a front that freezes); 5, Late

Table 7.5 Results Measured on the Engine

Case and condition	Result
1, switchover	96.0 % full at 0.860 s, valve 54.1 %
1, handover step	≤0.18 % in 4 shots, ≤0.44 % in 20
1, peaks	melt 87.4 MPa, mean cavity 47.5 MPa
1, part	24.107 g, 99.545 mm, 10.86 s, 32.8 kJ
1, hold 1 / 1.5 / ≥2.5 s	24.063 / 24.093 / 24.107 g
2, new lot, automatic	valve 84.7 %, 147.9 MPa, 23.755–23.769 g
2, manual, valve 40 %	short, 65.0 % at the time limit, 15.6 g
2, correction to 15 / 30 MPa	near 23.96 g / within 0.03 g
3, off / 40 / 120 MPa/g	weighed spread 0.065 / 0.073 / 0.080 g
4, switch at 30 mm	at 50.5 %, short at 98.1 %, 23.412 g
4, 20 mm / 80 MPa at 30 mm	24.038 g / 24.383 g heavy
5, switch at 11.5 mm	103.6 %, 94.6 MPa, flash, 24.560 g
5, cavity pressure 15 MPa	100.4 %, 56.7 MPa, 24.150 g
5, pressure loop 0.1, 300 ms	77.1 MPa, 24.574 g heavy
6, coolant 10 %, part 10 / 20	99.418 / 99.330 mm, 24.097–24.107 g
6, cooling 6.5 s from part 10	99.490–99.556 mm, 13.36 s, 35.7 kJ
7, bias 22 MPa	reads 60.0, true 38.0 MPa, 23.821 g
7, correction to 15 / 25 MPa	0.087 g light / within 0.03 g
8, cold mold, part 1 / 7	99.976 / 99.647 mm, 24.125 / 24.113 g
8, cooling time 3 s	four oversize parts instead of six

switchover (a screw that rams a full cavity); 6, Cooling restriction (right mass, wrong length); 7, Pressure sensor bias (an outer loop on the product); 8, Cold mold (start-up parts); and the challenge, a 30-shot production campaign that tests the authority of a correction. Cases 1 to 3 and 6 to 8 are experiments, cases 4 and 5 demonstrations. Their outcomes are listed in Table 7.5. All numbers here come from runs of the application's engine at each scenario's own seed, with any setting named for a case changed before the run starts. Part masses are true masses unless called weighed, and a value quoted for a shot that repeats is taken from the third shot.

7.5.1 One Cycle and Its Handover

From the second shot of the nominal recipe on, the switchover comes 0.860 s into the shot with the cavity 96.0 % full and the valve at 54.1 %. The melt ahead of the screw peaks at 87.4 MPa and the mean cavity pressure at 47.5 MPa, well under the 80 MPa flash limit. The first output of the pressure loop differs from the last output of the speed loop by at most 0.18 % in the first four shots and 0.44 % in twenty, one proportional increment on the change in the transducer reading over 2 ms. A pressure loop resuming from its own last output, the 29.8 % that held 60 MPa at the end of the previous hold, would instead move the valve by 24.3 % in one period. After the handover the melt pressure settles within ±2 MPa of its setpoint in 0.27 s. A gate-freeze study run on the simulator traces the curve a molder draws to set the hold time. Holding the pressure for 1 s makes a 24.063 g part and 1.5 s a 24.093 g part, while 2.5 s or more makes the 24.107 g part, ejected at 91.4 °C and 99.545 mm long, every 10.86 s for 32.8 kJ.

> ✦ DID YOU KNOW? **Holding Longer Adds Nothing**
>
> From the simulator: in the nominal recipe a hold of 1 s makes a 24.063 g part and a hold of 1.5 s a 24.093 g part, but every hold of 2.5 s or more makes the same 24.107 g part (Table 7.5). Once the gate has frozen, pressure on the screw no longer reaches the part.

7.5.2 The Resin Changes

In case 2 a lot 75 % more viscous reaches the hopper before the fourth shot. To hold 40 mm/s the speed loop opens the valve from 54.1 to 84.7 %, close to its 90 % limit, and the melt ahead of the screw peaks at 147.9 MPa instead of 87.4 MPa. The fill takes 0.874 to 0.876 s instead of 0.860 s and

looks the same. The part does not. It weighs 23.755 to 23.769 g, 1.37 to 1.43 % below the target, because the thicker melt loses more pressure on its way to the cavity during the pack. The cavity sensor peaks at 20.5 to 22.0 MPa instead of 41.4 MPa, and the gate freezes at 3.09 s instead of 3.42 s. The fill at the switchover also falls from 96.0 to 93.5 or 93.7 %, because the melt ahead of the screw is compressed further at the same screw position. With the valve fixed at 40 % in manual mode from the start, the same melt has filled only 65.0 % of the cavity when the 3 s fill-time limit forces the switchover, and every part of the new lot is short, 15.6 g on the fourth shot. Switched to manual after the first fill instead, the valve holds the 53.9 % the speed loop last applied, and the melt stops at 93.9 % of the cavity, a 22.3 g short shot. The speed loop hides the lot from the fill but not from the scale.

Pause. Try Exercise 7.1 before reading on.

The correction at its defaults raises the pack setpoint 4 MPa per part, brings the parts back inside the tolerance from the sixth, and stops at its 15 MPa limit with the parts near 23.96 g. With a 30 MPa limit it settles near 25.4 MPa, and from the tenth part the weighed masses stay within 0.03 g of the target.

TRY IT IN THE APP 2 · A Thicker Resin Lot

Load the scenario and press Run. Watch the injection valve trend as the new lot reaches the fourth shot: the speed loop opens the valve from 54.1 to 84.7 %, and the fill takes 0.874 to 0.876 s, hardly longer than before. Then watch the finished parts: from the fourth shot every part comes out about 1.4 % light.

Recycled feed is the opposite disturbance. In case 3 the regrind variation of 15 % spreads the lot factor from 0.73 to 1.40 over 24 parts and the peak melt pressure from 71.4 to 119.7 MPa, and the weighed masses spread by 0.065 g (1σ), all inside the tolerance. The correction at its default gain widens the spread to 0.073 g, and at 120 MPa/g with a 10 MPa step to 0.080 g, an ordering that holds at seven other seeds as well. Over 195 parts with the step and total limits widened to 20 and 50 MPa, the spread without correction is 0.0655 g, and it grows by 2, 6, 15 and 43 % at gains of 10, 20, 40 and 80 MPa/g, against 4, 8, 17 and 49 % from the recursion of Equation (7.13). A correction that reacts to a disturbance already gone can only add variance, and the more gain the more it adds.

7.5.3 Where to Switch Over

Moving the switchover to 30 mm hands the valve to the pressure loop at 50.5 % fill, 0.464 s into the shot, with a pack setpoint of 60 MPa below the 87.4 MPa the speed loop needs to finish the fill at speed. Under pressure control the front slows as the filled length grows, the skin gains on it, and the cavity reaches only 98.1 %. The part is short at 23.412 g, and the mean cavity pressure never passes 21.7 MPa. At 20 mm the switchover comes at 79.6 % and the part is good at 24.038 g. An 80 MPa pack at 30 mm fills the cavity instead but packs the part 1.2 % heavy (24.383 g), because the setpoint that finishes the fill also packs the part.

Moving it to 11.5 mm keeps the speed loop pushing at 40 mm/s, or 28 cm^3/s, into a cavity that is already full. The cavity is 103.6 % full at the switchover, the melt is compressed within milliseconds, and the mean cavity pressure peaks at 94.6 MPa, above the 80 MPa the clamp can hold. Every part flashes and is 1.91 % heavy (24.560 g). Switching over when the cavity sensor reads 15 MPa ends the fill at 100.4 % with a 56.7 MPa peak and a part in tolerance (24.150 g), which is the melt itself reporting that the cavity is full. A sluggish pressure loop does similar damage at the nominal 14.5 mm. With 0.1 % per MPa and 300 ms the melt pressure overshoots to 100.8 MPa after the handover, the mean cavity pressure peaks at 77.1 MPa, and the part is 1.97 % heavy (24.574 g). The switchover is correct only as a pair, a moment and a loop that takes the valve cleanly at that moment.



Figure 7.3 The injection molding application at the end of the challenge campaign, run with the suggested correction of Table 7.6 (40 MPa per g, 5 MPa per part, 40 MPa in total). The figure is a composite of one screen: the badges, the plant view and its panels above the trends. (1) Shot 35 has just been ejected at a weighed 24.15 g, and the badge counts 29 good parts of 35 because the first six, from the cold mold, are oversize. (2) The screw has recovered to 48.2 mm, and the valve holds 3 % against the back pressure. (3) The scope draws this shot's cavity pressure over the last one, both under the 80 MPa flash line. (4) The histogram counts 30 of 30 campaign parts in tolerance, and parts 28 to 35 are all OK. (5) The trends over the last 2.0 min show eleven shots of screw speed, valve and melt pressure, with the drifting transducer reading 21.9 MPa against a true 6.0 MPa, the weighed mass stepping about the 24.10 g target, the pack setpoint moved by the correction to the 90.6 MPa in force for the last shot, and the mold at 68 °C. The header and the scorecard of the full screen, cut here, show $t = 6.30$ min and report the challenge target met, 30 of 30 in tolerance, no short, flashed or light part, no pressure excursion and a 10.87 s cycle.

> **TRY IT IN THE APP** **5 · Late Switchover**
>
> Load the scenario, press Run and watch the cavity pressure trace against the flash line and the label of each finished part: every part flashes and is 1.91 % heavy. Set Switch to pack on to Cavity P and watch the next part. The fill now ends at 100.4 %, and the part, at 24.150 g, is in tolerance.

7.5.4 Mass and Length Part Ways

In case 6 the coolant flow falls to 10 % at 12 s. The mold warms by about 1.1 K per shot, from 69.0 °C at the second part to 78.0 °C at the tenth, and the part mass stays between 24.097 and 24.107 g. The tenth part leaves the mold at 99.7 °C, goes on shrinking free of it, and measures 99.418 mm, the first out of tolerance. By the twentieth part the mold is at 84.6 °C and the part 99.330 mm long. Raising

Table 7.6 Challenge Campaign, Parts 6 to 35. Gain G in MPa/g, Limits and Correction in MPa, Worst Weighed Deviation From the Target

G, Δ_{shot}, Δ_{total}	In tolerance	Worst (%)	Peak c
fixed recipe	9 of 30	2.29	0
40, 4, 15 (defaults)	26 of 30	1.37	15.0
40, 5, 20	29 of 30	1.09	20.0
40, 5, 25	30 of 30	0.81	25.0
40, 5, 40 (suggested)	30 of 30	0.69	34.2
20, 5, 40	30 of 30	0.79	32.3
80, 5, 40	30 of 30	0.74	36.3
120, 5, 40	30 of 30	0.79	34.7

the cooling time to 6.5 s after the tenth part ejects the parts at 90.4 to 94.8 °C and brings them back to 99.490 to 99.556 mm, for a 13.36 s cycle and 35.7 kJ per part instead of 10.86 s and 32.8 kJ.

Case 8 starts from a mold at 30 °C, which reaches 37.0 °C by the first ejection and 60.7 °C by the seventh. The masses are right, 24.125 g on the first part and 24.113 g on the seventh, within 0.08 % of the warm-mold part, because a colder mold both densifies the melt and freezes the gate sooner. The length is not. The first part, ejected at 63.1 °C, did most of its shrinking held in the mold and measures 99.976 mm, the first six parts are oversize, and the seventh, at 99.647 mm, is the first in tolerance. A 3 s cooling time ejects the parts warmer and cuts the oversize run to four. The correction stays within 1.7 MPa of zero in case 8 and within 1 MPa in case 6, because its only input, the mass, was never wrong. A scheme that monitors part weight alone passes all of these parts.

7.5.5 A Biased Transducer

In case 7 the melt pressure transducer starts to read 22 MPa high at 26 s. In the middle of the fifth shot's hold it reads 60.0 MPa while the melt is at 38.0 MPa, and the cavity sensor, a separate instrument, peaks at 29.5 MPa instead of 41.4 MPa. The parts fall from 24.107 to 23.821 g, 1.16 % below the target, while the pressure loop holds its reading perfectly. The correction never reads the transducer, but it reads the scale. At its defaults it raises the setpoint 4 MPa per part, restores the tolerance from the fifth part, and stops at its 15 MPa limit with the parts 0.087 g light. With a 25 MPa limit it settles at 21.7 MPa and holds the weighed parts within 0.03 g of the target from the ninth. An outer loop on an independent measurement corrects the biased sensor of an inner loop, within the authority it is given, and the size of the correction becomes the diagnosis. Past the stability bound the same case shows the step limit at work. At 200 MPa/g with a 20 MPa step and a 50 MPa total the correction alternates between 10.6 and 30.6 MPa from part to part and the weighed mass between 23.95 and 24.22 g, while the default 4 MPa step confines the swing to one step, 19.2 to 23.2 MPa at first, and the weighed parts to within 0.05 g of the target.

7.5.6 The Production Campaign

The challenge starts from a 30 °C mold with 8 % of regrind variation. From the twelfth shot the lot setting rises in five steps from 1.1 to 1.5, and the transducer drifts 2.5 MPa per minute, 15.75 MPa by the end of the 378 s campaign. Both disturbances make the parts lighter, and neither is visible to the speed and pressure loops. The outcomes are listed in Table 7.6 and drawn in Figure 7.4.

The fixed recipe keeps 9 of the 30 campaign parts in tolerance, 20 short of the target. The first light part is shot 15, the worst weighed part is 2.29 % light, and the last weighs 23.761 g. The correction at its defaults keeps 26, because it pins at its 15 MPa limit and the last parts drift out. With 5 MPa per part and a 40 MPa total it keeps all 30, the worst 0.69 % from the target, with no flash and no shot at the pressure limit. The same settings carried through a restart give the same result. The correction peaks at 34.2 MPa after the 33rd part and is 30.6 MPa while the last part is molded, the 90.6 MPa setpoint shown in Figure 7.3. That part weighs 24.15 g, and the run stops as it is weighed, before the

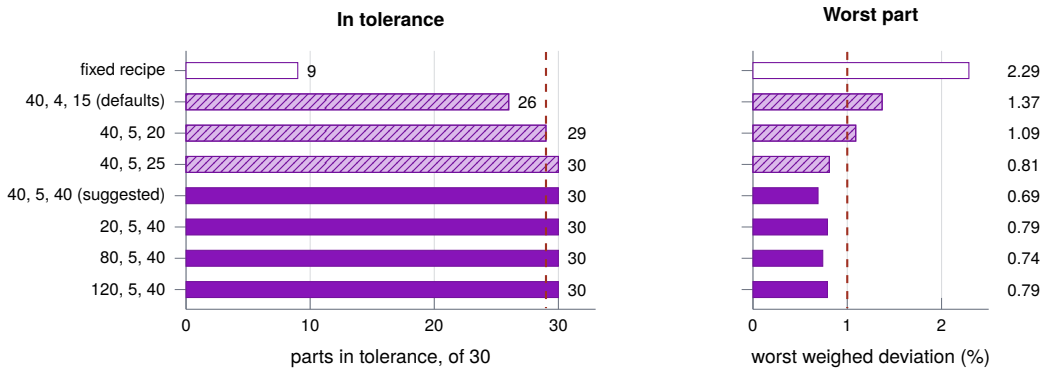


Figure 7.4 The challenge campaign of Table 7.6, labeled by gain G in MPa/g, step limit and total limit in MPa. The dashed lines are the target of 29 parts and the 1 % tolerance. Every setting allowed 40 MPa in total (solid bars) meets the target, while the default 15 MPa total (hatched) keeps 26 parts.

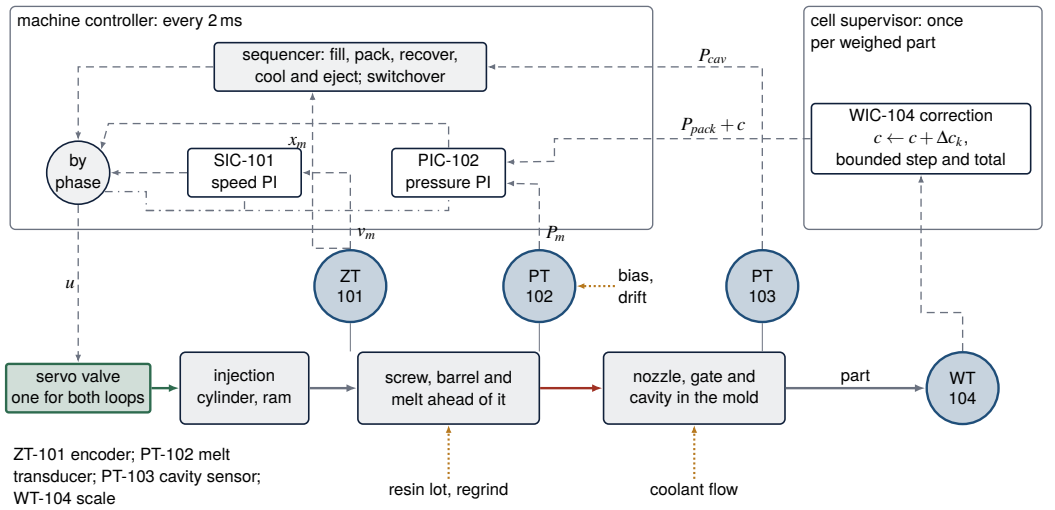


Figure 7.5 The chapter's loops on plant hardware. One servo valve drives the ram, and the sequencer hands its command u by phase to the speed loop SIC-101 on the encoder speed or to the pressure loop PIC-102 on the melt transducer, while the idle loop tracks the applied command (dash-dot). The switchover fires on the screw position, the cavity pressure or the time. After each part the scale WT-104 feeds the bounded correction that moves the pack setpoint. Tags are the book's.

correction acts on it. The mean cavity pressure peaks at 56.3 MPa and the valve at 84.4 % against its 90 % limit. A 20 MPa total already meets the target with 29 of 30 at the shipped seed, and at a 40 MPa total the worst part varies only from 0.69 to 0.79 % over gains of 20 to 120 MPa/g, smallest at the default, where the correction follows the drift without chasing much of the shot-to-shot variation. The cold mold still makes the first six parts oversize whatever the correction does, which is why the target scores mass and leaves the start-up parts out. Over five other seeds the fixed recipe keeps 9 to 11 parts, the default correction 26 to 29, the 20 MPa total 28 to 30 and the suggested settings 30, the only setting that meets the target at all six.

7.6 Technology in Practice

What Measures It

The loops of this chapter read three instruments inside the shot and one after it: an encoder on the screw, a transducer in the melt ahead of the screw, a pressure sensor in the cavity and a scale. A doctoral study of cavity pressure control weighed the candidates for the controlled variable. The hydraulic pressure is too far from the cavity to see the disturbances in between, the screw speed drops to about zero in packing, and the melt flow rate has no practical transducer, so it chose the cavity pressure, while noting that it is highly nonlinear and time varying [16]. In the model the encoder rounds to 1 µm, the cavity sensor adds 0.1 MPa of noise and the scale 0.01 g, and none of the four sees the gate freeze that decides the part.

> **TECHNOLOGY** **Sensor: Pressure Transducers From Pump to Cavity**
>
> Most electronic pressure sensors bond a strain gauge to a diaphragm. Pressure bows the diaphragm, the gauge's resistance changes, and a bridge circuit turns the change into a voltage. A metal gauge pressed past its elastic limit loses accuracy and can give the false impression that it still works [19]. One instrumented research machine carried transducers on the pump, ahead of the servo valve, on the injection cylinder, in the nozzle and in the cavity near the gate, each calibrated on a dead weight tester before installation [16]. A sensor in the cavity has to be built into the mold, which a review counts among the destructive ways of monitoring [20]. The model's melt transducer, with its 5 ms lag, 0.2 MPa of noise and an optional bias, stands for the nozzle sensor.

What Moves It

On a hydraulic machine the servo valve is one element of a circuit with a pump, relief valves and solenoid valves that divert the oil in the direction each phase of the cycle needs [16]. In this chapter the sequencer does the diverting in software: the valve command goes to the speed loop, the pressure loop or the back-pressure opening by phase, and the valve itself never changes. Figure 7.5 draws the loops on that hardware.

> **TECHNOLOGY** **Final Element: A Servo Valve on the Injection Cylinder**
>
> One research machine was refitted with a hydraulic servo valve for its injection cylinder, and with a second one for packing, because one valve did not give enough room to maneuver. The openings were proportional to the supplied current, and only 0.5 to 99.5 percent of the range was used, to keep clear of the dead zone. A servo valve is a complex, high-order, nonlinear device, but below about 100 cycles per second a first-order lag describes it adequately [16]. Its spool is machined to tight tolerances, so even small concentrations of particles in the oil can drastically shorten its life [19]. The chapter's valve is that first-order lag, 20 ms, shared by both loops.

Where the Controller Runs

The research machine of the thesis was sequenced by timers, limit switches and relays, and a microcomputer added data acquisition, machine sequencing and the control of each phase. It sampled fast signals such as pressures and slow ones such as barrel temperatures on separate boards, and the models of ram speed and nozzle pressure it reviewed were identified at a sampling period of 10 ms [16]. The app's machine controller runs every 2 ms. Control loops in industry run with cycle times from milliseconds to minutes, and in discrete part manufacturing an intermediate supervisor can oversee each cell with its machine controller [21]. Read as plant hardware, the speed and pressure loops and the phase sequencer belong to the machine controller, while the once-per-part correction from the scale fits the cell's supervisory layer.

IN PRACTICE **Where Practice Puts the Switchover**

Traditionally the phases of a shot were switched open loop, by limit switches, relays and timers, which leaves a large variability in part weight and dimensions. Switching on ram position is unreliable when a cushion of melt is left ahead of the screw, because the position at the end of filling then varies from shot to shot, and most process engineers avoid running without a cushion, because it risks short shots. One working rule keeps the ram under position or speed control until the cavity is about 90 to 98 percent full and then switches. The rate of rise of the cavity pressure marks the end of filling well [16]. The nominal recipe of this chapter switches at 96.0 % fill.

What Is Modernizing It

The first development is the drive. A national energy study of plastics counts the switch from hydraulic to all-electric molding machines as a 74 % saving of the machine's energy, the largest percentage saving of the processes it studied [18]. The second is cooling, which takes 50 to 80 % of the cycle. Conformal channels that follow the shape of the cavity, some made by additive manufacturing, are reported to shorten the cooling phase by 15 to 50 % [20]. The third is the control of quality. The same review collects norm-optimal iterative learning control of the cavity pressure, model predictive control combined with a neural network, quality prediction from in-mold sensors by support vector regression, and monitoring that reads the machine's own hydraulic pressure and screw position or measures the cavity pressure by ultrasound instead of a sensor built into the mold [20].

7.7 Limitations

One cavity is fed straight from a hot nozzle, without the runners, multi-cavity balance and check-ring leakage that change the shot delivered from one cycle to the next.

The temperatures are lumped and the skin is a geometric construction, so the model reproduces the timing of gate freeze but not the thickness distribution that a conduction solution gives [7, 8].

The fill uses one speed setpoint. Velocity profiles over the stroke, profiled hold pressures and a switchover point corrected from shot to shot are not modeled, and the correction acts on the pack pressure alone.

The drive has no inertia or hydraulic resonance, dynamics that controllers for real electrohydraulic machines must address [13].

The length is a proxy with an illustrative tolerance, without warpage or anisotropic shrinkage. The scale weighs every part at once, while a sampled or delayed weight would add dead time to the correction and move its stability bound, as Equation (1.9) shows for the integral form. The vial filler of Chapter 9 meets that delay: its checkweigher samples one vial per head every few cycles and reports it seconds later.

OPEN CHALLENGE **Closing the Quality Loop**

A molding machine nests three loops: machine control of speed, pressure and temperature, process control of the pressure and temperature in the mold, and a set point loop fed back from part quality. The machine loop is the most developed and the quality loop the least, its development having started only in the decade before the review. Learning, predictive and data-driven controllers on these loops have given good results, yet they are not widely used, because they lack standardization [20]. Progress would be a common framework that closes the quality loop for any machine, mold and resin, as this chapter's scale closes it for one part and one measurement.

7.8 Conclusion

Injection molding hands one valve between two loops inside every shot and fixes the part at a moment neither loop observes, and this chapter made both properties measurable on the simulator. Tracking the applied valve keeps the handover to one proportional increment, at most 0.44 % over twenty shots. A resin lot 75 % more viscous leaves the fill looking the same and the part about 1.4 % light, and

in manual mode it short-shots at 65 %. A switchover at 50.5 % fill makes a short part, and one at 103.6 % flashes at 94.6 MPa. A warmer or colder mold moves the length out of tolerance while the mass stays on target.

The correction from the scale removes what persists, within its authority, and widens the spread of what does not, by the 17 % its default loop gain of 0.55 predicts against 15 % measured over 195 parts. In the campaign the fixed recipe keeps 9 of 30 parts, the default correction 26, and a correction allowed 40 MPa all 30.

The model is lumped, with one cavity, a linear PVT law, a length proxy and a scale without delay. A fuller model would add runners and cavities, velocity and hold profiles, a delayed weight and a corrected switchover point. Within those limits the application shows how a product that no loop measures is still held to specification.

KEY POINTS

- Two loops take turns on one valve in every shot, and the part is fixed at a moment neither of them observes.
- Tracking the applied valve keeps the handover to one proportional increment, at most 0.44 % over twenty shots.
- A resin lot 75 % more viscous leaves the fill looking the same and the part about 1.4 % light.
- The correction from the scale removes what persists, within its authority, and widens the spread of what does not.
- In the campaign the fixed recipe keeps 9 of 30 parts, the default correction 26, and a correction allowed 40 MPa all 30.

Where else this applies. The converter of Volume 16, Chapter 4 corrects its heat-loss factor between heats with the same exponentially weighted update, and the lithography line of Volume 14, Chapter 6 corrects each lot's recipe while its measurements wait in a queue.

References

[1] R. Hanus, M. Kinnaert, and J.-L. Henrotte, "Conditioning technique, a general anti-windup and bumpless transfer method," *Automatica*, vol. 23, no. 6, pp. 729–739, 1987.

[2] S. F. Graebe and A. L. B. Ahlén, "Dynamic transfer among alternative controllers and its relation to antiwindup controller design," *IEEE Transactions on Control Systems Technology*, vol. 4, no. 1, pp. 92–99, 1996.

[3] Y. Wang, F. Gao, and F. J. Doyle, "Survey on iterative learning control, repetitive control, and run-to-run control," *Journal of Process Control*, vol. 19, no. 10, pp. 1589–1600, 2009.

[4] Z. Chen and L.-S. Turng, "A review of current developments in process and quality control for injection molding," *Advances in Polymer Technology*, vol. 24, no. 3, pp. 165–182, 2005.

[5] D. O. Kazmer, S. Velusamy, S. Westerdale, S. Johnston, and R. X. Gao, "A comparison of seven filling to packing switchover methods for injection molding," *Polymer Engineering & Science*, vol. 50, no. 10, pp. 2031–2043, 2010.

[6] Y. Yang, X. Chen, N. Lu, and F. Gao, *Injection Molding Process Control, Monitoring, and Optimization*. Carl Hanser Verlag, 2016.

[7] M. R. Kamal and S. Kenig, "The injection molding of thermoplastics part I: Theoretical model," *Polymer Engineering & Science*, vol. 12, no. 4, pp. 294–301, 1972.

[8] M. Rafizadeh, W. I. Patterson, and M. R. Kamal, "Physically-Based Model of Thermoplastics Injection Molding for Control Applications," *International Polymer Processing*, vol. 11, no. 4, pp. 352–362, 1996.

[9] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science Advances*, vol. 3, no. 7, p. e1700782, 2017.

[10] A. R. Agrawal, I. O. Pandelidis, and M. Pecht, "Injection-molding process control—A review," *Polymer Engineering & Science*, vol. 27, no. 18, pp. 1345–1357, 1987.

[11] C. A. Hieber and S. F. Shen, "A finite-element/finite-difference simulation of the injection-molding filling process," *Journal of Non-Newtonian Fluid Mechanics*, vol. 7, no. 1, pp. 1–32, 1980.

[12] Y. Yang and F. Gao, "Cycle-to-cycle and within-cycle adaptive control of nozzle pressure during packing-holding for thermoplastic injection molding," *Polymer Engineering & Science*, vol. 39, no. 10, pp. 2042–2063, 1999.

[13] H. Havlicsek and A. Alleyne, "Nonlinear control of an electrohydraulic injection molding machine via iterative adaptive learning," *IEEE/ASME Transactions on Mechatronics*, vol. 4, no. 3, pp. 312–323, 1999.

[14] Y. Yang and F. Gao, "Adaptive control of the filling velocity of thermoplastics injection molding," *Control Engineering Practice*, vol. 8, no. 11, pp. 1285–1296, 2000.

[15] F. Gao, Y. Yang, and C. Shao, "Robust iterative learning control with applications to injection molding process," *Chemical Engineering Science*, vol. 56, no. 24, pp. 7025–7034, 2001.

[16] M. Rafizadeh, "Physically-based dynamic model for the control of cavity pressure in thermoplastics injection molding," Ph.D. dissertation, McGill University, Department of Chemical Engineering, Montreal, 1996.

[17] A. Ingolfsson and E. Sachs, "Stability and Sensitivity of an EWMA Controller," *Journal of Quality Technology*, vol. 25, no. 4, pp. 271–287, 1993.

[18] U.S. Department of Energy, Advanced Manufacturing Office, "Bandwidth study on energy use and potential energy savings opportunities in U.S. plastics and rubber manufacturing," U.S. Department of Energy, Washington, DC, Report DOE/EE-1573, Sep. 2017, oSTI 1512369.

[19] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[20] O. Ogorodnyk and K. Martinsen, "Monitoring and control for thermoplastics injection molding: A review," *Procedia CIRP*, vol. 67, pp. 380–385, 2018.

[21] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

7.9 Exercises

7.1 Concept check · 2 min

In scenario 2 a resin lot 75 % more viscous reaches the hopper before the fourth shot. In automatic mode the speed loop still fills at 40 mm/s, and the fill takes 0.874 to 0.876 s instead of 0.860 s. Every part of the new lot is nevertheless about 1.4 % light. Why?

(a) The speed loop runs into its valve limit, so less melt enters the cavity before the switchover

(b) The scale reads the thicker resin low, because a more viscous melt is less dense at room temperature

(c) The thicker melt loses more pressure on its way to the cavity during the pack, and the gate freezes sooner

(d) The pressure loop cannot reach its setpoint through the thicker melt, so the sequencer ends the hold early

7.2 Concept check · 1 min

At the switchover of the nominal recipe the pressure loop takes the valve from the speed loop. By how much does its first output differ from the last output of the speed loop?

(a) By $K_{c,p}$ times the gap between the pack setpoint and the melt pressure reading

(b) By the 24.3 % between its output at the end of the last hold and the valve now

(c) By nothing at all, because the clamp holds the emitted output at the previous command

(d) By its own proportional increment over one period, $-K_{c,p}(P_{m,k} - P_{m,k-1})$

7.3 Concept check · 2 min

In scenario 3 the lot factor of each shot is drawn afresh, with no correlation from shot to shot. What does the shot-to-shot correction at its default gain, loop gain $k = 0.55$, do to the spread of the weighed masses?

(a) It widens it by about 17 % in standard deviation, because each step answers a lot already gone
(b) It narrows it by the factor $1 - k$ per part, as it does for a disturbance that persists
(c) It leaves it unchanged, because the correction can move only the mean of the masses
(d) It narrows it at this gain, and it would widen it only for a loop gain above one

7.4 Theory and calculation · 6 min

In the nominal recipe a pack pressure of 50 MPa makes a 23.974 g part and one of 70 MPa a 24.247 g part. Use the recursion of Equation (7.13).

(a) Compute the sensitivity $\partial m/\partial P$ from the two parts and the loop gain k at the default gain $G = 40\,\text{MPa/g}$. *Checkpoint:* 0.013 65 g/MPa and $k = 0.546$.

(b) Find the gain G at which a persistent step is removed in one part, and the gain at which the correction stops converging. *Checkpoint:* 73 MPa/g and 146 MPa/g.

(c) For an uncorrelated disturbance, find the factor by which the correction widens the standard deviation at G = 20, 40 and 80 MPa/g, and compare with the increases of 6, 15 and 43 % measured over 195 parts in case 3. *Checkpoint:* 1.076, 1.173 and 1.484.

7.5 Theory and calculation · 6 min

In case 2 the parts of the new lot weigh 23.755 to 23.769 g without correction, 23.762 g on average. The correction runs at its defaults ($G = 40\,\text{MPa/g}$, $\Delta_{shot} = 4\,\text{MPa}$, $\Delta_{total} = 15\,\text{MPa}$, $n_{start} = 3$), and the fourth part is the first of the new lot. Take $\partial m/\partial P = 0.013\,65\,\text{g/MPa}$ and the lower tolerance $0.99 \times 24.10 = 23.859\,\text{g}$.

(a) Find the correction that would bring the new lot back to 24.10 g. *Checkpoint:* 24.8 MPa.

(b) Using Equation (7.12), list the correction in force for parts 5 to 8 and find the first part back in tolerance. *Checkpoint:* 4, 8, 12 and 15 MPa. The sixth part.

(c) Predict the mass at which the parts settle, and compare with the 25.4 MPa at which the correction settles when its total limit is 30 MPa. *Checkpoint:* 23.967 g.

7.6 Theory and calculation · 5 min

Use Equations (7.3) and (7.10) with the parameters of Tables 7.2 and 7.3. The length specification is $99.54 \pm 0.12\,\text{mm}$.

(a) Compute v_0 at 23 °C, the volume shrinkage of a 24.107 g part and its length when ejected at $T_{ej,0} = 91.3\,°\text{C}$. *Checkpoint:* $0.9503\,\text{cm}^3/\text{g}$, $s_V = 0.0454$, 99.546 mm.

(b) Find the highest and the lowest ejection temperatures at which that part stays in tolerance. *Checkpoint:* 99.7 °C and 83.7 °C.

(c) In case 8 the first part weighs 24.125 g and leaves the mold at 63.1 °C. Compute its length. *Checkpoint:* 99.976 mm.

7.7 App challenge · 9 min

Load scenario 4, Early switchover, which switches over at 30 mm. Use the default seed, make every change before pressing Run, and run each case to 35 s at 5× (three parts, 7 s of wall time). Read the label of the third part in the parts row and its true mass, column `mass_true_g` of the exported file at 35 s.

(a) Run the scenario as shipped, and read also the fill fraction in the short-shot note. *Checkpoint:* SHORT, 23.412 g, 98.1 %.

(b) Press the reload button, set Switchover position to 20 mm and run again. *Checkpoint:* OK, 24.038 g.

(c) Load scenario 5, Late switchover (11.5 mm), run it as shipped, and read also the largest `p_cav_mean_MPa` between 21 and 30 s. *Checkpoint:* FLASH, 24.560 g, 93.8 MPa.

(d) Press the reload button, set Switch to pack on to Cavity P and run again. *Checkpoint:* OK, 24.150 g, 56.6 MPa.

Accept masses within 0.003 g and pressures within 1 MPa.

7.8 App challenge · 8 min

Load scenario 2, A thicker resin lot. The new lot arrives at 26 s, and the fourth part, the first of the new lot, is weighed at about 41 s. Use the default seed.

(a) Run the scenario as shipped to 45 s. Read the label and the weighed mass of the fourth part. *Checkpoint:* LIGHT, 23.769 g.

(b) Press the reload button, switch to Manual before pressing Run, which holds the valve at its 40 % fill opening, and run to 45 s. Read the label and the mass of the fourth part and the fill fraction in the short-shot note. *Checkpoint:* SHORT, 15.62 g, 65.4 %.

(c) Press the reload button again, set Shot-to-shot correction to On and Largest total correction to 30 MPa before pressing Run, and run to 220 s at 20×. Read the first part back in tolerance, the Pack correction tile from the thirteenth part on, and the largest deviation of a weighed mass from 24.10 g from the tenth part on. *Checkpoint:* The sixth part, 25.1 to 25.7 MPa, 0.03 g.

7.9 App challenge · 8 min

Load the challenge, a 30-shot production campaign from a cold mold. The run stops itself when the 35th part is weighed, at 377.9 s, about 19 s of wall time at 20×. Use the default seed and make each change after pressing Reset and before pressing Run. For each run read the tiles Challenge target, Mass in ±1 % after 5 parts and Pack correction.

(a) Run the campaign as shipped, with the correction off. *Checkpoint:* Not met, 9 of 30, 0.0 MPa.

(b) Set Shot-to-shot correction to On at its defaults (40 MPa/g, 4 MPa per part, 15 MPa in total) and run again. *Checkpoint:* Not met, 26 of 30, 15.0 MPa.

(c) Press Reveal suggested tuning (40 MPa/g, 5 MPa per part, 40 MPa in total) and run again. *Checkpoint:* Met, 30 of 30, 30.6 MPa.

Explain from the disturbances why run (b) fails and run (c) passes.

7.10 Application area · 4 min

In case 6 the coolant flow of the mold falls to 10 %. The parts keep their mass, and a cooling time raised from 4.0 to 6.5 s restores their length at a cycle of 13.36 s instead of 10.86 s and 35.7 kJ per part instead of 32.8 kJ.

(a) Find the parts per hour at each cycle and the cost of the longer cooling in output and in energy per part. *Checkpoint:* 331 and 269 parts per hour, 19 % fewer, 8.8 % more energy.

(b) Name two measurements that would reveal the fault before the length gauge does, and the lasting fix.

7.11 Application area · 3 min

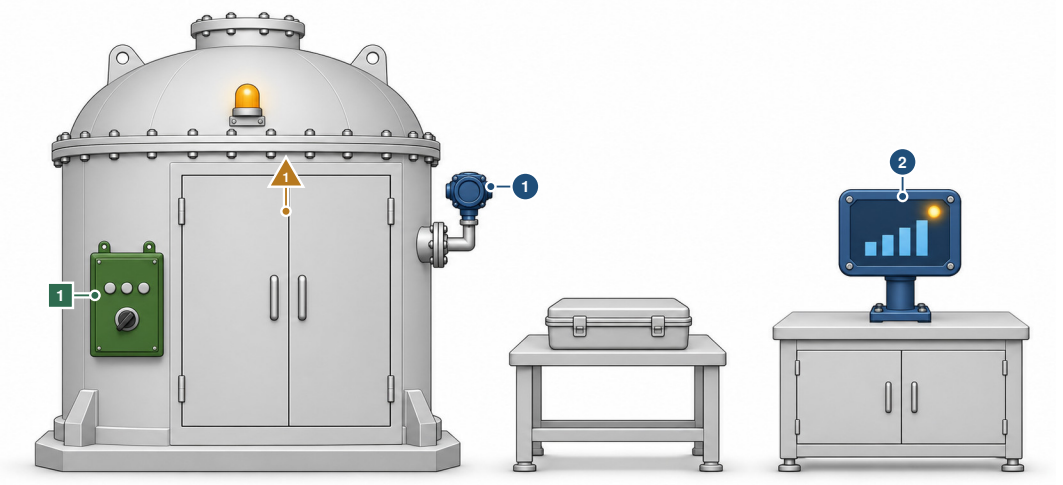
A plant moves its scale to the end of a conveyor, so that the weight of a part reaches the correction two shots later than in the app. The integral form then has a delay of $D = 2$ parts in Equation (1.9), which is stable for $k = \lambda\xi$ below 0.618. With $\partial m/\partial P = 0.0137\,\mathrm{g/MPa}$, find the largest stable gain G and judge the default of 40 MPa/g.

7.12 Review the agent · 5 min

An AI assistant is asked to make the correction of case 7, Pressure sensor bias, as fast as possible and replies as follows. "The loop gain is $k = G\,\partial m/\partial P$ with $\partial m/\partial P = 0.0137\,\mathrm{g/MPa}$, and the correction is stable for $0 < k < 2$. The fastest correction is the largest stable gain, so set $G = 145\,\mathrm{MPa/g}$ ($k = 1.99$) with the step limit at 20 MPa and the total at 50 MPa. The parts will be back on target within one shot of the fault." Find the flaw and say how you would check it.

CHAPTER 8

TRISO Coating Reactor Simulator for Inferential Control of an Unmeasured Product



The controller moves 1 furnace power

The controller reads 1 bed temperature thermocouple 2 layer thickness, measured on a sample a week later

What disturbs it 1 coating efficiency nobody can see

Notice that the coating grows behind closed doors, so the thickness is read only on a sample at a separate laboratory station after the batch. Illustration, not to scale.

IN PLAIN TERMS **Coating Fuel Grains Nobody Can Watch**

The main idea. Nuclear fuel grains, each about the size of a grain of sand, get four thin coats inside a very hot furnace. No instrument can see a coat grow, so a computer model adds up how much should have landed and ends each coat when the sum reaches its target.

The challenge. If less of the gas sticks than the model assumes, the model still reaches its target on time while the real coats come out thin. Every reading in the control room looks normal.

Why it matters. Weighing the whole coater as it gains mass catches the missing coating and saves a batch of fuel that would otherwise be rejected.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history and the first fact in Section 8.2, the second fact and the two try-its in Section 8.6, 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 8.6), Technology in Practice (Section 8.8), Limitations (Section 8.9) and the open challenge

The control question No instrument reads a coating thickness while a TRISO fuel batch is in the furnace, so each of its four layers is ended on an integrated model. Which disturbances can such an endpoint defend against, and what does one real measurement, a load cell, add?

Learning objectives

- Compute the particle count, bed depth and minimum spouting velocity of a charge, and explain why the carrier flow must be scheduled.
- Explain why a timer, an inferential endpoint and an endpoint corrected by the bed mass give the same batch on a nominal plant and different ones under a disturbance.
- Invert a measured mass gain through the particle geometry into a layer thickness.
- Identify which errors a gain correction cannot see, and the check that catches them.
- Relate the fluidization setpoint to the spread of the silicon carbide layer on which a batch is accepted.

Prerequisites Sections 1.4.2 (guards that end a phase), 1.4.4 (endpoint constraints) and 1.6.2 (the PID of the furnace loop), and Chapter 6 for a hold ended on an observer's estimate.

The app https://apmonitor.com/apps/triso-coating/

See also Volume 5, Chapter 10 for a soft sensor corrected by a delayed laboratory result, and Volume 5, Chapter 11 for a product temperature no instrument reaches. Volume 18, Chapter 11 for inferential measurements on another nuclear-materials process, a waste vitrification melter.

8.1 Introduction

A block diagram teaches a student that a controller holding its measurement is controlling the process. On a fluidized-bed coater growing nuclear fuel that statement is false, and the falsehood is expensive. The bed temperature holds its setpoint to a mean error of 2.7 °C through the four layers, all three flows hold theirs, the bed spouts, the integrated thickness estimate climbs a smooth curve to exactly the right number, every layer ends on schedule, and the batch is 28.4 µm thin. The simulator of this chapter reproduces that run, and it reproduces it beside the nominal run so that the two are visibly identical on every trend the operator has. What separates them is a coating efficiency of 0.72 rather than 1.00, and no instrument in the coater carries that number.

The misconception the simulator corrects is the substitution of the measured variable for the variable of interest. A coating thickness cannot be measured through the wall of a 1500 °C graphite furnace, so the controlled variable is computed from an integrated deposition model driven by a thermocouple, three mass flow controllers and an estimate of how much particle surface is in the pot. The application integrates both models, the plant's seven-class truth and the control system's single-particle estimate, and the controller sees only the second. A transfer function between precursor flow and thickness cannot pose the problem, because the error that matters is a parameter inside the estimator rather than a term in the plant.

The class of problem is closed-loop control on an unmeasured quality variable through a soft sensor, and it is general. Reviews of soft-sensor practice place the technique wherever the economically important variable is unmeasurable, expensive to measure, or measurable only after a delay [1, 2]. Semiconductor film thickness is measured between wafers rather than during a deposition, so the loop that sets the process runs run to run rather than within run [3], and polymer melt index is inferred from reactor conditions between laboratory samples hours apart and corrected by the sample when it comes [4]. Pharmaceutical tablet coating, distillation composition from tray temperatures and kiln free lime have the same shape, and Volume 5, Chapter 10 corrects a psychrometric soft sensor of powder moisture with a delayed laboratory result. Two features recur in all of them. Model error and estimator error are indistinguishable from process error until an independent measurement arrives, so the design question is never how the model is tuned but which of its parameters can move without any instrument noticing. And the batch is accepted on the variance rather than the mean, because a population of a hundred thousand particles is only as good as its thin tail, which is why the acceptance criterion here is a standard deviation that nothing on the operating screen displays.

This chapter sets out a browser-hosted TRISO coating reactor in the detail needed to rebuild it, and shows on that plant that an endpoint scheme defends against exactly the disturbances its estimator can see and against nothing else. The analysis rests on three measured results. A silicon carbide step run 50 °C cold gives 26.2 µm on a timer and 35.0 µm on the integrated estimate for 21.6 min of extra time. A coating efficiency of 0.72 leaves the same estimate 28.4 µm thin with all four layers rejected and nothing moving on any trend, and a load cell closes that gap to 0.79 µm for 28 extra minutes. And the silicon carbide standard deviation rises from 0.61 to 3.53 µm as the spouting ratio falls from 2.20 to 1.15 with every mean thickness within 0.15 µm of target. The implementation is a 34-state plant carrying a seven-class particle ensemble, a ten-phase recipe sequencer, an instrument chain, a cruder estimator model, four endpoint architectures and fourteen scripted scenarios, the last of which withholds the one parameter the reader most needs. The deposition kinetics, the Mathur-Gishler correlation and the correction of a fast model by a slow measurement are all established methods.

The chapter is organized as follows. Section 8.2 recalls coated particle fuel and why its coaters became a control problem. Section 8.3 sets down the states, the constants and the instruments in enough detail for the plant to be reproduced, and closes with the assumptions. Section 8.4 gives the thickness estimator the control system runs and its correction by the load cell. Section 8.5 describes the temperature loop, the fluidization schedule, the recipe sequencer with its four endpoint architectures and the acceptance criteria. Section 8.6 reports the measured batches: the nominal run, the temperature loop, the fluidization window, the central result on coating efficiency, what the correction cannot see, variability, errors upstream of the instruments and the graded challenge. Section 8.8 describes the measurements, gas trains and control platforms of real coaters, Sections 8.7 and 8.9 discuss the structure of the result and state the limitations, Section 8.10 summarizes, and the exercises follow in Section 8.11.

8.2 Coated Particle Fuel and Its Coaters

Coated particle fuel began as a way to keep fission products inside a fuel element that had no cladding. The Dragon project, a cooperative European high-temperature reactor program operated at Winfrith, was built to test fuel for a helium-cooled reactor and had to solve the containment problem inside the fuel particle itself [5]. The pyrolytic-carbon-coated particle came out of that work, and the four-layer arrangement that followed placed a porous buffer against the kernel, a dense inner pyrocarbon over it, a silicon carbide shell over that and a dense outer pyrocarbon on the outside. The silicon carbide layer is the pressure boundary and the barrier to metallic fission products, and it is the layer that decides whether the fuel is any good [6].

Two reactor programs then ran the fuel at scale and produced the operating record the fabrication specifications were written from. The AVR pebble-bed reactor at Jülich raised its average hot-gas temperature to 950 °C and reported the fission-product release from coated-particle fuel under that duty [7], and Fort St. Vrain in Colorado ran prismatic coated-particle fuel as a commercial unit [6]. The German program that grew out of AVR converted to low-enriched uranium dioxide TRISO and, over two decades of irradiation and heating tests, established that particles made to its process released almost nothing below 1600 °C [8]. A direct comparison of the German and United States campaigns found the difference to be in fabrication rather than in the design of the particle, and identified the coating conditions, the kernel and the resulting layer properties as what separated fuel that performed from fuel that did not [9]. That result is why a coater is a control problem rather than a furnace with a timer on it.

The United States answered it with a qualification program rather than with a single recipe. The Advanced Gas Reactor fuel development and qualification campaign, run from Idaho National Laboratory with Oak Ridge National Laboratory, was laid out as a sequence of irradiations, AGR-1 through AGR-5/6/7, each preceded by a fabrication campaign whose product was characterized before it went into a capsule [10]. Its recipes, its coaters and the result of its first irradiation are set

out in the history box that follows. The layer thicknesses used here are the AGR-2 specification and the coater is a 75 mm laboratory unit of the size those campaigns ran development batches in.

> ⧖ HISTORY **2002 to 2018: Several Recipes Before One**
>
> A national program to qualify coated particle fuel for licensing began in 2002 [6]. Earlier campaigns in the United States had coated their particles under one nominal set of conditions, and the fuel often performed below what was required under irradiation. The first irradiation of the new program therefore carried a baseline recipe and three variants: a hotter inner pyrocarbon step, a richer inner pyrocarbon gas, and a silicon carbide step at a lower temperature with argon mixed into its hydrogen. Those particles came from a laboratory coater of 2 inches, and the later irradiations from a pilot-scale coater of 6 inches [11]. The first irradiation reached a peak burnup of 19.6 % fissions per initial metal atom with no failed particle detected among the roughly 300 000 irradiated [12].

Acceptance moved from the mean to the distribution over the same period, and for a mechanical reason. A batch is a hundred thousand independent pressure vessels, the internal gas pressure they hold is common to all of them, and the through-wall stress in each is set by its own silicon carbide thickness, so the probability that any particle fails is fixed by the thin tail of the thickness distribution rather than by its mean. Characterization for the AGR campaigns is correspondingly statistical, reporting a mean and a dispersion measured on a sample of particles from every coating batch [6], and a specification written that way is not defended by any loop that holds a mean. Fabrication has continued to be studied as a multiscale problem, coupling the bed hydrodynamics that decide how uniformly the particles see the gas to the chemistry that decides how fast the layer grows [13]. What has not changed since Dragon is that the growing layer cannot be seen. The batch is sectioned, polished and measured about a week after it leaves the coater, so the feedback loop that actually sets the process is a week long and everything inside the batch is open loop with respect to the property that matters. That is the situation inferential control exists for.

> ✦ DID YOU KNOW? **Ten Thousand Particles to a Pebble**
>
> The most common fuel pebble holds around ten thousand coated particles in a sphere 60 mm across. One fuel line built for a pebble-bed demonstration plant was designed for 300 000 spheres a year, and more than 500 000 had been made for its first core [6].

8.3 Coater Model

Coater CV-301 is a conical spouted bed of 75 mm diameter at the cylindrical section with an 8 mm gas inlet nozzle in the bottom of the cone, standing inside a graphite resistance furnace whose susceptor surrounds the vessel. The charge is 200 g of 500 µm uranium oxycarbide kernels, 279 865 particles presenting 0.220 m^2 of surface in a 7.16 mm bed. Carrier gas and precursor enter through the nozzle from three mass flow controllers, off-gas and soot leave through the freeboard and a filter, thermocouple TE-301 reads the bed, differential pressure cell PDT-301 reads across the bed, and load cell WT-301 carries the whole coater. Figure 8.1 shows the arrangement, the instruments and every symbol used in the equations that follow. The clock is in minutes, because a coating run is three hours long and the silicon carbide layer alone is an hour of it.

Table 8.1 classifies what the control system interacts with. The first entry is the one the application is about. A layer thickness is a controlled variable with no transmitter, so it appears only as the estimate $\hat{\delta}_L$, and the spouting ratio R is in the same position for a different reason. Every unmeasured disturbance listed is a quantity the estimator either assumes or reads through an instrument that does not see it. Thirty-four states are integrated, six for the plant and twenty-eight for the particle ensemble.

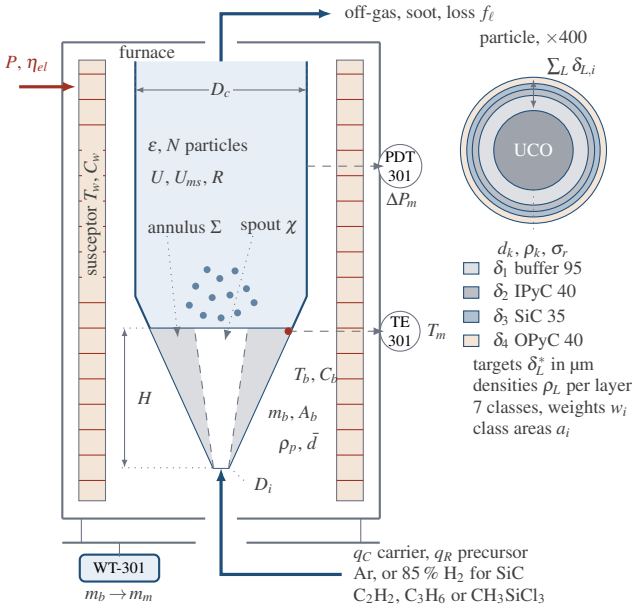


Figure 8.1 Conical spouted-bed coater CV-301 in cutaway inside its graphite furnace, with one particle magnified beside it. Carrier gas q_C and precursor q_R enter through the nozzle of diameter D_i and spout a bed of depth H in a column of diameter D_c, throwing a fountain of particles that rain back down the dense annulus. Superficial velocity U against the minimum spouting velocity U_{ms} sets the ratio R, which sets the circulation quality χ and the segregation strength Σ that shares the deposition flux unevenly among the seven size classes. Furnace power P heats the susceptor at T_w, which heats the bed at T_b. Only three instruments read the pot, thermocouple TE-301, differential pressure cell PDT-301 across the bed and load cell WT-301 under the whole coater, and none of them reports a thickness $\delta_{L,i}$.

Table 8.1 Classification of Variables. Nominal Values Are Those at the Start of the Buffer Layer, Where Every Scenario That Coats Begins.

Symbol	Role	Nominal	Range
$\hat{\delta}_L$	CV, inferred only	0	0–150 µm
$R = U/U_{ms}$	CV, inferred only	2.20	0.80–4.00 SP
T_b (TE-301)	CV, measured	1250 °C	1100–1650 °C SP
P	MV	2.04 kW	0–12 kW
q_C	MV	18.4 slpm	0–300 slpm
q_R	MV, recipe only	5.5 slpm	0–40 slpm
η_d coating efficiency	DV, unmeas.	1×	0.40–1.30×
ε_m charge error	DV, unmeas.	0 %	±20 %
γ_C, γ_R delivered	DV, unmeas.	1×	0.4–1.3, 0.5–1.4×
η_{el} element condition	DV, unmeas.	1×	0.35–1.10×
b_{TC} thermocouple bias	DV, unmeas.	0 °C	±80 °C
T_w, T_b	states	1275, 1250 °C	0–2600, 2400 °C
$\bar{q}_C$, $\bar{q}_R$	states	18.4, 5.5 slpm	0–300, 40 slpm
f_ℓ, a loss, agglomeration	states	0	0–0.80, 0–1.50
$\delta_{L,i}$, 4×7	states	0	m

8.3.1 The Particle Ensemble

The bed is represented by $N_c = 7$ size classes on a Gaussian grid, $z_i = -3, -2, \ldots, 3$, with weights and kernel diameters given by Equation (8.1).

$$w_i = \frac{\mathrm{e}^{-z_i^2/2}}{\sum_j \mathrm{e}^{-z_j^2/2}}, \qquad d_{k,i} = d_k\,(1 + z_i\,\sigma_r) \tag{8.1}$$

where d_k is the mean kernel diameter and σ_r the relative standard deviation of the kernel size distribution. Each class has a kernel mass $m_{k,i} = \rho_k \pi d_{k,i}^3/6$ and four coating thicknesses. Writing $r_{0,i} = d_{k,i}/2$ and $r_{L,i} = r_{L-1,i} + \delta_{L,i}$ for $L = 1 \ldots 4$, the class mass, diameter and surface area follow from Equation (8.2).

$$\begin{aligned} m_i &= m_{k,i} + \tfrac{4}{3}\pi \sum_{L=1}^{4} \rho_L \left(r_{L,i}^3 - r_{L-1,i}^3\right) \\ d_i &= 2 r_{4,i}, \qquad a_i = \pi d_i^2 \end{aligned} \tag{8.2}$$

where ρ_L is the density of layer L. The number of particles actually in the pot is set by the charge as it was weighed, by the weighing error and by the fraction carried over, as shown in Equation (8.3).

$$N = \frac{m_0\,(1 + \varepsilon_m/100)}{\bar{m}_k}\,(1 - f_\ell), \qquad \bar{m}_k = \sum_i w_i m_{k,i} \tag{8.3}$$

The bed mass, surface area, mean particle density and bed depth are the class averages multiplied by that count, as shown in Equation (8.4).

$$\begin{aligned} m_b &= N \sum_i w_i m_i, \qquad A_b = N \sum_i w_i a_i \\ \rho_p &= \frac{\sum_i w_i m_i}{\sum_i w_i \pi d_i^3/6}, \qquad H = \frac{m_b}{\rho_p (1 - \varepsilon) A_{col}} \end{aligned} \tag{8.4}$$

where ε is the loose-packed voidage and $A_{col} = \pi D_c^2/4$ is the cylindrical cross-section. A_b is the denominator of the whole deposition problem, and it runs from 0.220 to 0.745 m^2 through a batch.

8.3.2 Gas, Fluidization and the Spouting Ratio

The carrier is argon for the three carbon layers and 85 % hydrogen for silicon carbide, and its molar mass and molar heat capacity are mixed linearly in the hydrogen fraction f_{H_2}. The precursor is not a trace, because acetylene is nearly a quarter of the buffer gas and methyltrichlorosilane is 149 g/mol against hydrogen's 2, so it is carried explicitly in the mixture properties of Equation (8.5).

$$\begin{aligned} q_C &= \tilde{q}_C\,\gamma_C, \quad q_R = \tilde{q}_R\,\gamma_R, \quad q_T = q_C + q_R, \quad y_R = \frac{q_R}{q_T} \\ M_{mix} &= \frac{q_C M_c + q_R M_r}{q_T}, \qquad \rho_g = \frac{P_{atm} M_{mix}}{\mathscr{R}\,T}, \\ U &= \frac{q_T \times 10^{-3}\,(T/T_0)}{60\,A_{col}} \end{aligned} \tag{8.5}$$

where $\tilde{q}_C$ and $\tilde{q}_R$ are the flows the mass flow controllers indicate in standard liters per minute, γ_C and γ_R are the fractions actually delivered to the cone, M_c and M_r are the carrier and precursor molar masses, T is the bed temperature in kelvin, T_0 is 273.15 K, U is the superficial velocity in m/s and the factor T/T_0 converts standard liters to bed conditions. The minimum spouting velocity is the Mathur-Gishler correlation for a conical-cylindrical spouted bed [14], as shown in Equation (8.6).

$$U_{ms} = \frac{\bar{d}}{D_c}\left(\frac{D_i}{D_c}\right)^{1/3} \sqrt{\frac{2gH(\rho_p - \rho_g)}{\rho_g}}, \qquad R = \frac{U}{U_{ms}} \tag{8.6}$$

where $\bar{d} = \sum_i w_i d_i$ is the mean particle diameter, D_i the inlet nozzle diameter and R the spouting ratio. Equation (8.6) states the control problem in the hydrodynamics on its own. The requirement rises linearly with the particle diameter, with the square root of the bed depth and with the square root of the particle-to-gas density ratio, and the coating is moving all three. Below $R = 1$ the spout collapses, and above the carry-over ratio R_{carry} the fountain clears the freeboard.

The bed pressure drop is the one hydrodynamic measurement available, and Equation (8.7) shows why it is nearly useless for control.

$$\Delta P = \frac{f_s\, m_b\, g}{1000\, A_{col}} \underbrace{\min\!\left(R^{0.9}, 1\right)}_{\text{packed branch}} \underbrace{\left[1 + \Delta p_{pk}\, \mathrm{e}^{-((R-1)/0.20)^2}\right]}_{\text{onset peak}} \tag{8.7}$$

where f_s is the fraction of the bed weight the gas carries when the bed is spouting, Δp_{pk} is the fractional overshoot at the onset of spouting, and the result is in kPa. Above $R = 1$ the first bracket is unity and the second decays, so the reading settles at about three quarters of the bed weight per unit area and then barely moves with flow. It says whether the bed is spouting. It says nothing about how hard, which is why the fluidization loop of Section 8.5 runs on a computed ratio instead. Circulation quality is a normalized distance from the fully developed spout, as shown in Equation (8.8).

$$\chi = \operatorname{clamp}\!\left(\frac{R - 1}{R_{good} - 1}, 0, 1\right) \tag{8.8}$$

where R_{good} is the ratio at which circulation is fully developed.

8.3.3 Deposition Kinetics

Deposition is two limits in series. The kinetic limit is Arrhenius in bed temperature and a power law in precursor mole fraction, which is the form fitted to pyrocarbon deposition from light hydrocarbons [15] and to silicon carbide from methyltrichlorosilane [16]. The supply limit is the precursor arriving, multiplied by the fraction of it that lands on a particle, divided by the total particle surface area. Equation (8.9) gives both and combines them.

$$\begin{aligned} s &= \operatorname{clamp}\!\left(\frac{y_R - y_{soot}}{w_{soot}}, 0, 1\right), \qquad \eta = \eta_L\, \eta_d\, (1 - k_s s) \\ J_{sup} &= \frac{\eta\, Y_L\, q_R}{A_b} \\ J_{kin} &= k_L \exp\!\left[-\frac{E_{a,L}}{\mathscr{R}}\left(\frac{1}{T} - \frac{1}{T_{ref,L}}\right)\right]\left(\frac{y_R}{y_{ref,L}}\right)^{n_L} \\ J &= \frac{J_{kin} J_{sup}}{J_{kin} + J_{sup}}, \qquad \dot{m}_{dep} = J A_b \end{aligned} \tag{8.9}$$

where J is the bed-average deposition flux in kg/(m^2 min), s is the soot fraction, y_{soot} is the hydrocarbon mole fraction at which carbon begins nucleating in the gas phase, w_{soot} is the width of that transition, k_s is the efficiency lost to fully developed sooting, η_L is the coating efficiency the recipe assumes for layer L, η_d is the coating-efficiency disturbance, Y_L is the mass of deposit a standard liter of precursor could make if every molecule of it landed on a particle, k_L is the kinetic flux at the reference condition, $E_{a,L}$ is the activation energy, $T_{ref,L}$ and $y_{ref,L}$ are the reference temperature and mole fraction, and n_L is the reaction order. A harmonic mean is the standard way to write a process that is kinetically controlled when it is cold and feed-limited when it is hot, and it is why a temperature excursion changes the rate at all. η_d is the disturbance the estimator has no variable for, because Section 8.4 uses η_L alone.

The bed-average flux is shared out unevenly among the classes. A fully developed spout turns the bed over every few seconds and every particle sees the same average gas, while a slack bed leaves

the annulus as a dead zone in which the heavier particles, which sink, spend proportionally longer. Equation (8.10) gives the strength of that effect and the per-class flux factors.

$$\begin{aligned} \Sigma &= \sigma_{min} + (\sigma_{max} - \sigma_{min})(1-\chi)^{3/2} + \sigma_s\, s \\ f_i &= \max\!\left(0.05,\; 1 - \Sigma\,\frac{m_i - \sum_j w_j m_j}{\sum_j w_j m_j}\right) \cdot \frac{\sum_j w_j a_j}{\sum_j w_j a_j f_j^{0}} \end{aligned} \tag{8.10}$$

where Σ is the segregation strength, σ_{min} and σ_{max} are its values at full circulation and at minimum spouting, σ_s is the extra segregation contributed by a sooting bed, and f_i^0 is the unnormalized factor. The final quotient normalizes the factors so that the area-weighted mean is unity, which leaves the bed depositing JA_b in total and lets the classes disagree only about who receives it. The active layer then grows by Equation (8.11).

$$\frac{\mathrm{d}\delta_{L,i}}{\mathrm{d}t} = \frac{J\, f_i}{\rho_L}, \qquad \frac{\mathrm{d}\delta_{K,i}}{\mathrm{d}t} = 0 \;\text{ for } K \neq L \tag{8.11}$$

Equation (8.11) is where poor fluidization becomes a standard deviation rather than a slower mean. The bed mean and spread reported to the scorecard are the class averages of Equation (8.12).

$$\bar{\delta}_L = \sum_i w_i \delta_{L,i}, \qquad \sigma_L = \left[\sum_i w_i \left(\delta_{L,i} - \bar{\delta}_L\right)^2\right]^{1/2} \tag{8.12}$$

8.3.4 Furnace, Bed Energy and Actuators

Power enters a graphite susceptor, crosses to the bed by conduction and radiation lumped into one temperature-dependent conductance, and leaves as radiation from the susceptor and as sensible heat in the gas. Equations (8.13) and (8.14) are the two energy balances.

$$UA = UA_0 \left(\frac{T_w + T_0}{1773.15\,\mathrm{K}}\right)^3 \tag{8.13}$$

$$\begin{aligned} C_w \frac{\mathrm{d}T_w}{\mathrm{d}t} &= 60\,P\,\eta_{el} - UA\,(T_w - T_b) - Q_{rad} \\ Q_{rad} &= k_{rad}\,\phi\left[(T_w{+}T_0)^4 - (T_a{+}T_0)^4\right] \\ C_b \frac{\mathrm{d}T_b}{\mathrm{d}t} &= UA\,(T_w - T_b) - C_g\,(T_b - T_g) + \Delta H_L\,\dot{m}_{dep} \end{aligned} \tag{8.14}$$

where T_w is the susceptor temperature, C_w and C_b are the susceptor and bed heat capacities in kJ/K, P is the furnace power in kW, the factor 60 converts it to kJ/min, η_{el} is the element condition, Q_{rad} is the susceptor's radiative loss with coefficient k_{rad} and multiplier ϕ, T_a is room temperature, T_g is the inlet gas temperature and ΔH_L is the heat released per kilogram of deposit. The gas heat capacity flow is $C_g = (q_T/V_m)\,C_{p,mol}/1000$ in kJ/(K min) with V_m the standard molar volume. The cube in (8.13) carries the radiative part of the wall-to-bed transfer, and the fourth power in Q_{rad} is the loss that makes the loop gain a function of temperature. Both balances carry kJ/min on both sides.

Pause. Try Exercise 8.2 before reading on.

The two gas trains are first-order lags on the commanded flows, as shown in Equation (8.15), and the carry-over and agglomeration states are given by Equation (8.16).

$$\tau_{MFC}\frac{\mathrm{d}\tilde{q}_C}{\mathrm{d}t} = u_C - \tilde{q}_C, \qquad \tau_{MFC}\frac{\mathrm{d}\tilde{q}_R}{\mathrm{d}t} = u_R - \tilde{q}_R \tag{8.15}$$

$$\begin{aligned} \frac{\mathrm{d}f_\ell}{\mathrm{d}t} &= k_{el}\,\max(0,\, R - R_{carry})^{3/2}\,(1 - f_\ell) \\ \frac{\mathrm{d}a}{\mathrm{d}t} &= k_{ag}\,\max(0,\, 1-R)\begin{cases} 1 & \text{depositing} \\ 0.25 & \text{otherwise} \end{cases} \end{aligned} \tag{8.16}$$

Table 8.2 The Four-Layer Recipe. Targets Are the AGR-2 Specification. Every Entry in the Upper Block Is Adjustable From the Screen.

	Buffer	IPyC	SiC	OPyC
Deposit	porous PyC	dense PyC	SiC	dense PyC
Precursor	C_2H_2	C_2H_2/C_3H_6	CH_3SiCl_3	C_3H_6
Target δ_L^* (µm)	95	40	35	40
T_{sp} (°C)	1250	1300	1500	1300
q_R (slpm)	5.5	4.0	4.5	5.0
Hold time (min)	26.0	24.0	59.5	24.0
ρ_L (kg/m^3)	1050	1900	3200	1900
k_L (10^{-3} kg/(m^2 min))	13.00	6.300	3.050	5.100
$E_{a,L}$ (kJ/mol)	190	200	230	200
$T_{ref,L}$ (°C)	1250	1300	1500	1300
$y_{ref,L}$	0.240	0.0878	0.0698	0.0603
Y_L (10^{-3} kg/L)	1.0719	1.3399	1.7893	1.6078
η_L assumed; n_L	0.35; 1.0	0.50; 1.0	0.35; 0.8	0.50; 1.0
f_{H_2}; M_r (g/mol)	0; 26.0	0; 34.1	0.85; 149.5	0; 42.1

Particles that leave take their surface area out of Equation (8.3), so the ones that remain coat faster than the model expects. The agglomeration index declares the batch lost at $a = 1$.

8.3.5 Instruments

Three instruments and two flow indications are all the control system has. The thermocouple carries a bias, the load cell is filtered hard because the bed is bouncing on the gas, and both carry first-order-filtered Gaussian noise scaled so that the stationary standard deviation matches the specified value (0.8 °C and 0.5 g by default), as shown in Equation (8.17).

$$\begin{aligned} \tau_{TC}\frac{\mathrm{d}T_m}{\mathrm{d}t} &= T_b + b_{TC} + n_T(t) - T_m \\ \tau_{DP}\frac{\mathrm{d}\Delta P_m}{\mathrm{d}t} &= \Delta P - \Delta P_m \\ \tau_{cell}\frac{\mathrm{d}m_m}{\mathrm{d}t} &= m_b + n_m(t) - m_m \end{aligned} \tag{8.17}$$

where T_m, ΔP_m and m_m are the measured bed temperature, bed pressure drop and bed mass. Each filter is advanced at the control scan by the exact zero-order-hold update $x_k = \alpha x_{k-1} + (1-\alpha)u_k$ with $\alpha = \exp(-\Delta t/\tau)$, so no filter can go unstable at any sample period. The mass flow controllers report $\tilde{q}_C$ and $\tilde{q}_R$, which are the flows through the meters and not the flows into the cone.

🕮 **Pause.** Try Exercise 8.1 before reading on.

8.3.6 Parameters, Initialization and Numerics

The recipe is given in Table 8.2 and the remaining constants in Table 8.3. The pre-exponential factors k_L are fitted so that each layer takes the time a real run takes at the recipe's flow and temperature, while the activation energies, the reference temperatures and the stoichiometric yields are physical.

Every scenario opens at a genuine steady state. The thicknesses are set to the targets of any layer the chosen start phase has already passed, the bed temperature to the recipe value for the current phase, the precursor flow to the recipe value, and the carrier flow to the value that puts the bed at the spouting ratio setpoint. That flow has to be found iteratively, because the velocity required depends on the gas density, which depends on the composition, which depends on how much carrier the answer turns out to be. Six passes of Equation (8.6) put the opening ratio within a part in ten thousand of the setpoint, except at the start of the silicon carbide layer, where methyltrichlorosilane dominates the mixture mass and the residual is 0.07 %. The susceptor temperature is then set to the value that holds the bed, obtained by solving the steady form of Equations (8.13) and (8.14) in four passes, which

Table 8.3 Model Constants. Every Constant Except Room Temperature Is Adjustable in the Model Dialog. The Lower Block Sets the Segregation, Carry-Over, Agglomeration and Pressure-Drop Laws.

Symbol	Meaning	Value	Unit	Vis.
D_c, D_i	coater, nozzle diameter	0.075, 0.008	m	adj.
ε	loose-packed voidage	0.42	—	adj.
R_{good}, R_{carry}	circulation, carry-over	2.00, 3.00	—	adj.
m_0	charge as weighed	0.200	kg	adj.
d_k	mean kernel diameter	500	µm	adj.
σ_r	kernel size spread	0.024	—	adj.
ρ_k	UCO kernel density	10900	kg/m^3	adj.
y_{soot}, w_{soot}, k_s	soot onset, width, cost	.30, .18, .55	—	adj.
ΔH_{pyc}, ΔH_{sic}	heat of deposition	9450, 900	kJ/kg	adj.
C_w, C_b	susceptor, bed capacity	8.0, 0.40	kJ/K	adj.
UA_0	conductance, 1500 °C	1.50	kJ/K·min	adj.
k_{rad}	radiative loss	1.7004×10^{-11}	kJ/min·K^4	adj.
P_{max}	furnace rating	12.0	kW	adj.
T_g	inlet gas temperature	120	°C	adj.
τ_{TC}, τ_{MFC}, τ_{DP}	instrument lags	.05, .10, .15	min	adj.
τ_{cell}, τ_{gas}, τ_{obs}	cell, gas, trim	1.0, 1.2, 3.0	min	adj.
m_{gate}	correction noise gate	0.006	kg	adj.
t_{soak}, t_{cool}, t_{max}	soak, cool, abort	3.0, 20, 150	min	adj.
T_{band}	at-temperature band	8.0	°C	adj.
σ^*_{SiC}	SiC variability limit	2.0	µm	adj.
σ_{min}, σ_{max}, σ_s	segregation strengths	.30, 2.60, .90	—	adj.
k_{ag}, k_{el}	agglomeration, carry-over	0.80, 0.020	1/min	adj.
f_s, Δp_{pk}	bed weight, onset peak	0.72, 0.30	—	adj.
T_a	room temperature	30	°C	hidden

returns $T_w = 1275\,^{\circ}\mathrm{C}$ and $P = 2.04\,\mathrm{kW}$ at the buffer condition. Both initializations are checked by a startup test that evaluates every derivative at the opening state of scenario 2 and requires the largest relative value to be below 2×10^{-3}.

The integrator is classical fourth-order Runge-Kutta at a fixed 0.04 min (2.4 s) step, which is well inside the 6 s flow controller lag, the fastest dynamics it integrates. The control system declares and executes a scan every three physics steps, 0.12 min (7.2 s), which is close to what a batch coater's distributed control system runs. The instrument filters, including the 3 s thermocouple, are advanced at that scan, and the historian keeps five samples per simulated minute. Initialization does not advance the thickness estimator or its load-cell trim before plant time passes. Noise comes from the framework's seeded generator (Section 1.6.1), here with seed 20 260 906, so a noisy scenario replays identically. After every step the thicknesses are floored at zero, the flow states are clipped to their actuator spans, f_ℓ to 0.80 and a to 1.50.

8.3.7 Assumptions

The model is detailed where the control lesson lives and thin everywhere else, and each simplification fails somewhere.

Deposition is a lumped two-limit fit rather than a mechanism. There is no gas-phase chemistry, no intermediate species and no boundary-layer transport in Equation (8.9). The activation energies of 190 to 230 kJ/mol and the stoichiometric yields are physical, but the pre-exponential factors are fitted to reproduce layer durations, so the model cannot support a claim about a real machine's deposition rate.

Coating efficiency is one constant per layer times a disturbance. On a real coater it depends on temperature, gas velocity, wall condition and cone geometry and drifts within a run. Making it a constant is what allows the lesson to be posed cleanly, and it is also a simplification of exactly the quantity the lesson is about.

Layer anisotropy, density and crystallinity are absent. The optical anisotropy of the inner pyrocarbon matters as much as its thickness, because an anisotropic layer shrinks unevenly under irradiation and cracks [9]. Nothing here predicts whether a batch would survive a reactor.

Seven size classes on a Gaussian grid, not a population balance. There is no attrition, no breakage, no bimodal charge and no representation of the far tail of the distribution that a real qualification is written on [6]. The seven classes reproduce a mean, a spread and the direction in which poor fluidization moves them.

One correlation and a two-parameter circulation map for the hydrodynamics. Equations (8.6) to (8.10) carry no computational fluid dynamics, no explicit spout-annulus exchange and no cone-angle dependence, and the segregation coefficients are fitted rather than derived from a residence-time distribution [17]. The slump and carry-over flows are the right order of magnitude and the right trend, not design numbers.

The furnace is two capacities and a radiative loss. There is no gradient down the cone and no separate wall node in the deposition zone, so wall deposition, one of the real reasons coating efficiency is not what it should be, appears only inside the lumped efficiency of Equation (8.9).

The thermocouple reads the bed directly. A bias is offered as a disturbance, but on a real coater the well reads something between the gas and the graphite and a 40 °C discrepancy is ordinary, which through Equation (8.9) moves the kinetic rate by 30 to 47 %, depending on the layer and the sign.

Residual precursor swept out during a purge deposits nothing, although its molar mass is still carried in Equation (8.5) because dropping it at the changeover would halve the mixture density in one step. In reality a small amount of the last layer continues to grow through the purge.

The numerics are more repeatable than a control system. An identical trajectory on every machine is convenient and false.

8.4 The Thickness Estimator

Everything in this section is what the control system believes rather than what the plant is doing. The estimator is cruder than the plant in five separate ways. It carries one mean particle instead of seven classes, the charge as weighed rather than as charged, the nominal coating efficiency η_L rather than $\eta_L \eta_d$, no soot term at all, and the flows as indicated by the mass flow controllers rather than as delivered. Every one is a place the estimate can walk away from the truth.

8.4.1 One Mean Particle

From the four estimated thicknesses $\hat{\delta}_L$ the estimator builds a single particle and multiplies it by a particle count taken from the batch ticket, as shown in Equation (8.18).

$$\begin{aligned} \hat{n} &= \frac{m_0}{\bar{m}_k}, \qquad \hat{r}_0 = \frac{d_k}{2}, \qquad \hat{r}_L = \hat{r}_{L-1} + \hat{\delta}_L \\ \hat{m}_p &= \bar{m}_k + \tfrac{4}{3}\pi \sum_{L=1}^{4} \rho_L \left(\hat{r}_L^3 - \hat{r}_{L-1}^3 \right) \\ \hat{d} &= 2\hat{r}_4, \qquad \hat{A} = \hat{n}\pi\hat{d}^2, \qquad \hat{m}_b = \hat{n}\,\hat{m}_p \end{aligned} \tag{8.18}$$

The estimated particle density and bed depth follow from $\hat{m}_p$ and $\hat{d}$ as in Equation (8.4). Comparing (8.18) with (8.3) locates the errors precisely. The charge error ε_m and the carry-over fraction f_ℓ are both absent from $\hat{n}$, and the class spread is absent from $\hat{A}$.

8.4.2 The Integrated Estimate

The estimator runs its own copy of Equation (8.9) on measured and indicated quantities, as shown in Equation (8.19).

$$\begin{aligned} \hat{y}_R &= \frac{\tilde{q}_R}{\tilde{q}_C + \tilde{q}_R}, \qquad \hat{J}_{sup} = \frac{\eta_L Y_L \tilde{q}_R}{\hat{A}} \\ \hat{J}_{kin} &= k_L \exp\left[-\frac{E_{a,L}}{\mathscr{R}}\left(\frac{1}{T_m} - \frac{1}{T_{ref,L}}\right)\right]\left(\frac{\hat{y}_R}{y_{ref,L}}\right)^{n_L} \\ \frac{\mathrm{d}\hat{\delta}_L^{r}}{\mathrm{d}t} &= \frac{1}{\rho_L}\frac{\hat{J}_{kin}\hat{J}_{sup}}{\hat{J}_{kin} + \hat{J}_{sup}}, \qquad \hat{\delta}_L = \hat{\delta}_{L,0} + \kappa\left(\hat{\delta}_L^{r} - \hat{\delta}_{L,0}^{r}\right) \end{aligned} \tag{8.19}$$

where $\hat{\delta}_L^{r}$ is the uncorrected integral, κ is the load-cell correction factor of Section 8.4.3, and the zero subscripts are the values at which the current layer began. Both integrals are kept, because the correction has to be formed against the uncorrected one. Equation (8.19) contains the measured bed temperature T_m, so a run that is 50 °C cold extends itself, and it contains no η_d, so a run whose coating efficiency is 0.72 does not. The quantity the application exists to display is the difference between the two integrals, summed over the batch so that the divergence reads as one curve rather than four, as shown in Equation (8.20).

$$G = \sum_{L=1}^{4} \hat{\delta}_L - \sum_{L=1}^{4} \bar{\delta}_L \tag{8.20}$$

G is the inferential gap. No instrument on a real coater produces it, and it is on the screen only because the application is the plant as well as the control system.

8.4.3 The Bed-Mass Correction

The bed goes from 200 g to 383 g over a nominal run, and that gain is a direct measurement of how much material landed on the particles. It is the only measurement in the coater that bears on the controlled variable. The correction inverts the estimator's own mass model to find the thickness that accounts for the measured gain, as shown in Equation (8.21).

$$\hat{m}_b\left(\hat{\delta}_1, \ldots, \delta_L^{m}, \ldots, \hat{\delta}_4\right) = \hat{m}_{b,0} + (m_m - m_{m,0}) \tag{8.21}$$

where δ_L^{m} is the unknown, $\hat{m}_{b,0}$ and $m_{m,0}$ are the modeled and measured bed masses at the start of the layer, and the left side is Equation (8.18). It is monotone in δ_L^{m}, so twenty passes of bisection on $[0, 4\delta_L^{*}]$ are exact to better than a nanometer. The correction factor is the ratio of Equation (8.22).

$$\kappa \leftarrow \mathrm{clamp}\left(\frac{\delta_L^{m} - \hat{\delta}_{L,0}}{\tilde{\delta}_L^{r} - \hat{\delta}_{L,0}^{r}},\ 0.25,\ 3\right), \qquad \tau_{cell}\frac{\mathrm{d}\tilde{\delta}_L^{r}}{\mathrm{d}t} = \hat{\delta}_L^{r} - \tilde{\delta}_L^{r} \tag{8.22}$$

Three details in Equation (8.22) are deliberate and all three are general. The measured mass is inverted through the geometry rather than used as a mass ratio, because mass is not linear in thickness and a shell on a 250 µm kernel radius grows by a factor of 1.38 by the time the buffer is finished. The denominator is the *uncorrected* integral, because against the corrected one the loop would be chasing its own tail and would converge on the square root of the error instead of the error. And the model is put through the same 1 min filter as the load cell, which is what $\tilde{\delta}_L^{r}$ is, because comparing a filtered measurement against an unfiltered model reads short by whatever arrived during the filter's time constant.

Two gates protect the correction. It stays at unity until the modeled mass gain exceeds $m_{gate} = 6\,\mathrm{g}$, since a two-gram measurement over a two-gram model is not information, and on the first gated scan it snaps to the ratio and is filtered with τ_{obs} after that so it does not spend the first minutes of

every layer walking away from one. At every changeover the finished layer is re-based on the cell by solving Equation (8.21) once more with $\hat{\delta}_{L,0}$ substituted for the accumulated estimate. That reading is taken three load-cell time constants into the following purge rather than at the instant the precursor is shut off, with nothing depositing and the filter caught up, which is also when an operator would read it. Because the correction compares a measured gain against a modeled gain, it can only see errors that change the gain, and Section 8.6 shows what that excludes.

8.5 Control System

Three actuators carry three very different kinds of loop. Furnace power holds bed temperature in an ordinary PID on a thermocouple. Carrier flow holds a spouting ratio that nobody measures. Precursor flow has no loop at all, and is the recipe's number for whichever step the sequencer is in.

8.5.1 Bed Temperature

TIC-301 is the incremental, or velocity, form of PID of Section 1.6.2, with the derivative taken on the filtered measurement rather than on the error, executed every $\Delta t_c = 0.12\,\mathrm{min}$. Each scan forms increments, as shown in Equation (8.23), and accumulates them into the furnace command, as shown in Equation (8.24). In the symbols of Chapter 1, s_k is the running sum S_k of Equation (1.15), c_k the output before limiting of Equation (1.16) and I_k the integral of Equation (1.17).

$$\begin{aligned} e_k &= T_{sp,k} - T_{m,k}, & \Delta u_k^P &= K_c\,(e_k - e_{k-1}) \\ \Delta u_k^I &= \frac{K_c}{\tau_I}\,e_k\,\Delta t_c, & \Delta u_k^D &= -K_c\,\tau_D\left(\dot{T}_{m,k}^{f} - \dot{T}_{m,k-1}^{f}\right) \end{aligned} \tag{8.23}$$

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D, & c_k &= s_k + I_{k-1} \\ u_k &= \mathrm{clamp}(c_k,\,0,\,P_{max}) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{8.24}$$

where T_m^f is the measurement passed through a first-order filter of time constant 0.5 min and $\dot{T}_{m,k}^{f}$ is its backward difference over one scan, s_k accumulates the proportional and derivative increments without limit, c_k is the power the algorithm demands, u_k is the power applied, and the integral I used at scan k was accumulated through scan $k-1$. I is seeded at the steady power of the opening state, so every scenario starts without a bump. The loop is reverse acting, because the process gain is positive and a bed above setpoint calls for less power. The derivative filter is advanced by the same zero-order-hold update as the instruments, so it is unconditionally stable at the 7.2 s scan. Anti-windup is the back-calculation term of Equation (8.24), with tracking time $T_t = \tau_I$, so while the command is clipped each scan removes $\Delta t_c/\tau_I = 3\,\%$ of the excess from I alone. Manual mode tracks the operator's power as Section 1.6.2 describes, so the transfer is bumpless in both directions. The default settings are $K_c = 0.130\,\mathrm{kW/^\circ C}$, $\tau_I = 4.0\,\mathrm{min}$ and $\tau_D = 0.60\,\mathrm{min}$, and the output carries no rate limit because a resistance furnace has none.

The setpoint is ramped rather than stepped, at a rate that is itself a recipe parameter, as shown in Equation (8.25).

$$T_{sp,k+1} = T_{sp,k} + \mathrm{clamp}\left(T_{tar} - T_{sp,k},\; -\dot{T}\,\Delta t_c,\; \dot{T}\,\Delta t_c\right) \tag{8.25}$$

with $\dot{T} = 15\,^\circ\mathrm{C/min}$ by default. The target T_{tar} belongs to the phase and is re-read every scan rather than latched at the changeover, so moving a recipe temperature while a step is running does what the operator expects.

8.5.2 Fluidization as a Schedule With a Trim

The spouting ratio has no transmitter, so FIC-301 holds the ratio computed from the indicated flows and the estimated bed. Because the relationship between flow and ratio is algebraic and known, the

honest structure is a schedule carrying the flow and feedback closing only the residual, as shown in Equation (8.26).

$$\begin{aligned} \hat{R} &= \frac{\hat{U}}{\hat{U}_{ms}\left(\hat{d}, \hat{H}, \hat{\rho}_p, \hat{\rho}_g\right)} \\ q_{C,ff}:\ \hat{R}\left(q_{C,ff}\right) &= R_{sp} \quad \text{by six passes of (8.6)} \\ u_C &= \text{clamp}\left(q_{C,ff} + \Delta q_C,\ 0,\ 300\,\text{slpm}\right) \end{aligned} \tag{8.26}$$

where $\hat{U}$ and $\hat{U}_{ms}$ are evaluated from $\tilde{q}_C$, $\tilde{q}_R$, the measured temperature T_m and the estimated bed of Equation (8.18), and Δq_C is the trim. The trim is a PI controller of the same form as (8.24) with $\tau_D = 0$, output limits of ± 120 slpm, a slew limit of 120 slpm/min applied after the clamp, back-calculation against the slewed output, and default settings $K_c = 3.0$ slpm per unit of ratio and $\tau_I = 1.5$ min. The final clamp of Equation (8.26) lies outside the trim, so its integral would not see it, although no shipped scenario reaches either limit of that clamp. Every quantity in (8.26) comes from upstream of the cone, so the loop is exact when the delivered flow and the estimated bed are right and arithmetically correct but physically false when they are not.

8.5.3 The Recipe Sequencer

A coating recipe is not a setpoint, it is a state machine. Ten phases run in order, heat-up, buffer, purge, inner pyrocarbon, purge, silicon carbide, purge, outer pyrocarbon, cool-down and complete, and every transition changes the temperature target, the carrier composition and the precursor at once. The carrier changeover and the precursor ramp are both first-order with time constant $\tau_{gas} = 1.2$ min rather than steps, and that is not decoration. Replacing argon with hydrogen moves U_{ms} by a factor of more than two through the ρ_g term of (8.6), and methyltrichlorosilane at 6 to 8 % of the stream carries 57 to 64 % of the mixture's mass, so cutting it in one step would more than halve the gas density within one scan and drop the bed out from under a flow controller that had done nothing wrong.

A heat-up or a purge ends when the bed is within T_{band} of a target the ramp has already reached and the phase has run longer than its soak or purge time, and a cool-down ends on its allowance. What ends a *deposition* phase, the guard of Section 1.4.2, is the design decision, and four architectures are selectable.

- MANUAL. Nothing advances on its own. The operator steps the recipe from the panel, and in Manual mode also holds all three actuators.
- TIMER. The phase ends at $t_{phase} \geq$ the hold time, which is how every coater built before about 1990 decided when to stop.
- INFERENTIAL. The phase ends at $\hat{\delta}_L \geq \delta_L^*$, with $\kappa \equiv 1$.
- INFERENTIAL + MASS. The same condition with κ from Equation (8.22) and the end-of-layer re-basing active.

No phase may run forever. A phase that has not met its condition after $t_{max} = 150$ min is abandoned with a latched fault, because a step that cannot finish is a fault rather than a reason to wait. The mechanism does select. Derating the elements to $\eta_{el} = 0.35$ leaves the bed unable to reach 1500 °C, the purge ahead of the silicon carbide layer is abandoned at 150 min, the layer itself is abandoned 150 min later with 11.4 µm on the particles, and the batch finishes at 422 min instead of 191. The same choice among a clock, a measured event and an estimate ends primary drying in Chapter 5 and each column loading in Chapter 11.

8.5.4 Acceptance

The batch is graded once, after it comes out, on the six criteria of Equation (8.27).

$$\begin{aligned} &\left|\bar{\delta}_1 - 95\right| \le 8, \quad \left|\bar{\delta}_2 - 40\right| \le 3, \quad \left|\bar{\delta}_3 - 35\right| \le 2, \\ &\left|\bar{\delta}_4 - 40\right| \le 3, \quad \sigma_3 \le 2.0\,\mu\text{m}, \quad f_\ell \le 0.02, \quad a < 1 \end{aligned} \tag{8.27}$$

The fifth condition is the one that decides most real batches and the only one with no corresponding number anywhere on the operating screen.

8.6 Results

Fourteen scenarios are arranged as a laboratory sequence, with the briefings, completion rules and challenge of Section 1.6.4. Scenario 1 coats the buffer by hand and scenario 2 step-tests the temperature loop. Scenarios 3 and 4 find the slump and carry-over walls on a fresh and on a coated bed. Scenarios 5 and 6 run the whole recipe on the timer and on the inferential endpoint. Scenarios 7, 8 and 10 to 13 each load one fault: a cold silicon carbide step, a low coating efficiency, a mis-weighed charge, excess acetylene, a slack bed and a carrier leak. Scenario 9 repeats scenario 8 with the load-cell correction on. Scenario 14 is the graded challenge. Each fixes the starting phase, the endpoint architecture, the tuning, the disturbances and any timed events. The challenge is judged on the case as given, so changing a model parameter or a disturbance voids the attempt, as does an extra learner disturbance event when it fires. Every number below was produced by driving the shipped model headlessly at the physics step, exactly as the browser steps it, until the batch completed.

8.6.1 The Nominal Batch

The design point is 200 g of 500 µm kernels, 279 865 particles presenting 0.2199 m^2 in a 7.16 mm bed at 10 900 kg/m^3, with $U_{ms} = 0.2282$ m/s reached by 18.4 slpm of argon alongside 5.5 slpm of acetylene at a spouting ratio of 2.20, and 2.04 kW of furnace holding 1250 °C with the susceptor at 1275 °C. The opening buffer rate is 5.10 µm/min. The sequencer walks all ten phases, starting the buffer at 6.6 min, the inner pyrocarbon at 36.7 min, the silicon carbide at 74.3 min and the outer pyrocarbon at 147.3 min, and finishes at 191.4 min on 12.0 kWh. The particles leave 921 µm across, the bed weighs 383 g and presents 0.745 m^2, so the denominator of Equation (8.9) rose by a factor of 3.4 while the recipe ran.

The timer and the inferential endpoint produce the same batch on a nominal plant, which is worth stating as a result. The first two rows of Table 8.4 give 95.2, 39.8, 35.1 and 40.0 µm against 95, 40, 35 and 40 on the timer and 95.2, 40.1, 35.0 and 40.1 µm on the integrated estimate, both accepted, in 191.1 and 191.4 min. Over 210 µm of coating the gap of Equation (8.20) ends at 0.09 µm and never exceeds 0.35 µm, and that residual is not drift but the difference between the seven-class plant and the one-particle estimator. An endpoint scheme that changed the answer on a nominal plant would be a scheme with a defect in it.

8.6.2 The Temperature Loop at Both Ends of the Recipe

The bed temperature loop is the only conventional one on the machine, and its process gain is not a constant. Solving the steady form of Equations (8.13) and (8.14) over a 50 °C interval with the carrier flow rescheduled by Equation (8.26) and deposition active gives 285 °C/kW at the 1250 °C buffer condition and 132 °C/kW at the 1500 °C silicon carbide condition, a factor of 2.2, because the radiative loss Q_{rad} grows far faster than the transferred heat does. The dominant time constant moves the other way for the same reason, from $C_w/UA = 8.0$ min at the buffer condition to 4.7 min at the silicon carbide condition. One set of settings therefore has to work across a gain that more than doubles, and the default $K_c = 0.130$ kW/°C with $\tau_I = 4.0$ min closes a 50 °C setpoint move to 63 % in 3.1 min at either end. Furnace power peaks at 8.31 kW of the 12 kW rating during the ramp to 1500 °C, and derating the elements to $\eta_{el} = 0.66$ puts the command against that ceiling for 4.6 min while still completing the batch in 192 min.

Table 8.4 Measured Batch Outcomes. Thicknesses and the Inferential Gap G Are in µm. Scenarios 7 and 12 Start at the Purge Ahead of the Silicon Carbide Layer and Scenarios 11 and 13 at the Buffer, So Their Batch Times Are Not Comparable With a Full Run.

#	Case	Buffer	IPyC	SiC	OPyC	σ_{SiC}	G	t (min)	Verdict
5	Timer, nominal plant	95.2	39.8	35.1	40.0	0.53	−0.08	191.1	accepted
6	Inferential, nominal plant	95.2	40.1	35.0	40.1	0.52	0.09	191.4	accepted
7	Timer, SiC 50 °C cold	95.0	40.0	26.2	41.6	0.46	−0.04	117.7	rejected, SiC mean
7	Inferential, SiC 50 °C cold	95.0	40.0	35.0	40.1	0.61	0.01	139.4	accepted
8	Inferential, $\eta_d = 0.72$	79.7	34.9	32.0	35.4	0.52	28.35	191.4	rejected, all four means
9	Inferential + mass, $\eta_d = 0.72$	97.2	41.2	34.8	39.9	0.52	−0.79	218.9	accepted
10	Inferential + mass, $\varepsilon_m = +20\%$	83.8	36.7	31.6	36.7	0.50	22.00	185.8	rejected, all four means
11	Inferential, acetylene 12 slpm	79.3	41.9	36.0	41.3	0.49	11.82	170.3	rejected, buffer mean
12	Inferential, $R_{sp} = 1.30$	95.0	40.0	35.1	40.2	2.89	−0.04	100.8	rejected, SiC variability
13	Inferential, $\chi_C = 0.62$	105.2	44.6	40.9	44.0	0.96	−24.35	183.8	rejected, all four means
14	Challenge as loaded	77.3	46.2	35.2	52.9	0.54	−34.52	325.6	rejected, three means and loss
14	Challenge, suggested settings	96.5	40.0	34.3	40.1	0.51	−1.47	208.7	accepted

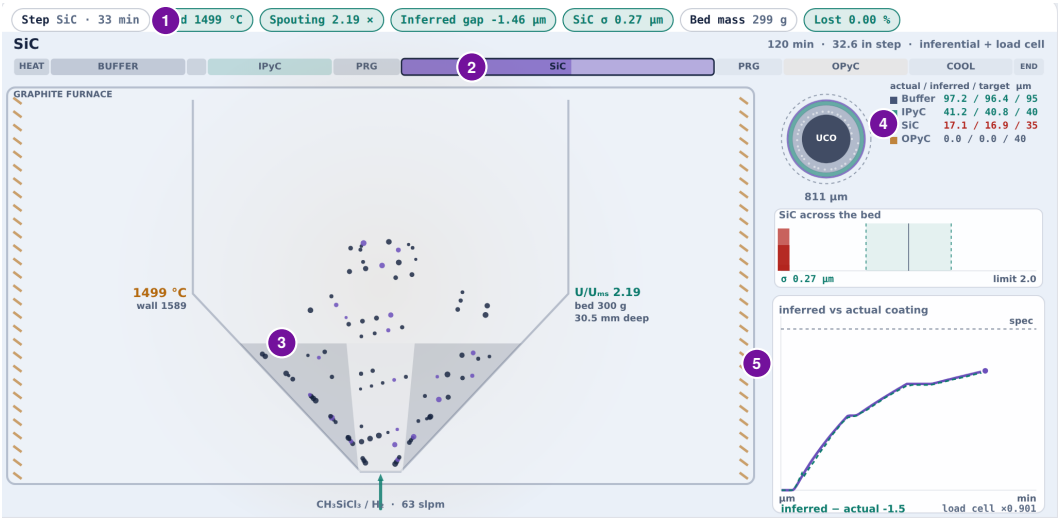


Figure 8.2 Scenario 9 at $t = 120$ min, thirty-three minutes into the silicon carbide layer, with the coating efficiency at 0.72 and the load-cell correction working (plant view cropped from the app's screen). The badges (1) read an inferential gap of −1.46 µm, which ends at −0.79 µm (Table 8.4), a spread of 0.27 µm, a load cell at 299 g against the 200 g charge and nothing lost. The recipe strip (2) shows the sequencer in the silicon carbide step, and the furnace view (3) shows the spout and the annulus. The layer table (4) reads the quantity the machine cannot measure beside the quantity it infers: 97.2 against 96.4 µm on the buffer, 41.2 against 40.8 on the inner pyrocarbon, and 17.1 against 16.9 µm of the 35 µm silicon carbide target so far, on a particle 811 µm across. The coating plot (5) carries the correction factor derived from the bed mass, 0.901, which is the mechanism this chapter argues for caught in the act. The trends of the full screen, cut here, read the bed at 1499.2 °C (thermocouple 1499.3) against a 1500 °C setpoint on 6.1 kW, and the spouting ratio at 2.201 inferred against 2.188 actual on 58.1 slpm of hydrogen carrier. The spread is far inside the 2.0 µm acceptance limit.

The gain that matters is the other one. In the exponential of Equation (8.9) at the silicon carbide condition, +50 °C multiplies the kinetic rate by 1.534 and −50 °C leaves 0.636 of it. Bed temperature is not a comfort variable here. It is a term inside an exponential in the quantity that has no instrument,

Table 8.5 The Fluidization Window at Four Points in the Recipe. Carrier Flows Are Those That Put the Bed at $R = 1$, at the Setpoint of 2.20 and at the Carry-Over Ratio of 3.00, With the Recipe Precursor Flowing.

	Buffer start	SiC start	SiC end	OPyC end
Carrier	Ar	85 % H_2	85 % H_2	Ar
T_b (°C)	1250	1500	1500	1300
$\bar{d}$ (µm)	500	770	840	920
H (mm)	7.16	26.1	33.9	44.6
ρ_p (kg/m^3)	10900	3988	3807	3352
m_b (g)	200.0	267.0	330.8	382.6
A_b (m^2)	0.220	0.521	0.621	0.744
U_{ms} (m/s)	0.228	0.601	0.789	0.565
q_C at slump (slpm)	5.9	11.0	17.1	20.9
q_C at $R = 2.20$ (slpm)	18.4	49.5	66.4	52.2
q_C at carry-over (slpm)	26.7	78.4	102.3	73.0

and scenario 7 prices it: the timer runs the layer for its 59.5 min either way and delivers 26.2 µm against a target of 35 with a tolerance of 2, while the inferential endpoint delivers 35.0 µm by running the layer for 81.1 min. That converts a quality failure into a schedule cost of 21.6 min.

8.6.3 Fluidization Through the Batch

Table 8.5 evaluates Equation (8.6) at four points in the recipe and reports the carrier flow at each wall of the operating window.

The minimum spouting velocity rises from 0.228 to 0.565 m/s between the first minute and the last, a factor of 2.47, and the carrier flow that holds the setpoint rises from 18.4 to 52.2 slpm. Three of the four terms in Equation (8.6) move. The particles grow 84 % in diameter and U_{ms} is linear in that, the bed becomes six times deeper and U_{ms} goes as its square root, and working the other way the particle density falls from 10900 to 3352 kg/m^3 as low-density carbon replaces uranium oxycarbide, which takes back a factor of 1.8. The requirement is not monotone, because the 85 % hydrogen carrier of the silicon carbide layer is five times lighter than argon at 7.71 g/mol against 39.95, and takes U_{ms} to 0.789 m/s at the end of that layer before the return to argon brings it back to 0.565. The carrier command peaks at 88.6 slpm in the purge ahead of that layer, where hydrogen has replaced argon and no precursor is in the stream to weigh the mixture down.

The consequence is that the flow which spouts the fresh charge slumps the finished bed. The window runs from 5.9 to 26.7 slpm on the fresh charge and from 20.9 to 73.0 slpm on the finished bed, overlapping only between 20.9 and 26.7 slpm. No fixed number in a recipe can hold the bed for three hours, which is why the fluidization loop has to be the schedule of Equation (8.26) and why it inherits every error the thickness estimator carries.

8.6.4 The Central Result

Scenario 8 sets $\eta_d = 0.72$ and changes nothing else. The recipe assumes that 35 % of the acetylene and 35 % of the methyltrichlorosilane land on a particle, and today 0.72 of that does. Every loop stays on setpoint, the batch takes the same 191.4 min as the nominal one, each layer ends on schedule, the estimate finishes at 95.2, 40.1, 35.0 and 40.1 µm exactly on target, and the particles carry 79.7, 34.9, 32.0 and 35.4 µm. The gap is 28.35 µm, 13 % of the coating, and all four layers are rejected. Nothing on any trend moved, because Equation (8.19) has no term in which η_d appears.

Table 8.6 Inferential Gap G and Silicon Carbide Mean Against Coating Efficiency, With and Without the Load-Cell Correction. Both Columns Use the Same Plant and the Same Recipe.

η_d	1.00	0.90	0.80	0.72	0.60	0.50
Inferential endpoint						
G (µm)	0.09	9.31	19.45	28.35	43.35	57.77
SiC mean (µm)	35.02	34.07	33.00	32.02	30.31	28.58
Batch (min)	191	191	191	191	191	191
Verdict	pass	pass	fail	fail	fail	fail
Inferential endpoint with bed-mass correction						
G (µm)	−0.86	−2.13	−1.08	−0.79	−1.44	−1.75
SiC mean (µm)	34.45	34.97	34.58	34.84	35.02	34.35
Batch (min)	191	199	207	219	236	255
Verdict	pass	pass	pass	pass	pass	pass

TRY IT IN THE APP **8 · The Estimate Lies, and Nothing on the Screen Looks Wrong**

Load the scenario and press Run. Watch the total coating trend: the estimate climbs to each target and every layer ends on schedule, in the same 191.4 min as the nominal batch. Then read the scorecard. The particles carry 79.7, 34.9, 32.0 and 35.4 µm, and all four layers are rejected.

Scenario 9 runs the identical plant with the load-cell correction of Section 8.4.3 switched on and obtains 97.2, 41.2, 34.8 and 39.9 µm, accepted, with a gap of −0.79 µm against 28.35, a reduction of 97 %, and a correction factor settling at $\kappa = 0.890$. The cost is 28 min, 218.9 min against 191.4, and 1.7 kWh of extra furnace energy, 13.8 against 12.1 kWh, because each layer now runs until the corrected estimate reaches target. Table 8.6 sweeps the disturbance to show the structure of the result rather than one instance of it.

The uncorrected gap is nearly linear in the efficiency shortfall and the batch time never moves, because the estimate reaches target on schedule whatever the plant does, as Figure 8.3 draws. The corrected gap stays inside 2.2 µm across the whole range while the batch time absorbs the disturbance, from 191 to 255 min. The corrected gap moves by about 1 µm with the load-cell and thermocouple noise seed, from −0.98 to +0.62 µm over six seeds at $\eta_d = 0.72$, so the differences between its entries are within that spread. That is the shape of a correctly built inferential scheme. An integrated model supplies resolution and speed, one real measurement supplies the level, and the schedule rather than the product carries the disturbance.

DID YOU KNOW? **The Clock Cannot Tell**

From the simulator: on the inferential endpoint every batch takes 191 min, whether the coating efficiency is 1.00 or 0.50, and the silicon carbide mean falls from 35.02 to 28.58 µm. With the bed-mass correction the mean stays near target and the batch stretches from 191 to 255 min (Table 8.6).

8.6.5 What the Correction Cannot See

Scenario 10 charges 20 % more kernels than the batch ticket records, with the correction switched on, and the batch is rejected on all four means at 83.8, 36.7, 31.6 and 36.7 µm with the estimate 22.00 µm high. The reason is structural. A mis-charge scales the particle count, and through Equations (8.3) and (8.4) it scales the surface area and the mass gain by the same factor, so the measured gain matches the modeled gain, κ stays near unity, and every thickness is reported 20 % high because it is divided by a count that came off a piece of paper. The measurement that would have caught it is the same load cell read differently. Its *absolute* value is 240 g against a ticket of 200, and in a run that starts from the bare charge the application writes that comparison as a latched warning at 0.04 min, the

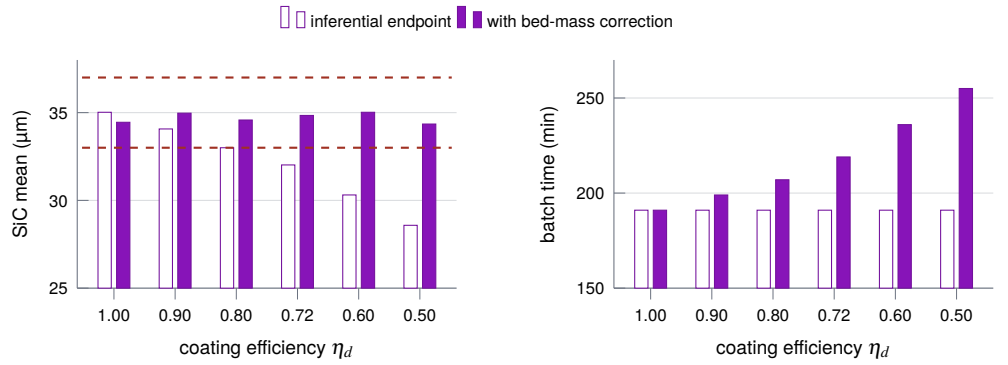


Figure 8.3 The efficiency sweep of Table 8.6. Without the correction (white bars) the batch always takes 191 min, the silicon carbide mean falls with the efficiency to the edge of the acceptance band (dashed) and below it, and every batch from 0.80 down fails. With it (solid bars) the mean stays near target and the batch time absorbs the disturbance.

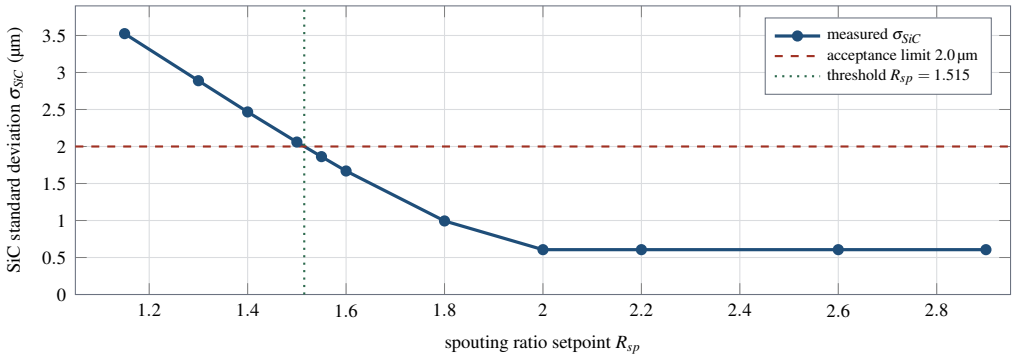


Figure 8.4 Silicon carbide layer variability against the fluidization setpoint, measured on the shipped model at setpoints on the slider grid. Every point is a silicon carbide layer whose mean thickness is within 0.15 µm of the 35 µm target, so the acceptance decision turns entirely on the width of the distribution. The spread saturates at 0.606 µm above $R_{good} = 2.00$, where circulation is fully developed. The threshold line comes from model runs at 1.51 and 1.52, between grid points.

first physics step. It saves the opening load-cell reading before deposition and compares it with the expected opening bed mass from the ticket and any completed layers. The same check therefore applies to cases that open partway through the recipe without mistaking deposited mass for a charge error. A correction built on a gain can only see errors that change the gain, so errors in the baseline have to be caught by looking at the baseline. Every soft sensor needs both, and the gross-error check is the cheaper of the two.

🕮 **Pause.** Try Exercise 8.3 before reading on.

8.6.6 Variability, and Why the Mean Is Not the Specification

Scenario 12 runs the silicon carbide layer at a spouting ratio of 1.30 instead of 2.20. The bed is still spouting, no fluidization alarm is raised, and the layer reaches 35.1 µm, dead on target. The batch is rejected, on a standard deviation of 2.89 µm against a limit of 2.0, with every mean on the scorecard inside tolerance. Figure 8.4 sweeps the setpoint.

The measured spreads are 3.53, 2.89, 2.47, 2.06, 1.67, 0.99 and 0.61 µm at setpoints of 1.15, 1.30, 1.40, 1.50, 1.60, 1.80 and 2.20, and the silicon carbide mean stays between 35.00 and 35.14 µm

across all seven. On the slider's 0.05 grid the batch is rejected at 1.50 and accepted at 1.55 with $\sigma_{SiC} = 1.86\,\mu m$, and model runs between those grid points put the crossing at 1.515, where the segregation strength of Equation (8.10) passes $\Sigma = 1.08$. Above $R_{good} = 2.00$ the curve is flat at 0.606 µm, because χ has saturated and the only remaining spread is the kernel size distribution feeding through Equation (8.2). The fluidization setpoint is therefore a product quality variable, and the operator's instinct to run the gas gently is the wrong instinct.

> **TRY IT IN THE APP 12 · The Bed Goes Slack During the SiC Layer**
>
> Load the scenario, press Run and watch the histogram beside the magnified particle. The layer ends at 35.1 µm, on target, and the batch is rejected on a spread of 2.89 µm. Press Reset, set the spouting ratio setpoint to 1.55 before the layer starts and run again: the spread is 1.86 µm and the batch is accepted.

8.6.7 Errors Upstream of Every Instrument

Scenario 13 delivers 62 % of the indicated carrier to the cone. The inferred ratio of Equation (8.26) averages 2.200 over the whole run, exactly on setpoint, while the true ratio is 1.432 twenty minutes in. Nothing in the control system has done anything wrong, because the flow meter measures what passes through the flow meter and the loop divides that by an estimated minimum spouting velocity to obtain a number that is arithmetically correct and physically false. What it costs is the spread and the mean together. The buffer standard deviation rises from 1.95 to 7.97 µm because the bed circulates at less than half the intended vigor for the entire layer, and every layer comes out over target at 105.2, 44.6, 40.9 and 44.0 µm because a smaller total flow is a richer gas. Taking the delivery to 45 % before the run slumps the bed at 27.0 min, accumulates 72.3 min below minimum spouting and agglomerates the batch at 162.0 min. The differential pressure cell is the one instrument downstream of the leak. By Equation (8.7) it cannot report the ratio, but it collapses when the bed does, and that is the general point. An inferential measurement is only as good as its most upstream input, and the defense is an independent measurement further down, however crude.

Scenario 11 raises the acetylene from 5.5 to 12 slpm to speed the buffer up, which takes the hydrocarbon mole fraction to 33 % and past the soot onset of Equation (8.9). The companion run is the same case with the acetylene set back to 5.5 slpm before Run or carried through Reset. Both paths initialize at the selected flow and produce no soot. The layer does finish faster, in 11.7 min instead of 25.0, and it finishes at 79.3 µm instead of 95.1. 15.0 g of soot leaves through the vent and the estimate reads 95.3 µm, 16.0 µm high, because Equation (8.19) has no soot term. Precursor utilization needs its observation window stated. Read at the end of the buffer step it falls from 21.3 % to 16.4, and read over the whole batch it falls from 19.8 % to 18.5, because the three later layers are untouched. Two consequences are worse than the thickness. The buffer standard deviation rises from 1.95 to 4.11 µm through the $\sigma_s s$ term of Equation (8.10), and the error propagates, because the estimator now believes the particles are larger than they are and carries an oversized $\hat{A}$ into every later layer.

The fluidization trim works against a process gain that falls through the batch. Differentiating Equation (8.6) numerically gives 0.096 units of ratio per slpm on the fresh charge, 0.028 at the start of the silicon carbide layer and 0.038 on the finished bed, so a fixed trim gain is three times more aggressive at the beginning than in the middle. Sweeping K_c over a nominal batch, the carrier command travels 177 slpm at the default value of 3.0 and 180 slpm at 20, with the extra reversals coming from thermocouple noise rather than from the loop. At $K_c = 22$ the travel jumps to 1170 slpm (927 to 1247 over five noise seeds) and the peak inferred-ratio error from 0.19 to 0.39, and at $K_c = 30$, the top of the slider, the travel is 5258 slpm. The onset of sustained oscillation is therefore near $K_c = 22$, and the batch stays acceptable throughout the sweep because the layers are long compared with the cycle. The app's hint on this gain once put the onset at 8 and now puts it at 22.

8.6.8 The Graded Challenge

The challenge loads four faults at once, an unknown coating efficiency of 0.78, a timer endpoint, a spouting ratio setpoint of 3.30 and a deliberately weak temperature tuning, and asks for an accepted batch in under 240 min. As loaded it gives 77.3, 46.2, 35.2 and 52.9 µm in 325.6 min, rejected on three means and on particle loss, because the fountain is over the freeboard for the whole run. The suggested settings, the inferential endpoint with the bed-mass correction at a spouting ratio of 2.20 and the tuned temperature loop, give 96.5, 40.0, 34.3 and 40.1 µm with $\sigma_{SiC} = 0.51$ µm, accepted in 208.7 min. The challenge carries 3 °C of thermocouple noise, and over eight noise seeds the batch as loaded is rejected on the same four criteria in 317.7 to 329.4 min while the suggested settings are accepted in 208.1 to 210.8 min. Removing the three fixes one at a time separates what each is for. The plain inferential endpoint at a ratio of 3.30 is rejected on three means and on loss. Adding the bed-mass correction while leaving the ratio at 3.30 is still rejected, on the buffer, silicon carbide and outer pyrocarbon means at 134.3, 46.3 and 47.7 µm and on 27.3 % particle loss (27.3 to 34.7 % over the seeds), because carry-over removes surface area from Equation (8.3) faster than the correction can follow it. Fixing the ratio as well gives an accepted batch, but in 329.2 min, far past the deadline, because the weak tuning makes every ramp crawl, and tuning the loop brings the same batch to 208.7 min, 31.3 min inside the deadline whether the settings are made before Run or carried through Reset. Two habits follow. Fix the measurement before the controller, because no tuning recovers information the loop does not have, and separate the faults that cost quality from the faults that cost time, because they are answered by different things.

8.7 Discussion

The four architectures are not four controllers. Manual, timer and inferential are all open loop with respect to thickness, and what the fourth closes is not a thickness loop but a model-correction loop driven by a load cell. That is the honest shape of a great deal of industrial advanced control. The improvement comes from a measurement nobody had thought to use, wired into a model that was already there, and the load cell costs a few thousand dollars and converts a rejected batch of nuclear fuel into an accepted one. When an inferential scheme is presented for approval the useful question is never how the model is tuned. It is which parameter in the model can move without any instrument noticing, and what measurement would notice.

The economics of the plant force the structure. A coating run is three hours and the metallography that says whether it worked is a week of sectioning, polishing and image analysis on a sample of particles [6]. The feedback loop that actually sets the process is therefore a week long, and everything inside the batch is open loop with respect to the property that matters. Every in-situ measurement worth having on such a plant is a way of shortening that loop, and the ranking of candidates follows from Table 8.6 rather than from instrument catalogs. The load cell is first because it bears directly on deposited mass. An off-gas analyzer would be second, because unreacted precursor and soot are the difference between the two branches of Equation (8.9). A pyrometer down the freeboard would be third, because it addresses a quantity the thermocouple already reports imperfectly rather than one nothing reports.

The same equations support methods beyond the implemented loops. The estimator of Section 8.4 is an open-loop integration with one scalar correction, and the natural replacement is a moving horizon estimator on Equations (8.9) to (8.14) that reconciles the load cell, the thermocouple and the bed pressure drop at once and estimates η_d as a parameter rather than correcting its consequence. The endpoint decision is a terminal-constraint optimal control problem, minimizing batch time subject to four terminal thicknesses, a variance limit, the carry-over limit of Equation (8.16) and the furnace rating, and its solution would schedule temperature and precursor flow together rather than holding both at recipe values.

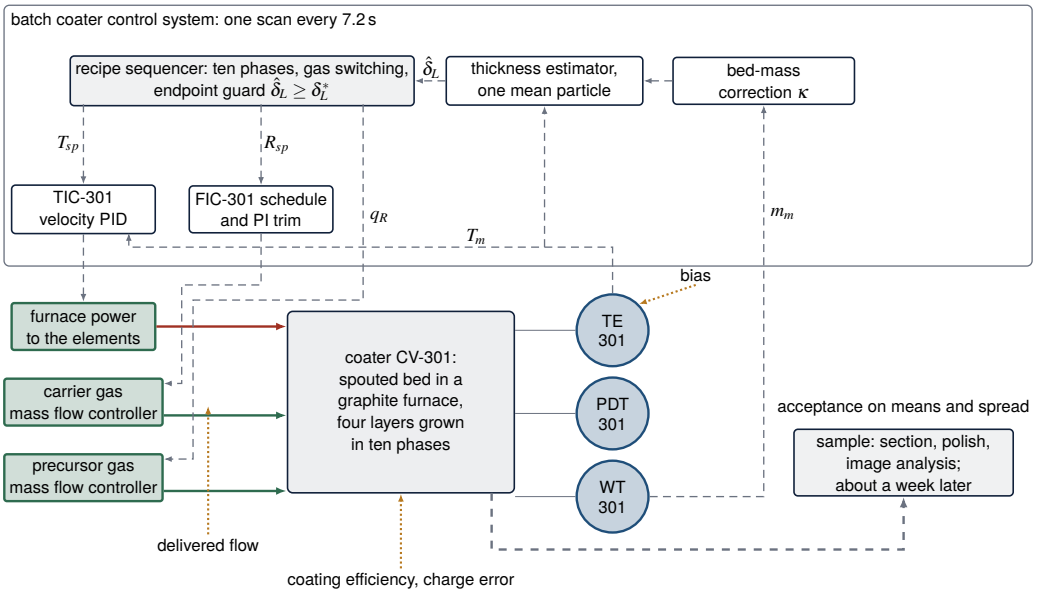


Figure 8.5 The chapter's loops on plant hardware. TIC-301 holds the bed temperature on the thermocouple through furnace power, FIC-301 schedules the carrier flow with a trim, and the sequencer sets the targets, switches the gases and ends each layer when the estimate reaches it. The load cell feeds the correction of the estimate. The thickness itself is read only on a sample, after the batch (dashed). Tags are the chapter's.

8.8 Technology in Practice

What Measures It

The coater of this chapter carries a thermocouple in the bed, a differential pressure cell across it and a load cell under the whole vessel, beside the flows its mass flow controllers indicate. None of them reads a thickness. In a qualification program the thicknesses are measured on a random sample of the coated particles, by sectioning, polishing and image analysis, and specifications are written on both the mean and the tail of each property [18]. In the model the thermocouple has a 3 s lag and an optional bias, and the load cell is filtered at 1 min. Those two signals are all the estimator and its correction use.

> **TECHNOLOGY** **Sensor: Layer Thickness on a Polished Section**
>
> The thickness this chapter controls is measured only after coating, on a random sample of particles. The particles are sectioned and polished to near their midplane, and optical micrographs of the sections are analyzed by automated image analysis. Kernel diameters are read from shadowgraphs [18]. The image processing finds each layer boundary and corrects for the distance of the cut from the midplane. The first method read 80 particles mounted in a fixed grid, and an automated stage later extended it to mounts of hundreds of particles in random positions [11]. Buffer and inner pyrocarbon densities come from hot samples of the bed, taken before the next layer is deposited [18]. In the model the true thicknesses $\delta_{L,i}$ are known only to the simulator, and the chapter places the measurement about a week after the batch.

What Moves It

Furnace power heats a graphite susceptor, and the gas enters through the nozzle. Acetylene deposits the low-density buffer, propylene mixed with acetylene and argon deposits the dense pyrocarbon layers, and methyltrichlorosilane in hydrogen deposits the silicon carbide, with or without argon [6].

The program's list of coating improvements adds argon dilution during the silicon carbide step, changes to the coater retort to raise yields, and a better vaporizer for methyltrichlorosilane that drew on computer chip manufacture [11]. In the chapter's recipe every changeover switches the carrier and the precursor together, as a first-order ramp rather than a step. Figure 8.5 draws the loops on that hardware.

> **TECHNOLOGY Final Element: Thermal Mass Flow Controllers**
>
> Among the coating improvements of the qualification program were mass flow controllers that set the gas flows during the deposition of each layer [11]. A thermal mass flowmeter carries the gas past a heated sensor and an unheated one, and the heat the flow takes away measures its mass flow. A small mass flow controller adds its own throttling valve and control electronics. The reading depends on the specific heat of the gas, so a calibration holds for one composition only, and a change between laminar and turbulent flow shifts it in some designs [19]. In the model each gas train is a first-order lag of 6 s, and the controllers report the flow through the meter, not the flow that reaches the cone.

Where the Controller Runs

A distributed control system places a supervisory level over local controllers that hold their process conditions at set points, and loops in industry run with cycle times from milliseconds to minutes [20]. The app's control system scans every 7.2 s, which the chapter takes as close to what a batch coater's distributed control system runs. Read as plant hardware, TIC-301 and the fluidization schedule FIC-301 are local regulatory loops. The recipe sequencer, with its endpoint guards, the thickness estimator and the bed-mass correction, belongs to the supervisory level that ends each phase and hands the next targets down.

> **IN PRACTICE Acceptance on a Confidence Interval**
>
> A coated particle specification limits the mean of each property and the tail of its population. For one program's silicon carbide layer the mean had to lie between 32 and 38 µm, and at most 1 % of the particles could be 28 µm or thinner, both shown at 95 % confidence from a random sample. A lot whose property lies close to a limit needs many more particles measured before it can be accepted. Quick checks such as coating thickness are often run on every batch, so that one bad batch does not spoil a composite of several [18]. This chapter writes the tail as a silicon carbide standard deviation of at most 2.0 µm.

What Is Modernizing It

The first development is automated characterization. Layer thicknesses are now measured on mounts of hundreds of particles by image analysis, and computed tomography counts the particles in a fuel compact in minutes, although its instruments are expensive and may not be found in industrial plants [11]. The second is a closer link between the bed and the layer. One study fitted three nozzle geometries to a laboratory coater, mapped their particle circulation by particle image velocimetry at room temperature, and found clear differences in the structure of the pyrocarbon each one deposited [21]. The third is modeling and data. Computational fluid dynamics can now track every particle in the bed during coating, machine learning is mining past coating work for the link between process and properties, and the field has asked for automation and for more checks during fabrication to lower the cost of quality control [22].

8.9 Limitations

The model is a demonstration of inferential control and not a coater design tool. With lumped kinetics, seven size classes and a correlation-based hydrodynamic treatment it cannot support a claim about a real machine's deposition rate, coating efficiency, particle size distribution or fluidization limits, and the slump and carry-over flows of Table 8.5 should be read as the right trend rather than as design numbers. Nothing here addresses whether a batch would survive irradiation, because layer

anisotropy, density and crystallinity are absent and those properties matter as much as thickness [9]. The population is too coarse for the far tail a real qualification is written on, so Figure 8.4 demonstrates a mechanism without predicting a failure fraction. The carry-over state is clamped at 80 %, so the challenge scenario as loaded reports a particle loss that is a bound rather than a measurement. Disturbances arrive as commanded steps rather than as the drifting behavior of a plant, and the coating efficiency the application turns on is a constant in the plant as well as in the estimator. Three elements are stylized on purpose. A coating efficiency that steps from 1.00 to 0.72 at the touch of a button is not something a coater does, but a step is the disturbance whose absence from every trend can be shown in one run. A charge weighed 20 % heavy is beyond what a real weighing error produces, although it separates a baseline error from a gain error cleanly. And the true thicknesses are displayed beside the estimated ones, which no coater can do, because the gap of Equation (8.20) is the object of the lesson, and covering that trace turns the simulator back into the real problem. Where the individual assumptions break is set out in Section 8.3.7.

> **OPEN CHALLENGE** **From the Mean to Every Particle**
>
> A retrospective of the qualification program asks that quality control move from bulk properties to particle-to-particle distributions, which would reduce the conservatism of a specification's mean. Buffer density is measured by interrupting the coating after the buffer and gives one average over many particles, and inner pyrocarbon density needs particles from interrupted runs or hot samples of the bed. Nondestructive methods that an industrial plant can deploy are wanted [11]. A workshop of the field asked for more checks during fabrication, such as particle diameter during coating [22]. Progress would be a measurement made inside the run, which would give an endpoint like this chapter's a second correcting signal beside the load cell.

8.10 Conclusion

This chapter has treated a TRISO coating reactor as an interactive laboratory for control of a quality variable that has no transmitter. The simulator integrates thirty-four states, a susceptor and a bed temperature, two mass flow controller lags, particle carry-over, an agglomeration index and one thickness per layer for each of seven particle size classes, with deposition written as the harmonic mean of an Arrhenius kinetic branch and a precursor supply branch divided by a surface area that rises from 0.220 to 0.745 m^2 through the batch. A cruder model, one mean particle at the charge as weighed and the nominal coating efficiency, is integrated alongside it and is the only thickness the control system sees. On a nominal plant the two end within 0.09 µm of each other over 210 µm of coating and the timer and the inferential endpoint deliver the same batch in 191 min. A silicon carbide step run 50 °C cold is inside the estimator's model, so the timer gives 26.2 µm and is rejected while the inferential endpoint gives 35.0 µm for 21.6 min of extra time. A coating efficiency of 0.72 is outside it and leaves the batch 28.4 µm thin with all four layers rejected, every setpoint held and every layer ending on schedule. A load cell closes that gap to 0.79 µm for 28 min, and stays inside 2.2 µm for efficiencies from 1.00 down to 0.50 while the batch time absorbs the disturbance from 191 to 255 min. The same correction is blind to a 20 % charge error, which scales the modeled and measured mass gain together and is visible only in the cell's absolute reading of 240 g against a 200 g ticket. Underneath all of it the minimum spouting velocity rises by a factor of 2.47 and the silicon carbide standard deviation rises from 0.61 to 3.53 µm as the fluidization setpoint falls from 2.20 to 1.15 with every mean thickness on target. Further work would add a moving horizon estimator on the same equations, an off-gas analyzer as a second correcting measurement and a population balance in place of the seven classes. The TRISO coating reactor is an accessible browser platform for learning what a soft sensor can and cannot defend.

KEY POINTS

- A controller can hold every measurement on its setpoint while the coating it never sees comes out thin.
- On a nominal plant the timer and the inferential endpoint give the same batch, and a cold silicon carbide step separates them.
- A coating efficiency of 0.72 leaves the batch 28.4 µm thin with all four layers rejected, and a load cell closes that gap to 0.79 µm for 28 min.
- The same correction is blind to a 20 % charge error, which only the cell's absolute reading reveals.
- The silicon carbide spread rises from 0.61 to 3.53 µm as the fluidization setpoint falls, with every mean on target.

Where else this applies. Film thickness on semiconductor wafers is measured between wafers, so its loop runs from run to run [3], as in the lithography line of Volume 14, Chapter 6.

References

[1] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[2] F. A. A. Souza, R. Araújo, and J. Mendes, "Review of soft sensor methods for regression applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 152, pp. 69–79, 2016.

[3] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer, C. Bode, S. B. Hwang, K. S. Balakrishnan, and J. Hahn, "Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities," *Automatica*, vol. 36, no. 11, pp. 1567–1603, 2000.

[4] M. Ogawa, M. Ohshima, K. Morinaga, and F. Watanabe, "Quality inferential control of an industrial high density polyethylene process," *Journal of Process Control*, vol. 9, no. 1, pp. 51–59, 1999.

[5] S. Weintroub, "The Dragon high-temperature reactor project," *Nature*, vol. 197, no. 4864, pp. 240–241, 1963.

[6] P. A. Demkowicz, B. Liu, and J. D. Hunn, "Coated particle fuel: Historical perspectives and current progress," *Journal of Nuclear Materials*, vol. 515, pp. 434–450, 2019.

[7] H. Knüfer, "Preliminary operating experiences with the AVR at an average hot-gas temperature of 950 °C," *Nuclear Engineering and Design*, vol. 34, no. 1, pp. 73–83, 1975.

[8] H. Nickel, H. Nabielek, G. Pott, and A. W. Mehner, "Long time experience with the development of HTR fuel elements in Germany," *Nuclear Engineering and Design*, vol. 217, no. 1-2, pp. 141–151, 2002.

[9] D. A. Petti, J. Buongiorno, J. T. Maki, R. R. Hobbins, and G. K. Miller, "Key differences in the fabrication, irradiation and high temperature accident testing of US and German TRISO-coated particle fuel, and their implications on fuel performance," *Nuclear Engineering and Design*, vol. 222, no. 2-3, pp. 281–297, 2003.

[10] D. Petti, J. Maki, J. Hunn, P. Pappano, C. Barnes, J. Saurwein, S. Nagley, J. Kendall, and R. Hobbins, "The DOE advanced gas reactor fuel development and qualification program," *JOM*, vol. 62, no. 9, pp. 62–66, 2010.

[11] P. A. Demkowicz, D. A. Petti, J. D. Stempien, J. Palmer, W. F. Skerjanc, D. M. Scates, G. Hawkes, T. Gerczak, and G. Helmreich, "Advanced gas reactor fuel development and qualification program retrospective and lessons learned," Idaho National Laboratory, Idaho Falls, ID, Report INL/RTP-25-06381, Revision 0, 2026, oSTI 17429367.

[12] P. A. Demkowicz, J. D. Hunn, D. A. Petti, and R. N. Morris, "Key results from irradiation and post-irradiation examination of AGR-1 UCO TRISO fuel," *Nuclear Engineering and Design*, vol. 329, pp. 102–109, 2018.

[13] Z. Yan, L. Jiang, Y. Tian, R. Liu, Y. Shao, B. Liu, and M. Liu, "Multi-scale study of fluidized bed-chemical vapour deposition process in nuclear fuel coated particle fabrication for high-temperature gas-cooled reactor: A review," *The Canadian Journal of Chemical Engineering*, vol. 103, no. 4, pp. 1502–1517, 2025.

[14] K. B. Mathur and P. E. Gishler, "A technique for contacting gases with coarse solid particles," *AIChE Journal*, vol. 1, no. 2, pp. 157–164, 1955.

[15] Z. J. Hu and K. J. Hüttinger, "Mechanisms of carbon deposition—a kinetic approach," *Carbon*, vol. 40, no. 4, pp. 624–628, 2002.

[16] H. Sone, T. Kaneko, and N. Miyakawa, "In situ measurements and growth kinetics of silicon carbide chemical vapor deposition from methyltrichlorosilane," *Journal of Crystal Growth*, vol. 219, no. 3, pp. 245–252, 2000.

[17] M. Olazar, M. J. San Jose, A. T. Aguayo, J. M. Arandes, and J. Bilbao, "Stable operation conditions for gas-solid contact regimes in conical spouted beds," *Industrial & Engineering Chemistry Research*, vol. 31, no. 7, pp. 1784–1792, 1992.

[18] P. A. Demkowicz, D. W. Marshall, and J. D. Hunn, "Advanced gas reactor fuel specification technical bases," Idaho National Laboratory, Idaho Falls, ID, Report INL/RPT-23-71992, Revision 0, 2023, oSTI 1983878.

[19] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[20] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[21] B. Conry, K. I. Montoya, R. S. Heldt, E. G. Herbert, and E. López-Honorato, "Linking spout fluidization hydrodynamics to pyrolytic carbon deposition characteristics in a fluidized bed chemical vapor deposition reactor," *Chemical Engineering Science*, vol. 334, p. 124319, 2026.

[22] P. A. Demkowicz and I. Glagolenko, "Advanced fuels campaign FY25 coated particle fuel workshop summary," Idaho National Laboratory, Idaho Falls, ID, Report INL/RPT-25-87367, 2025, oSTI 17121540.

8.11 Exercises

8.1 Concept check · 1 min

Coater CV-301 carries a thermocouple, three mass flow controllers, a differential pressure cell and a load cell. Which quantity does one of these instruments read during a batch?

(a) The mean silicon carbide thickness, through the integrated estimate

(b) The spouting ratio U/U_{ms}, through the fluidization loop FIC-301

(c) The mass of the bed, through load cell WT-301 under the whole coater

(d) The coating efficiency, through the gap between estimate and truth

8.2 Concept check · 1 min

Which statement about the balance below is right?

$$C_b \frac{\mathrm{d}T_b}{\mathrm{d}t} = UA\,(T_w - T_b) - C_g\,(T_b - T_g) + \Delta H_L\,\dot{m}_{dep}$$

(a) It is the susceptor balance, and its last term is the radiative loss to the room

(b) It is the growth law of one layer, and its last term is the deposit on one size class

(c) It is the thermocouple lag, and its last term is the bias added to the reading

(d) It is the bed balance, and its last term is the heat released as the deposit forms

8.3 Concept check · 2 min

In scenario 10 the coater holds 20 % more kernels than the batch ticket records, and the endpoint is Inferential + mass. Why does the bed-mass correction fail to rescue the batch?

(a) The deposit is spread over more particles, so the measured gain matches the model and the factor stays near one

(b) The load cell is filtered with a time constant of one minute, so the extra mass is reported too late to act on

(c) The factor is clamped between 0.25 and 3, and the factor this error needs lies outside that range

(d) The gate holds the factor at one until 6 g has been deposited, which this batch never reaches

8.4 Theory and calculation · 6 min

The silicon carbide layer has $E_a = 230\,\mathrm{kJ/mol}$ and $T_{ref} = 1500\,^{\circ}\mathrm{C}$, with $\mathscr{R} = 8.314\,\mathrm{J/(mol\,K)}$. In mid-layer of a nominal batch the two branches of Equation (8.9) are $J_{kin} = 3.09 \times 10^{-3}\,\mathrm{kg/(m^2\,min)}$ and $J_{sup} = 4.90 \times 10^{-3}\,\mathrm{kg/(m^2\,min)}$.

(a) Find the factor by which the kinetic branch changes when the bed runs at 1450 °C and at 1550 °C. *Checkpoint:* 0.636 and 1.534.

(b) Find the combined flux J at 1500 °C and at 1450 °C, with the supply branch unchanged, and their ratio. *Checkpoint:* 1.89×10^{-3} and 1.40×10^{-3} kg/(m^2 min), ratio 0.740.

(c) A 59.5 min hold lays 34.9 µm of silicon carbide at 1500 °C. Predict the thickness the same hold lays at 1450 °C, and the time the inferential endpoint needs there to reach 35.0 µm. Compare with the 26.2 µm and 81.1 min of scenario 7. *Checkpoint:* 25.8 µm and 80.6 min.

8.5 Theory and calculation · 7 min

A fresh charge is $m_0 = 200\,\mathrm{g}$ of uranium oxycarbide kernels of $d_k = 500\,\mu\mathrm{m}$ and $\rho_k = 10\,900\,\mathrm{kg/m^3}$ in a column of $D_c = 75\,\mathrm{mm}$ with an 8 mm nozzle and voidage $\varepsilon = 0.42$. At the buffer condition, 1250 °C, the gas density is $\rho_g = 0.294\,\mathrm{kg/m^3}$ and 5.5 slpm of acetylene flows. Take $g = 9.81\,\mathrm{m/s^2}$.

(a) Treating every kernel as a 500 µm sphere, find the particle count N and the surface area A_b. *Checkpoint:* 280 300 particles and 0.220 m^2.

(b) Find the bed depth H and the minimum spouting velocity U_{ms} of Equation (8.6). *Checkpoint:* 7.16 mm and 0.228 m/s.

(c) Find the carrier flow that holds a spouting ratio of 2.20, using Equation (8.5). *Checkpoint:* 18.4 slpm.

8.6 Theory and calculation · 8 min

The estimator's mean particle has a kernel radius $\hat{r}_0 = 250\,\mu\mathrm{m}$, the count is $\hat{n} = 279\,865$ and the buffer density is $\rho_1 = 1050\,\mathrm{kg/m^3}$.

(a) Find the bed mass gain the estimator expects for a 95 µm buffer. *Checkpoint:* 31.3 g.

(b) The load cell reports a gain of 22.5 g over the buffer. Invert it through the geometry, as Equation (8.21) does, and compare with the shortcut that scales 95 µm by the ratio of the gains. *Checkpoint:* 73.7 µm against 68.3 µm.

(c) In scenario 8 the true bed mass rises by 24.86 g over the buffer layer. Invert it and compare with the Buffer tile, 79.67 µm. *Checkpoint:* 79.6 µm.

8.7 App challenge · 9 min

Load scenario 7, The SiC step runs fifty degrees cold, which opens on the Timer endpoint with SiC temperature at 1450 °C. Use the default seed and the default speed of 20, which plays 20 min of the batch each second. The run pauses itself when the batch completes, and the notes record when each layer starts and ends.

(a) Run as loaded. Read SiC mean, and the SiC start and end times in the notes. *Checkpoint:* 26.22 µm, 4.0 min to 63.5 min, rejected.

(b) Press Reset, set Endpoint to Inferential and run again. Read the same quantities and the batch time. *Checkpoint:* 35.03 µm, 4.0 min to 85.1 min, accepted in 139 min.

(c) Find the ratio of the two layer durations, compare it with the ratio of rates in the theory exercise on the silicon carbide branches, and say which term of Equation (8.19) extends the layer. *Checkpoint:* 0.734.

Accept the thicknesses within 0.05 µm and the times within 0.1 min.

8.8 App challenge · 9 min

Load scenario 8, The estimate lies, and nothing on the screen looks wrong, in which the coating efficiency is 0.72 and the endpoint is Inferential. Use the default seed and speed, and run each batch until it stops by itself.

(a) Run as loaded. Read Buffer, IPyC / OPyC, SiC mean, Inferred − actual and the batch time in the last note. *Checkpoint:* 79.67, 34.9 / 35.4 and 32.02 µm, a gap of 28.35 µm, 191 min, rejected.

(b) Set Endpoint to Inferential + mass, press Reset and run again. Read the same tiles, then export the CSV file and read `k_corr` in the last row. *Checkpoint:* 97.17, 41.2 / 39.9 and 34.84 µm, a gap of −0.79 µm, accepted in 219 min, factor 0.890.

(c) Explain why the second batch takes longer, and why the first gave no warning on its temperature, spouting and coating trends. *Checkpoint:* Each layer runs until the corrected estimate reaches target, 28 min in all.

Accept thicknesses within 0.05 µm, the factor within 0.001.

8.9 App challenge · 8 min

Load scenario 12, The bed goes slack during the SiC layer, which starts at the purge before the silicon carbide layer with Spouting ratio setpoint at 1.30. Use the default seed and speed, and run each batch until it stops by itself.

(a) Run as loaded. Read SiC mean, SiC 1σ and the verdict. *Checkpoint:* 35.11 µm, 2.89 µm, rejected on SiC variability.

(b) Press Reset, set the setpoint to 1.50 and run. Repeat at 1.55. Make each setting before the layer starts at 4 min. *Checkpoint:* 2.06 µm rejected; 1.86 to 1.87 µm accepted.

(c) Explain why the mean stays on target while the spread changes, and why raising the setpoint above 2.0 no longer helps. *Checkpoint:* The spread is 0.61 µm at 2.0 and above.

Accept SiC 1σ within 0.01 µm of the stated values or range.

8.10 Application area · 3 min

A plant manager notes that every mean in scenario 12 at a ratio of 1.30 was in tolerance, and proposes dropping the silicon carbide spread from the acceptance criteria. Give two reasons from the chapter for keeping it.

8.11 Application area · 3 min

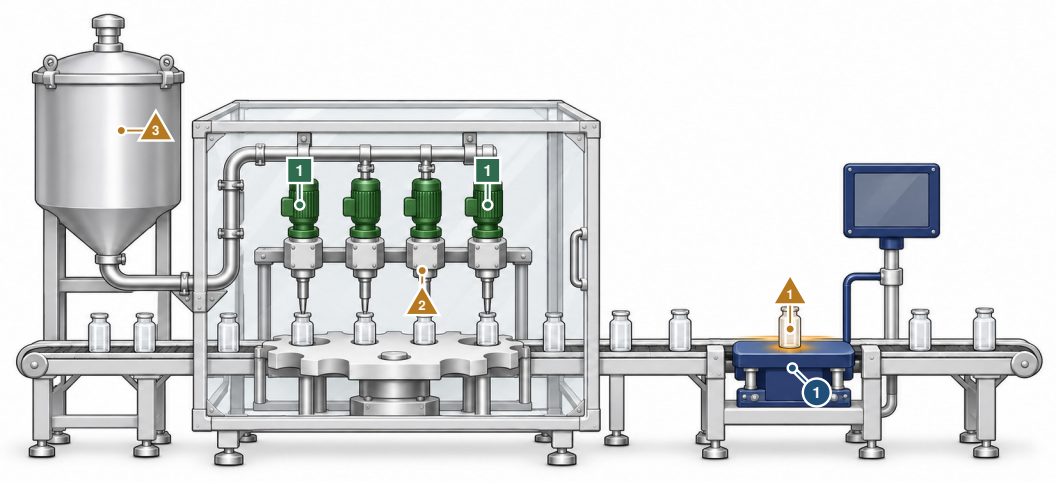
For each fault below, name the measurement on the coater that reveals it and say what that measurement compares: a coating efficiency of 0.72, a charge 20 % heavier than its ticket, and a carrier leak that delivers 62 % of the indicated flow.

8.12 Review the agent · 5 min

An AI assistant is asked how to pass the challenge, A batch with the efficiency unknown, and writes: “The coating efficiency is the only fault that costs quality, and only the load cell bears on it. Set Endpoint to Inferential + mass and leave the other settings as loaded. The correction will absorb the efficiency error, so the batch will be accepted, if slowly.” Find the flaw and say how you would check it.

CHAPTER 9

Aseptic Vial Filling Simulator With Drifting Heads and Late Checkweigher Results



The controller moves 1 the pump behind each of the four needles

The controller reads 1 the scale that weighs a few of the filled vials

What disturbs it 1 weights that arrive after more vials are filled 2 a pump tube that wears 3 a thicker batch of liquid

Notice the distance between the needles and the scale: by the time a vial is weighed, its needle has already filled the next ones. Illustration, not to scale.

IN PLAIN TERMS **Four Needles and a Scale That Reports Late**

The main idea. A machine fills small glass bottles of medicine through four needles, each fed by its own pump. A scale weighs some of the bottles, and a controller uses those weights to nudge each pump up or down.

The challenge. Each weight arrives after the same needle has already filled several more bottles. A controller that treats every late weight as fresh news counts the same error again and again, and the pumps start to swing.

Why it matters. A short bottle that is never weighed can still be shipped, so how often the scale weighs matters as much as the controller.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the first fact in the Introduction, the history in Section 9.2, the second try-it and the second fact in Section 9.3, the first try-it in Section 9.6, and the key points
Study	40 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 9.6), Technology in Practice (Section 9.7), Limitations (Section 9.8) and the open challenge

The control question A checkweigher weighs one vial per head every few cycles and reports it after later vials are filled. Which correction keeps every head on target without counting a late result twice, and what does the weighing interval decide that no correction can?

Learning objectives

- Explain why a pooled correction keeps the average fill right while one head fills under its target.
- Compute the cycle timing, pump speed ratio and spread factors from the line rate and the viscosity.
- Derive the stability limit of the integral correction with results in flight, and explain why the record-based estimate escapes it.
- Show from the dose records why canceling each whole error widens the fills.
- Choose the weighing interval, correction, guard band and target that release no vial outside its limits with the least giveaway.

Prerequisites Sections 1.4.5 (run-to-run control, the EWMA and integral forms and the metrology delay), 1.6.1 (engine and seed) and 1.6.4 (scenarios, scorecard and data). Chapter 7 applies the integral form to one machine with no result in flight.

The app https://apmonitor.com/apps/aseptic-vial-filling/

See also Volume 5, Chapter 6 for tablet weight held on a press, Volume 5, Chapter 9 for a correction from a valuation a year late, and Volume 5, Chapter 3 for control through a slow analyzer.

9.1 Introduction

Four peristaltic heads on the simulated line fill 72 vials a minute with 1.250 g of an antibody solution whose label minimum is 1.236 g. The whole margin between product given away and an underfilled vial is 14 mg, about 2.5 times the 5.7 mg spread of one replayed dose. In the first case of the application a correction that keeps the average of every weighed group on target lets one wearing head fill 10.4 mg under the target over its last 100 vials while it lifts the other three heads by 0.9 to 6.7 mg, and 32 underfilled vials leave the line accepted because three groups in four are never weighed. The average fill is right in that batch. One needle is not.

Aseptic filling of sterile drug and biological products follows current good manufacturing practice, for which a 2004 guidance replaced the 1987 guideline [1], and the dose is set by the filling pump. Peristaltic pumps are an industrial standard for filling high-concentration antibody formulations into vials and syringes, a precision target of 3 % at 0.3 mL has proved challenging for viscous liquids, and fill-weight variability tracked the height of the liquid drawn back at the end of each fill (the suck-back) rather than any pump setting [2]. In syringe filling the speed of that suck-back has also decided whether a filling needle clogged [3]. An open data descriptor reports 149 847 peristaltic doses from 0.1 to 2.0 mL taken under controlled conditions with calibrated weighing equipment [4, 5], and its authors have compared statistical and ensemble-learning compensation of the doses [6]. Balance readings have served as the feedback signal for the flow of a peristaltic pump [7], and adaptive filtering has been applied to the dynamic weighing of a checkweigher [8].

DID YOU KNOW? Product Worth Weighing

A 2025 dose study notes that biopharmaceutical products can cost up to 10 000 dollars a gram, so even small gains in dosing accuracy save product [4]. The suggested settings of this chapter's challenge give away 13.25 mg per accepted vial (Table 9.5).

Two beliefs are the simulator's target. The first is that a filling line has one fill weight, its average, when each head is its own pump, tube and needle. The second is that a correction which removes every measured error makes the fills tighter. A result is one dose with that dose's own noise, and it describes a vial filled tens of seconds earlier with a command that later corrections have already

changed. Adjusting a process after each deviation can increase the variance of its output, the lesson usually drawn from Deming's funnel, although active control to a target improves the output of the many processes whose disturbances are not of that kind [9], and which case applies rests on assumptions that often go unstated [10]. A measurement that arrives after later units have been processed is the metrology delay of run-to-run control, where delay combined with model error can destabilize a closed loop [11]. The same shape appears wherever discrete units are dosed from late, sampled measurements, from a silicon epitaxy reactor corrected run by run [12, 13] to a feed pump controlled from a balance [7]. Volume 5, Chapter 6 holds tablet weight on a rotary press, and Volume 5, Chapter 9 resets a contribution from a valuation a year late. A dead-time transfer function captures neither the sampling nor the identity of the unit a result describes.

However, open records of doses cannot show how another controller would have changed the doses they contain. In the compensated series of the dataset the command is a function of the compensation's own prediction, so command and disturbance are confounded, and only the uncompensated series recorded at a fixed command can be replayed through a different correction, under independence assumptions that have to be stated. This chapter sets fixed calibration, a pooled correction, a filtered per-head correction and a record-based per-head estimate side by side on a four-head filling simulator whose dose noise is replayed from those uncompensated records, with the tails of the fill distribution, the giveaway, the rejects and the recovery after a restart reported apart. Geometric moving averages [14], adjustment from past data [15], run-to-run control [13, 16] and its delay problem [11] are established, and the comparison rests on a declared action-response surrogate, not on prospective experiments. The simulator has a causal dispense model per head, a chronological split of the records into training and validation thirds, one write-once record per vial whose tare, gross weight, head and command stay aligned through every delay, and a bulk balance that closes to round-off, and the chapter gives every equation and parameter needed to build it again.

Section 9.2 follows the problem's history. Section 9.3 sets out the line: its cycle, the dispense, the drop, weighing, disposition and the parameters with their evidence. Section 9.4 splits the dose records and reads the noise-chasing lesson from them, and Section 9.5 gives the corrections as implemented. Section 9.6 measures them on eight cases and a challenge, the pipeline sweep and a policy comparison over seeds. Section 9.7 sets the line on real instruments and pumps, Section 9.8 states the limitations, Section 9.9 concludes, and the exercises of Section 9.10 run the cases.

9.2 Giveaway, Sampling and Feedback Adjustment

Setting a fill target is an older problem than correcting a filler. A graphical procedure published in 1977 chose the target of a process whose items are each checked against a lower specification limit, balancing the proportion sold at a reduced price against the give-away cost of setting the target further from the limit [17]. The simulated challenge poses that trade with one complication, since most vials are released on the evidence of the weighed vials around them rather than checked themselves.

The filter in every correction of the simulator came from control charts. A geometric moving average, which gives the newest observation the greatest weight and the older ones weights that fall geometrically, was proposed as a chart test in 1959 [14]. Three years later, keeping a variable such as the concentration of consecutive batches close to a target was posed as the estimation of the best adjustment from past data [15]. Whether to adjust at all was argued for decades afterward. Deming's funnel experiment illustrates that active control can increase the variance of a process output, and a 1990 note analyzed when that holds and how control to a target improves the output in the continuous process industries [9]. A 1992 discussion set statistical process monitoring beside feedback adjustment, illustrated the importance of assumptions that often go unstated and derived minimum-cost feedback schemes for simple models [10].

Semiconductor manufacturing turned these ideas into run-to-run control. A condition relating the EWMA weight to the estimated process gain was shown to keep an EWMA controller stable for a process that drifts or wanders [12]. A run-by-run controller that combined statistical process control

with feedback recovered a silicon epitaxy process within 3 runs after a disturbance and reduced its variation by a factor of 2.7 against historical data [13], and a 1997 review recast the methods from statistical and control viewpoints with self-tuning controllers that also handle trends and autocorrelation [16]. As production outran inspection, quality measurements arrived after later operations had started, and a sufficient condition for the stability of a multivariable EWMA controller under any length of metrology delay was derived from the eigenvalues of a model-mismatch matrix [11].

Filling followed a path of its own, from the 1987 guideline to the dose records of 2025, told in the history box below, and the simulator takes no number from the 2004 guidance [1]. The line in this chapter brings the two paths together, a run-to-run correction per head fed by a checkweigher whose results arrive late.

> ⧖ HISTORY **1987 to 2025: Filling Learns From Its Weights**
>
> A 2004 guidance on sterile drug products made by aseptic processing replaced the guideline of 1987. It counts any intervention or stoppage as a risk of contamination, and it names an on-line weight check device as one way to take a repeated manual task out of the critical area [1]. Bench studies of viscous antibody fills by peristaltic pump traced the variation of the suck-back height, which followed the fill weight, to two causes: a liquid stream that separated at an inconsistent point at the end of fill, and pressure variations on suck-back [2]. A dataset published in 2025 weighed one container tare and gross around every dose and calibrated the pump once at the start of every subset of 300 doses [4]. Compensation learned from such records raised dosing accuracy by 53.93 percent at 0.3 mL in a test on the real pump [6].

9.3 The Filling Line as Modeled

The simulated line is a teaching design drawn from published sources, and it copies no commercial filler. Figure 9.1 shows it. Qualified bulk, already sterile-filtered and transferred, waits in one vessel that feeds four peristaltic heads. Empty vials enter four at a time, one per lane, and an intermittent-motion track indexes the group one pitch per cycle. Every vial is tared on its lane's cell at the first pitch, filled under the needle of its lane's head at the second, weighed gross on the sampled cycles $D = 2$ pitches later while still open, given a stopper from a track fed by a vibratory bowl, checked by a presence sensor, and closed out as accepted, rejected or held. The tare and the gross weight belong to the same vial in the same closure state, so no stopper mass has to be compensated.

Each variable in Table 9.1 carries its role. The physical state is a set of event-driven quantities rather than integrated ones: the bulk mass, the liquid in vials, the drip tray, the purge cup, the hold-up in tubes and on needles and the accepted mass make six ledger states, and each head carries its gain G_h, its drop r_h, its relaxation a_h and its stroke counter n_h. Every vial has one sealed record with its head, group, empty mass, tare and its time, command, stroke, fill start and end, true liquid mass, bubble and drip flags, sampling flag, gross weight and its time, report time, net, stopper, sensor reading, aseptic flag, bracketing samples and disposition, each field written once.

9.3.1 The Cycle and Its Events

A cycle at line rate R (vials per minute, four per cycle) lasts $T_c = 60 \cdot 4/R$, of which t_{idx} is the index move and t_{set} a settle, so the fill window and the pump-speed ratio follow as shown in Equation (9.1).

$$T_f = T_c - t_{idx} - t_{set}, \qquad \frac{\omega}{\omega_{ref}} = \frac{T_f(R_{ref})}{T_f(R)} \tag{9.1}$$

where $R_{ref} = 60\,\text{vials/min}$ is the rate the dose statistics are referred to. At 72 vials/min, $T_c = 3.333\,\text{s}$, $T_f = 2.133\,\text{s}$ and the pump runs 1.3125 times its reference speed. A weighing cycle dwells at least $t_w = 3.0\,\text{s}$, which slows a line that weighs every group to 61.5 vials/min before stops. Each event of a cycle happens once, in order, at the process time it falls on: the index, the tare, four dispenses that read their heads' commands at their start, the gross weighing of a sampled group, the stoppers, the close-out, and each result's delivery a latency later. The time step of the engine (Section 1.6.1) is 0.1 s, a thirtieth of a cycle, and nothing is integrated between events.

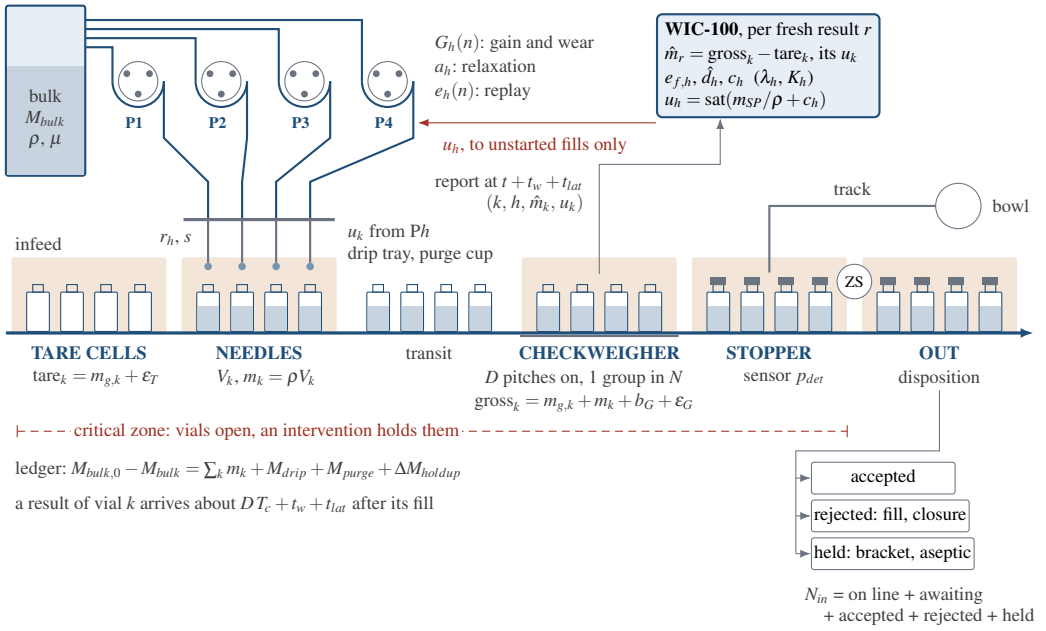


Figure 9.1 Four-head filling line as modeled, after the Line view of the simulator. The bulk (M_{bulk}, density ρ, viscosity μ) feeds heads P1 to P4 with calibration and wear gains G_h, relaxation a_h after a stop and replayed deviations $e_h(n)$ indexed by each head's stroke n. Under the needles vial k receives V_k for the command u_k its head held when the dispense started, with the hanging drop r_h set by the suck-back s. The checkweigher weighs one group in N with zero drift b_G, and WIC-100 reads only the reported records, net of each vial's own tare, and acts on dispenses that have not started. The critical zone runs from the tare cells to the stopper. Process model only, not a construction drawing. Not to scale.

Table 9.1 Variables, Their Roles and Their Reference Values

Variable	Role	Reference	Range
$\hat{m}_r$	CV, reported net of result r	—	1 mg steps
m_{SP}	setpoint, fill target (net)	1.250 g	1.230–1.290 g
u_h	MV, command of head h	1.2136 mL	1.10–1.32 mL
N	decision, weigh 1 group in N	4	1–20 cycles
R	decision, line rate	72	30–100 vials/min
s, n_p	decision, suck-back, purge shots	1, 0	0–3, 0–5
g_b	decision, reject guard band	0	0–5 mg
w_h	DV, tubing wear	0	0–5 %/1000 doses
μ	DV, bulk viscosity	1.0	0.8–12 mPa s
β_G	DV, gross-scale zero drift	0	±40 mg/h
x_e	DV, dose-noise multiplier	1	0–3
—	DV, events	—	lot, jam, glove, infeed
G_h, a_h, r_h	states, head (truth)	1, 0, 8 µL	—
b_G	state, scale zero (truth)	0	—
$e_{f,h}$, $\hat{d}_h$, c_h	states, controller	0	—
M_{bulk}	state, ledger	12 kg	—
m_k	output, true fill (truth)	—	—

9.3.2 The Dispense

The dispense is a causal surrogate, a calibrated volume command times the head's gain plus a recorded deviation, rather than a pump mechanism, as shown in Equation (9.2).

$$\begin{aligned} V_{p,k} &= u_k\, G_h(n)\, \frac{\gamma}{\gamma_0}\left(1 + a_h e^{-j/n_{rel}}\right) + x_e f_\sigma\, e_h(n) \\ V_k &= V_{p,k}\,(1-\beta_{air,k}) + r_h^- - r_h^+, \qquad m_k = \rho V_k \\ G_h(0) &= 1 + \sigma_{cal} z_h, \qquad G_h(n+1) = G_h(n) - 10^{-5} w_h \end{aligned} \tag{9.2}$$

where u_k is the command the head held when the dispense of vial k started, written into the record and never changed, n the head's own stroke count (a purge shot counts), j the doses since the last restart, r_h^- the drop on the needle before the fill and r_h^+ the drop it leaves, $\beta_{air,k}$ the share an air bubble displaces (zero without one), a_h the relaxation left by the last stop as a fraction of a dose, w_h the wear in percent per 1000 doses, x_e a dose-noise multiplier (1 by default) and z_h a unit normal drawn once per head and run. The response of a dose to its command, one gram of fill per gram asked after calibration, is a teaching assumption, because the records cannot identify it (Section 9.4). The pump speed and the viscosity change the delivery and widen the deviation as shown in Equation (9.3).

$$\begin{aligned} \gamma &= 1 + k_\omega\left(\frac{\omega}{\omega_{ref}} - 1\right) + k_\mu \ln\frac{\mu}{\mu_{ref}} \\ f_\sigma &= \left(\frac{\omega}{\omega_{ref}}\right)^{\beta_\omega}\left(\frac{\mu}{\mu_{ref}}\right)^{\beta_\mu} \end{aligned} \tag{9.3}$$

where γ_0 is γ at the rate and viscosity the heads were calibrated at, the start of the batch. At 72 vials/min on water $f_\sigma = 1.146$, and a lot at 6 mPa s takes 2.15 % off every head's delivery and widens its deviation 1.565 times.

9.3.3 Recorded Deviations and the Fitted Alternative

The default deviation $e_h(n)$ is replayed from the validation third of the dataset's uncompensated 1.2 mL series: 18 calibration segments of 300 doses, each less its own mean so that the calibration offset stays the separate state $G_h(0)$, converted from milligrams of water to volume at 998.2 kg/m^3. Head h starts at segment $[b+4(h-1)]$ mod 18 for a seed-chosen b and runs on through the following segments, indexed by its own stroke count so that a stop neither skips nor repeats a recorded dose. The fitted alternative is the mechanism behind the sign of the records' autocorrelation, a white pump term and a drop that leaves with the next fill, as shown in Equation (9.4).

$$e_k = p_k + r_{k-1} - r_k, \qquad \rho_1 = -\frac{\sigma_r^2}{\sigma_p^2 + 2\sigma_r^2} \tag{9.4}$$

where $p_k \sim N(0, \sigma_p^2)$, the drop is $\bar{r} + \sigma_r z$ and ρ_1 is the lag-1 autocorrelation. Fitting σ_p and σ_r to the training statistics gives 2.85 µL and 3.17 µL. Under the replay the drop is held at its mean $\bar{r} = 8$ µL, since the replayed deviation already carries the carry-over. Each stroke draws the same five random numbers whatever the settings, so a pair of runs that differ in a single setting meets identical disturbances.

9.3.4 Drop, Drips, Bubbles and Restarts

The suck-back setting s scales the drop by $\exp[-0.35(s-1)]$ and the drop grows with $(\mu/\mu_{ref})^{0.3}$. A drip, with probability $5\times10^{-4}\exp[-3(s-1)]$ per dose, drops the drop into the drip tray so the next vial lacks it, and a bubble, with probability $5\times10^{-4}\exp[2(s-1)]$, displaces $\beta_{air} = 4$ % of the dose (standard deviation 35 % of that). At the dataset's setting of 1 the two events are rare by construction. A pause longer than 10 s relaxes the tube by $a_h = a_{max}(1 - e^{-T_s/\tau_{rel}})$, with $a_{max} = -1.2$ % of a dose

and $\tau_{rel} = 300\,\text{s}$, which the next doses recover over $n_{rel} = 2$ doses through the factor of Equation (9.2), and a stop longer than $t_{drop} = 30\,\text{s}$ lets each hanging drop fall to the drip tray. A restart purge sends n_p doses per head to the purge cup before the first vial. All of these laws are teaching assumptions shaped by the qualitative findings on suck-back [2, 3].

> **TRY IT IN THE APP** **7 · Stop and Restart**
>
> Load the scenario, open the Restart view and press Run. The line stops twice for 4 min, and the first fills after a restart come out 10.2 to 23.7 mg under the target, with 8 vials rejected and 36 held. Press Reset, set Restart purge shots to 2 and run again: the first fills lie between −9.0 and +2.2 mg, with 2 rejected and 12 held.

9.3.5 Weighing and the Record

The tare and gross readings carry their own noise and quantization, and the gross scale's zero is a hidden state apart from any pump gain, as shown in Equation (9.5).

$$\begin{aligned} \text{tare}_k &= q\,\text{round}\big[(m_{g,k} + \sigma_s z_{T,k})/q\big] \\ \text{gross}_k &= q\,\text{round}\big[(m_{g,k} + m_k + b_G + \sigma_s z_{G,k})/q\big] \\ \hat{m}_k &= \text{gross}_k - \text{tare}_k, \qquad \frac{\mathrm{d}b_G}{\mathrm{d}t} = \beta_G \end{aligned} \tag{9.5}$$

where $m_{g,k}$ is the empty vial's mass (6.5 g, standard deviation 80 mg), $\sigma_s = 0.5\,\text{mg}$ is the dataset scale's stated repeatability bound used as a standard deviation, and $q = 1\,\text{mg}$. A weighing at time t is reported at $t + t_w + t_{lat}$, with $t_{lat} = 0.5\,\text{s}$ by default, 40 s in case 3 and 20 s in the challenge. The first group after a restart and the last group of the batch are always weighed. Without that restart check case 7 accepts 7 underfilled vials instead of 3, with fewer rejects in view. A 2 g reference weight on the gross scale reads b_G plus noise, the only instrument that tells a drifting zero from a drifting head.

9.3.6 Disposition and the Ledger

Each vial is closed out once at OUT by teaching rules. A stopper the sensor finds missing (probability of detection 0.995) rejects it for closure, and a vial open in the critical zone during a glove intervention is held, whatever it weighs. A weighed vial is accepted when its net lies within the fill limits narrowed by the guard band g_b and rejected otherwise. An unweighed vial is accepted only when the weighed vials of its own head before and after it, in fill order, both read inside the narrowed limits, and held otherwise, which leaves it awaiting until its later bracket is reported. The limits are teaching values, 1.236 g (the 1.20 mL label at 1.030 g/mL) and 104 % of it, 1.2854 g. The bulk balance is the ledger of Equation (9.6).

$$M_{bulk,0} - M_{bulk} = \sum_k m_k + M_{drip} + M_{purge} + \Delta M_{holdup} \tag{9.6}$$

where the hold-up is the liquid in four tubes of 2.26 mL and on four needles. The giveaway is the mean of m_k less the label minimum over the accepted vials, a quantity of teaching truth that no instrument on the line reads.

> **DID YOU KNOW?** **Stopping the Line Ships More Vials**
>
> From the simulator: when the stopper bowl jams for 3 min and the line keeps running, 148 vials leave unstoppered, all are found and rejected, and 2201 are accepted. Stopping on a low stopper track leaves none unstoppered and accepts 2354, although the line then makes 62.0 against 66.2 vials/min (Table 9.5).

9.3.7 Parameters and Evidence

Table 9.2 lists every parameter with its default and its label, measured (M) when read from the descriptor or its data, fitted (F) when estimated here from the training segments, and teaching (T)

Table 9.2 Parameters, Defaults and Evidence (M Measured, F Fitted, T Teaching)

Symbol	Meaning	Value	Unit	Ev.
$e_h(n)$	replayed deviation	18 seg.	—	M
σ_{cal}	offset one calibration leaves	0.363	% of dose	F
σ_p, σ_r	fitted pump, drop terms	2.85, 3.17	µL	F
s	suck-back of the records	1	device	M
σ_s, q	scale noise, resolution	0.5, 1	mg	T, M
ρ	bulk density	1.030	g/mL	T
V_{label}	label volume	1.20	mL	T
—	upper limit	104	% of label	T
g	fill per command, calibrated	ρ	g/mL	T
μ_{ref}	viscosity of the records	1.0	mPa s	T
k_ω, β_ω	speed: delivery, spread	−1.0 %, 0.5	—	T
k_μ, β_μ	viscosity: delivery, spread	−1.2 %, 0.25	—	T
$\bar{r}$	hanging drop	8	µL	T
a_{max}, τ_{rel}, n_{rel}	tube relaxation	−1.2, 300, 2	%, s, doses	T
t_{drop}	stop before the drop falls	30	s	T
t_{idx}, t_{set}	index move, settle	0.9, 0.3	s	T
R_{ref}	rate of the records	60	vials/min	T
t_w, D	weighing dwell, position	3.0, 2	s, pitches	T
t_{lat}	result latency	0.5	s	T
m_g	empty vial, mean and SD	6.5, 0.08	g	T
—	stopper track, bowl feed	60, 2.5	–, /s	T
—	track stop, restart levels	12, 40	stoppers	T
p_{det}	stopper detection	99.5	%	T
v	validity window	2.5	% of m_{SP}	T
Δu_{max}	largest correction step	1	% of u_{nom}	T
K_h, λ_h	gain, filter	0.50, 0.30	mL/g, —	T

when no source fixes it. The batch starts from a checkpoint, a primed and calibrated line with 12 kg of bulk and no vial filled.

9.3.8 Assumptions

The dispense is an empirical surrogate, so pressure, compliance, tube geometry and the roller pulsation are absent, and the model cannot say how a different tube or pump would behave. Its deviations are those of water on one pump and one container, and replaying them through another correction is valid only under three independence assumptions: a deviation does not depend on its command within the range the corrections move it, the four heads' deviations are independent of one another, and a deviation depends on the head's stroke count rather than on the line's timing or its stops, with a constant factor f_σ carrying the pump speed. The commands stay within 2.7 % of nominal outside cases 3 and 5, reach 3.4 % in case 5 and 6.7 % in case 3's oscillation. The drop, drips, bubbles, relaxation, viscosity and speed laws are teaching closures. The density does not change with the lot. Stoppering is a presence check, and crimping, container-closure integrity, visual inspection and contamination are not modeled. An aseptic hold is a discrete operational state, never inferred from a weight, so a run with no rejects is not evidence of sterility.

9.4 The Dose Records

The dataset's core 1.2 mL file holds 16 195 doses separated by 54 calibration markers into 54 segments, weighed at 0.001 g, and the replayed segments are used under the dataset's CC BY license [5]. The split is chronological, to avoid leakage between neighboring subsets: the first 36 segments (10 795 doses) train and the last 18 (5400 doses) validate, so that no validation segment lies between two training segments. The statistics are given in Table 9.3. The within-segment spread and the lag-1 autocorrelation differ between the thirds, which is why the application replays the validation third by default and uses the model fitted on the training third only as an option. The fitted model reproduces its training statistics and overstates the validation spread by 10 %.

Table 9.3 The 1.2 mL Records Split by Calibration Segment, and the Next-Dose Prediction Error of an EWMA Restarted at Each Calibration

Statistic, 1.2 mL	Training	Validation
segments, doses	36, 10 795	18, 5400
within-segment SD (mg)	5.31	4.81
SD of the segment means (mg)	4.35	3.90
lag-1, lag-2 autocorrelation	−0.356, +0.084	−0.228, −0.129
drift (mg per 100 doses)	−0.05 ± 0.32	−0.22 ± 0.19
first 3 doses less the rest (mg)	−0.02	+0.10
RMSE, target only, $\lambda = 0$ (mg)	6.82	6.12
RMSE, EWMA $\lambda = 0.05$ (mg)	5.47	4.95
RMSE, last dose, $\lambda = 1$ (mg)	9.33	7.98
Validation RMSE, other volumes	fixed / best / last	λ
0.2 mL (mg)	4.33 / 3.39 / 4.99	0.10
0.3 mL (mg)	5.46 / 3.70 / 5.96	0.05
2.0 mL (mg)	10.31 / 6.71 / 10.76	0.05

Successive doses are negatively correlated, a dose that runs high tends to be followed by one that runs low, and the spread of the segment means (4.35 mg on training) is what one recalibration leaves on a head, which fixes $\sigma_{cal} = 0.363\,\%$. A drop that hangs on the needle and leaves with the next fill gives that sign, consistent with the suck-back mechanism [2], although the records cannot tell it from a reading artifact. A gross reading reused as the next tare would give a lag-1 of about −0.01 at the stated scale repeatability of 0.5 mg, far from −0.36. Within a segment the slope spread matches the 0.35 mg per 100 doses that white noise gives in 300 doses, so no drift is identified, and the first doses after a calibration do not differ from the rest beyond –0.02(48) mg (standard error over 36 segments), which bounds the effect of a short pause but not of a stop of minutes.

The prediction rows of Table 9.3 are the data's own statement of the noise-chasing lesson. An EWMA of the dose error restarted at the target after each calibration, with its weight chosen on training, predicts the next validation dose to 4.95 mg, against 6.12 mg for the target alone and 7.98 mg for the last dose. The order holds at every volume in the lower rows. Under the independence assumption a correction that cancels the predicted error leaves exactly the prediction error, so these are also the corrected-dose spreads of a perfect actuator. They are prediction results, not closed-loop measurements. The records cannot identify the response of a dose to its command, because in the compensated files the command is $V_c = V_{tg}^2/V_p$, a deterministic function of the compensation's prediction V_p of the dose at target V_{tg}, so that $(V_c/V_{tg})V_p \equiv V_{tg}$ and command and disturbance cannot be separated.

9.5 The Corrections as Implemented

The fill-weight controller WIC-100 executes every 0.1 s but changes a command only at a fresh result. Its inbox holds, for each reported vial, the vial and head identifiers, the net $\hat{m}_r$ of that vial's own gross and tare, the command u_r that vial was filled with and the report time. It never reads a true fill, a pump gain or the scale's zero. A result is first checked for validity against the median of its head's last five results, valid or not, as shown in Equation (9.7).

$$\left|\hat{m}_r - \mathrm{med}_5(\hat{m}_h)\right| \le v\, m_{SP} \tag{9.7}$$

where the median is replaced by m_{SP} before three results and $v = 2.5\,\%$ (31.25 mg). An invalid result is counted and changes nothing. The median follows a real shift after three results, so the window cannot lock a head out, while a bubble's 50 mg short dose falls outside it. A window centered on the target instead froze the correction of case 3 for the rest of the batch, with 1020 results ignored and 282 vials accepted against 648. The command of head h is the recipe's $u_{nom} = m_{SP}/\rho$ plus a

correction c_h, updated by one of three laws, as shown in Equation (9.8).

$$\begin{aligned}
\text{per head: } & e_{f,h} \leftarrow \lambda_h(m_{SP} - \hat{m}_r) + (1-\lambda_h)e_{f,h} \\
& c_h \leftarrow c_h + \lim(K_h e_{f,h}) \\
\text{record: } & \hat{d}_h \leftarrow \hat{d}_h + \lambda_h(\hat{m}_r - \rho u_r - \hat{d}_h) \\
& c_h \leftarrow c_h + \lim\big(K_h\rho(-\hat{d}_h/\rho - c_h)\big) \\
\text{pooled: } & \bar{e}_f \leftarrow \bar{\lambda}\,(m_{SP} - \bar{m}_g) + (1-\bar{\lambda}\,)\bar{e}_f \\
& c_h \leftarrow c_h + \lim(\bar{K}\bar{e}_f) \quad \text{for every } h \\
& u_h = \mathrm{sat}(u_{nom} + c_h), \qquad c_h = u_h - u_{nom}
\end{aligned} \tag{9.8}$$

where lim caps a step at $\Delta u_{max} = 1\,\%$ of u_{nom}, sat clamps to 1.10–1.32 mL with the correction held at the clamp so nothing winds up, $\bar{m}_g$ is the mean of a weighed group's valid results and $\bar{\lambda}$, $\bar{K}$ are the means of the four heads' constants. Fixed calibration applies no update. The defaults are $K_h = 0.50\,\mathrm{mL/g}$ and $\lambda_h = 0.30$ on every head, and $u_{nom} = 1.213\,59\,\mathrm{mL}$ at the 1.250 g target. An update changes only dispenses that start after it, because a dispense reads its head's command at its start. In Manual each head's station sets its command, the correction follows the station and the filters keep filtering, so a return to Auto continues from the station's value.

The plain per-head law integrates a filtered error, and each result adds its error to a command that the corrections from earlier results have already moved. When several results of a head are in flight, the same error is counted once for each of them. With the filter at 1.00 it is the integral form of Section 1.4.5, whose characteristic polynomial with a delay of D results is Equation (1.9), and Chapter 7 applies the same form to pack pressure from each weighed part. The record-based law estimates the head's offset $\hat{d}_h$ from each result less the fill its own command would have given, so a late result is compared with the command that produced it and is never counted twice. That is the intercept form of the EWMA run-to-run controller, Equation (1.7) [12, 13], with a partial move of the command toward the estimate's target, and at rest it gives $\rho u_h + \hat{d}_h = m_{SP}$. A sufficient condition for such a controller to stay stable at any length of metrology delay places the model-mismatch eigenvalues within the unit circle centered at one [11]. Here the mismatch is the scalar ratio of a head's true delivery to the declared ρ per milliliter, which stays between 0.943 and 1.006 over every case as shipped, and the record-based immunity measured below is consistent with that condition rather than proved by it. The immunity rests on the declared response gain. A plant whose dose did not follow its command one for one would test it.

Pause. Try Exercise 9.2 before reading on.

The application keeps three models apart. The plant is everything in Section 9.3. The identified model is the data's dose statistics and the declared response $m = \rho u + d$. The controller's model is that declared response and nothing else, so no correction knows the wear, the viscosity, the drop or the scale's zero. An ideal-information policy, which sets each command before each dispense from the head's true gain and the delivery factor, is a teaching bound that reads the simulator's truth and is reported apart.

9.6 Measured Results

Each number in this section is a measurement on the application's own engine, driven headless at seed 20261021 (the default) through the step sequence the page uses, unless a range over seeds is given. Settings named in a row are applied before the batch begins. The giveaway and every count of accepted vials outside the limits are teaching truth. The exported CSV carries each head's reported results, commands and requested commands beside the true fills, gains and scale zero, whose columns carry a "true" suffix in their names, so a replay of the file through a correction has to leave them out by name.

Table 9.4 Code Verification: Identities Checked and Their Residuals

Identity	Result
Bulk ledger, Eq. (9.6), 13 runs with drips, purges, viscosity and scale drift	worst 3.2×10^{-14} relative
Sum of record fills against the in-vial state, hold-up against $4(V_{tube} + r_h)\rho$	0, 1.9×10^{-16}
Each dispense's command equals the correction standing at its start, 14 runs	36 000 dispenses, 0 differ
Updates citing an unreported result, results used twice, nets not of one record	0 of 10 219, 0 of 12 052, 0
Results held back (latency 10^6 s)	0 command updates
Counts reconcile: in = accepted + rejected + held + awaiting + on line	8667 instants, 0 fail
Replayed stroke order through two stops and purges	0 mismatches
Scale zero drifting 30 mg/h: pump gains	bit-identical
Step halved to 0.05 s, 12 runs	0 dispositions or readings changed, fills within 0.002 mg
Playback at 120× against stepping	identical records

9.6.1 Code Verification

The checks that the model does what Section 9.3 says are listed in Table 9.4. They verify bookkeeping and causality, not physics. The bulk ledger in particular closes by construction of the dispense step, so its residual shows that no event loses or doubles liquid. All 46 checks of the application's own test suite pass, its ten completion checks pass, and every case loads identically when chosen first, when chosen after another case and after Reset.

9.6.2 Eight Cases and a Challenge

The nine scenarios appear in Table 9.5 with the measured contrast each one teaches. All share the reference configuration of Table 9.1 and differ from it only where a row says so.

> **TRY IT IN THE APP** **1 · One Needle Drifting, a Pooled Correction**
>
> Load the scenario, open the Heads view and press Run. Watch head 3: over its last 100 vials it fills 10.4 mg under the target while the other three heads are lifted by 0.9 to 6.7 mg, and 32 underfilled vials are accepted. Press Reset, set Correction to Per-head correction and run again: head 3 ends within 0.7 mg.

The counts of accepted underfills in Table 9.5 are the default seed's and they depend on which stretch of the records each head replays. Over seeds 1 to 3 the head means, the spreads and the accepted and held counts keep the direction of every contrast, but the accepted underfills under the per-head correction rise from 2–4 at the default seed to 27–81 in cases 4, 6 and 8, and in case 1 at seeds 2 and 3 the per-head correction lets 36 and 51 underfills through against 30 and 46 under the pooled one. Its accepted count still rises in case 1 at every seed, from 1324–1530 to 1553–1579. The remedy of case 7 lifts the mean first fill after a restart by 8.7 to 13.9 mg at all four seeds, with the mean first fill moving from between −17.4 and −13.2 mg to between −6.6 and +0.1 mg. The scorecard leads with the accepted vials and the giveaway, and its per-head tiles (the integral of the held reported error, command travel and time at a limit) are diagnostics only. In case 4 the command travel of the four heads is 0.69 to 1.19 mL as shipped against 0.10 to 0.14 mL with the filtered correction, the same order as their true spreads.

Pause. Try Exercise 9.1 before reading on.

9.6.3 Pipeline Length and the Plain Law

Case 3 is where the two per-head laws part. With one group in two weighed and 40 s of latency, a result arrives 50.2 s after its fill, its head has filled 12.8 more vials meanwhile, and 5.9 of them were weighed, so about six results per head are always on their way. The plain law at its default gain swings head 1's command from 81 µL under its nominal 1.2136 mL to 78 µL over it. The sweep of Figure 9.2 varies the latency at four gains. At 0.50 mL/g the plain law holds a head spread of 6.05 mg at 0.5 s and 7.86 mg at 5 s, and oscillates from 20 s on (36.7 mg). At 0.25 mL/g it holds to 20 s (6.95 mg) and

Table 9.5 The Nine Scenarios and What Each One Measures at the Default Seed

#	Configuration and events	Measured contrast
1	1600 vials, head 3 wears 5 % per 1000 doses, pooled correction	head 3 −10.4 mg and heads 1, 2, 4 +6.7, +3.4, +0.9 mg over their last 100 vials, 32 underfills accepted. Per head: head 3 within 0.7 mg, 3 accepted, held 59 to 18
2	4800 vials, all heads wearing 1.5–4.5 %, fixed calibration	head 4 from −16.1 mg (10–15 min) to −62.1 mg (last 20 min), 2499 held, 713 rejected, 1588 accepted. Per head from 15 min: 4519 accepted
3	results 40 s late, 1 group in 2, per-head law	50.2 s from fill to result, 5.9 weighed vials of the head meanwhile. Head SD 47–48 mg, 648 accepted. Record-based 4.6–6.8 mg, 2394 accepted
4	filter 1.00, gain 0.97 mL/g	head SD 6.1–9.2 mg, 2141 accepted, 44 underfills accepted. At 0.30 and 0.50 mL/g: 5.0–7.1 mg, 2365, 2
5	a 6 mPa s lot at 300 s	head 1 −8.8 mg for two minutes, within 0.4 mg from 15 min, SD 5.1 to 9.7 mg, 66 underfills. At 56 vials/min 35, at 1.256 g 14 with giveaway 14.8 to 20.2 mg
6	suck-back 2.8	42 bubbles, head SD 8.2–10.0 mg, 33 underfills. At 1.0: 1 bubble, 2. At 0: 27 drips, 6
7	two 4 min infeed stops, no purge	first fills 10.2–23.7 mg under, 8 rejected, 36 held. With 2 purge shots: −9.0 to +2.2 mg, 2 rejected, 12 held, 19.8 g purged
8	stopper bowl jam 3 min, glove intervention, line keeps running	148 unstoppered, all found and rejected, 2201 accepted. Stopping on a low track: 0 unstoppered, 2354 accepted at 62.0 against 66.2 vials/min. 16 aseptic holds both ways
9	challenge: head 3 wearing, results 20 s late, a 4 min stop. Pooled, filter 1.00, gain 0.90 mL/g, 1 in 4	shipped 1598 accepted, 20.9 mg giveaway, 92 under and 18 over accepted. Suggested: 2329, 13.25 mg, 55.8 vials/min, none outside

fails at 30 s (17.5 mg), and at 0.10 mL/g it stays between 5.81 and 5.99 mg to 60 s. The record-based estimate at 0.50 mL/g stays between 5.81 and 5.98 mg at every latency. The same pattern holds when every group or one in four is weighed, and weighing every group lengthens the pipeline in results per head, so that the plain law at 0.10 mL/g reaches 61.3 mg at 60 s with every group weighed while the record-based estimate holds 5.7 mg. Gain at 0.97 mL/g, one result's error divided by the density, chases the noise even at 0.5 s (13.9 mg).

9.6.4 Four Corrections on the Challenge Line

The policy comparison of Table 9.6 uses the challenge line with the suggested settings (every group weighed, a 3 mg guard band, two purge shots and a 1.249 g target) and changes only the correction, then repeats the comparison with one group in four weighed. Each row runs 12 paired seeds on the validation replay. The seeds are not 12 independent noise realizations, since they start the replay on only 9 distinct segments of 18, while every seed draws its own calibration offsets and scale readings. The fill-distribution tail is reported as the 0.1st percentile of all true fills and as the share of true fills under the label minimum, before any disposition, and medians of 12 are rounded to whole vials. The training replay and the fitted model give the same pass count in every row and the same order, apart from the two near-equal rows.

Three results follow. The plain law at the default gain and the pooled correction oscillate on this line, because every vial weighed with 20 s of latency is a long pipeline, and their head spread of 53 to 61 mg sends 1835 to 1881 vials to the reject chute. Fixed calibration chases no noise, but it leaves head 3's wear and each head's calibration offset in place, so its head spread over the batch is 7.1 mg and 19 % of its fills fall under the label minimum. The record-based estimate and the plain law at 0.10 mL/g are equal within the seed spread, 2325 against 2324 accepted and 13.20 against 13.22 mg of giveaway, and both come within 18 vials and 0.06 mg of the ideal-information bound. On this line the record-based estimate's advantage disappears once the plain law is detuned for the pipeline it meets. Figure 9.3 draws the medians of the table as bars against the challenge bar.

A higher gain buys little here. With a 6 mPa s lot arriving at 600 s on the same line, the mean true fill error in the next two minutes is −11.0 mg under the record-based estimate at 0.50 mL/g

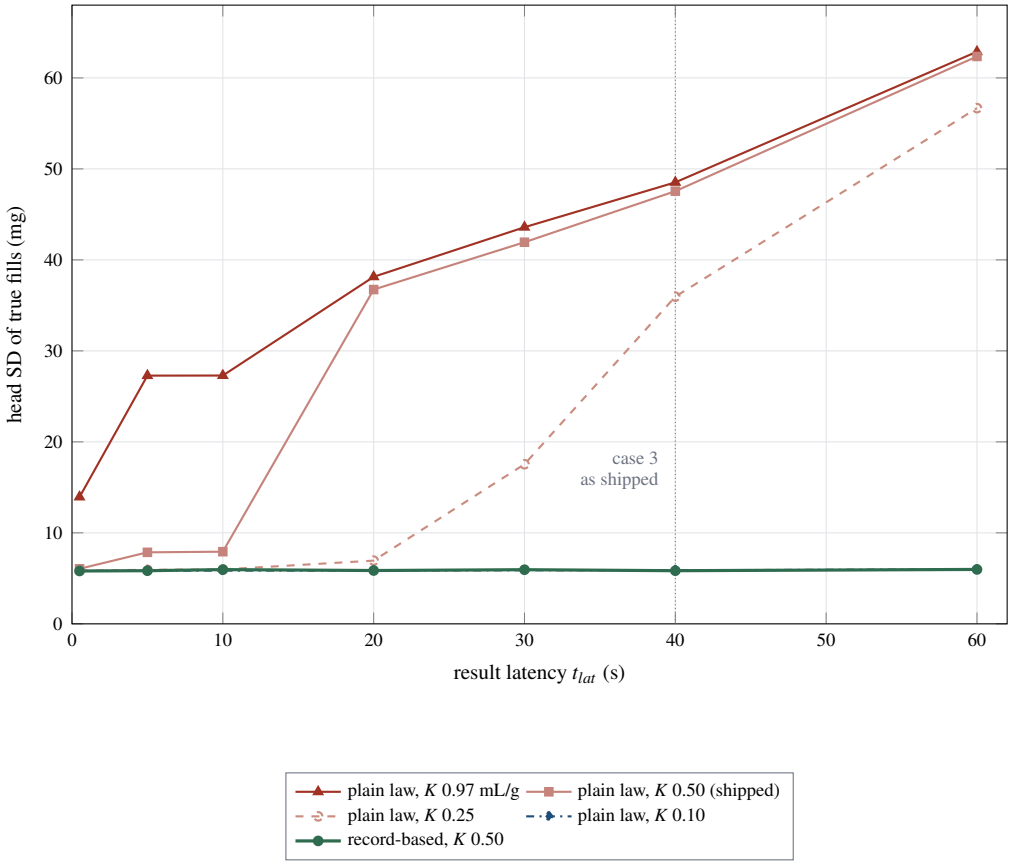


Figure 9.2 Mean head spread of the true fills against the checkweigher's result latency, with the settings of case 3 (one group in two weighed, filter 0.30, default seed). The plain per-head law loses stability at a latency that falls as its gain rises, while the record-based estimate is unaffected. The plain law's gains are marked by triangles and squares on solid lines, open circles on a dashed line and diamonds on a dash-dotted line, in falling order of gain, and the record-based estimate by dots on a thick solid line.

and −12.2 mg under the plain law at 0.10 mL/g, and over the two minutes after that −0.3 mg against +1.6 mg, with 2093 and 2094 vials accepted, while the plain law at 0.25 mL/g already oscillates (672 accepted, medians over 12 seeds). The record-based estimate's value is that its gain can be chosen for the disturbance without knowing the length of the pipeline. Fixed calibration stays 29.5 mg under. The restart recovery is the purge's, not the correction's. With the record-based estimate weighing every group, the first fill of each head after the 4 min stop is −14.6 mg with no purge, −5.1 mg with one shot and −1.2 mg with two (medians over the 12 seeds).

🕮 **Pause.** Try Exercise 9.3 before reading on.

9.6.5 What No Correction Sees

A gross scale whose zero drifts by 30 mg/h in case 1, with head 3's wear removed and the per-head correction on, leaves every pump gain bit-identical, and the correction lowers the commands by 11.1 µL, close to the bias over the density, 11.4 µL. The reported mean of the last 100 vials stays +0.6 mg from the target while the true mean is −9.7 mg, and the accepted underfills rise from 2 to 158. A correction that trusts its scale turns a scale bias into an underfill, and only the reference-weight

Table 9.6 Corrections on the Challenge Line Over 12 Paired Seeds, Validation Replay: Median (Range)

Correction	Pass	Accepted	Giveaway (mg)	Fill rejects	Held	Head SD (mg)	0.1st pct. (mg)	Under LSL (‰)
Every group weighed								
fixed calibration	0	1861 (1717–2124)	14.09	539	0	7.1	−39.9	193 (78–230)
pooled, K 0.50	0	566 (533–588)	22.87	1835	0	61.1	−127.6	427
per head, K 0.50	0	519 (504–551)	23.70	1881	0	53.2	−101.3	437
per head, K 0.10	12	2324 (2313–2352)	13.22	77	0	6.0	−16.1	3.1 (1.7–5.8)
record-based, K 0.50	12	2325 (2307–2348)	13.20	75	0	6.1	−16.2	4.2 (2.9–6.7)
ideal information	12	2342 (2325–2356)	13.26	59	0	5.9	−17.1	3.1 (2.1–4.2)
1 group in 4 weighed								
fixed calibration	0	1793 (1577–2021)	13.45	141	475	7.1	−39.9	193
per head, K 0.10	0	2225 (2137–2301)	12.71	31	146	6.1	−17.7	10.4 (4.6–24.2)
record-based, K 0.50	0	2226 (2146–2301)	12.97	31	144	6.2	−18.7	13.3 (4.2–26.7)
ideal information	0	2292 (2267–2340)	12.98	16	93	5.9	−17.1	3.1 (2.1–4.2)

Every row weighing every group accepts no vial truly outside the limits, so the tail appears as fill rejects. With one group in four, the accepted underfills (teaching truth) are 9 (5–23) under the record-based estimate, 8.5 (5–16) under the plain law at 0.10 mL/g and 6 (3–10) under ideal information, which is why no row passes. The pooled correction and the plain law at 0.50 mL/g accept 1752 and 1861 vials with one group in four and are not listed.

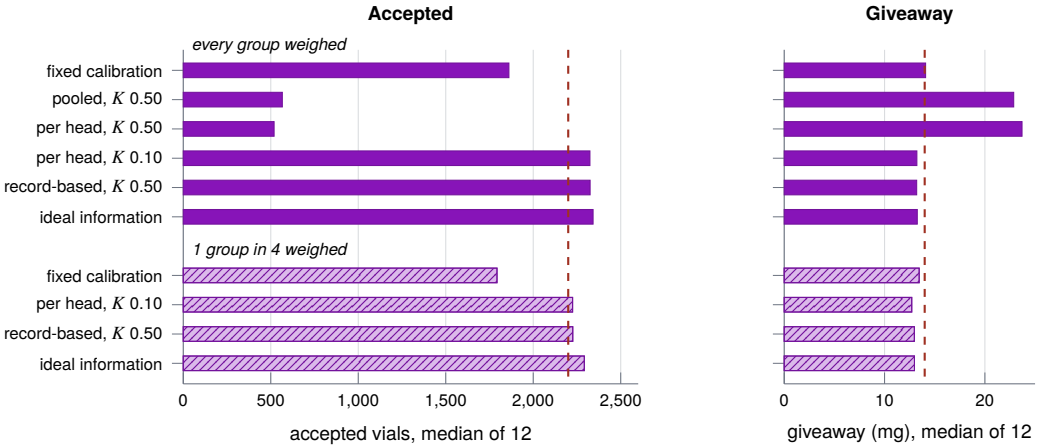


Figure 9.3 The corrections of Table 9.6 on the challenge line, solid with every group weighed and hatched with one group in four. The dashed lines are the challenge bar of 2200 accepted vials and 14 mg of giveaway. The pooled correction and the plain law at 0.50 mL/g lose most vials to the reject chute, while the detuned plain law and the record-based estimate sit 1 vial apart and within 18 vials of the ideal-information bound.

check separates the two. Viscosity and suck-back are the other limits, since they widen the spread that no command can narrow (cases 5 and 6), and they are paid for in line rate, in fill target or in the suck-back setting.

9.6.6 The Challenge

The challenge fills 2400 vials at 72 vials/min with head 3 wearing 4 % per 1000 doses, results 20 s late and a 4 min infeed stop at 15 min. Its target is at least 2200 vials accepted, at most 14 mg of giveaway, at least 55 vials/min with stops included, and no accepted vial truly outside its limits or without a stopper. The bar of 2200 was set so that the verdict holds on the validation replay, on the training replay and on the fitted model at 12 seeds each. As shipped (pooled, filter 1.00, gain

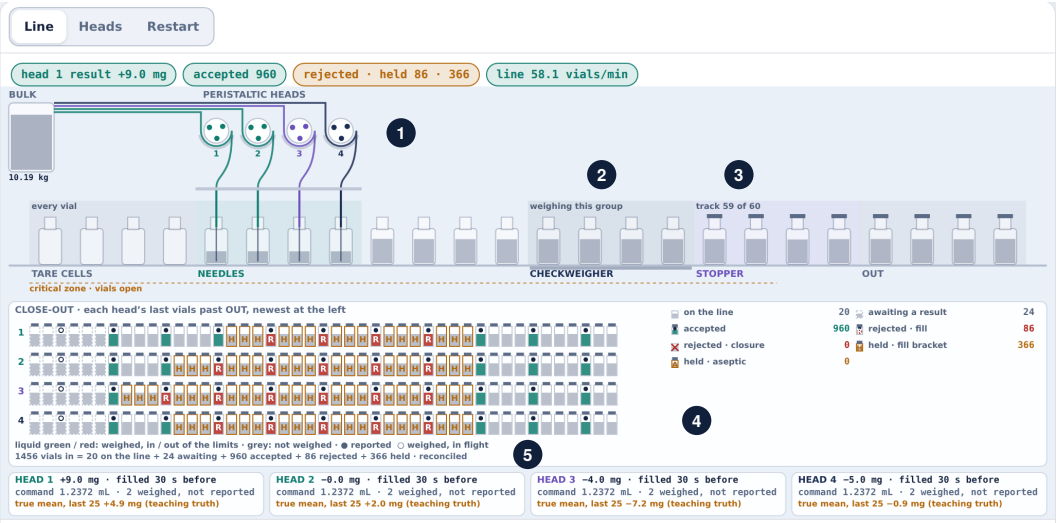


Figure 9.4 The challenge as shipped at $t = 1500$ s in Auto with the pooled correction, cropped to the Line view. The line is filling again after the infeed stop from 903.5 to 1140 s. (1) The bulk vessel feeds the four peristaltic heads, whose needles fill the group after the tare cells. (2) The checkweigher weighs a sampled group gross. (3) The stopper track leads to OUT, where each vial is closed out. (4) The close-out rows show each head's last vials past OUT, many held on their brackets, and reconcile 1456 vials as 20 on the line, 24 awaiting, 960 accepted, 86 rejected and 366 held. (5) Each head's chip shows a last result filled 30 s before its report with two weighed vials in flight and the command of 1.2372 mL at which all four heads stand, and head 3's true mean over its last 25 vials is −7.2 mg (teaching truth).

0.90 mL/g, one group in four, 1.250 g, no purge) the batch fails at all 12 seeds on every noise source, with 1598 accepted, 20.88 mg of giveaway and 92 under and 18 over accepted at the default seed. The suggested settings pass at all 12 on every source, accepting 2307 to 2348 on the validation replay, 2209 to 2350 on the training replay and 2248 to 2279 on the fitted model, at 55.8 vials/min. The settings of Figure 9.4 are the shipped ones at 1500 s.

Each piece of the suggested settings was ablated over six seeds in the application's benchmark file. Without the guard band 2 to 6 underfills pass at every seed, and over 12 seeds a band of 0, 1, 2, 3 and 4 mg passes 1, 5, 11, 12 and 12 of them. Without the purge the batch still passes, with 79 fill rejects against 75 (medians over 12 seeds), so the purge buys rejects rather than the verdict. Weighing one group in four instead fails at every seed even with ideal information, because unweighed underfills pass on their brackets, and filling fuller instead, one in four at 1.256 g, gives 19.40 to 19.95 mg of giveaway and still releases 1 to 4 underfills. Over 252 partial-weighing settings of the record-based and ideal policies, with one group in 2, 3 or 4, targets from 1.245 to 1.265 g and a guard band of 0 or 3 mg, the best passes one of six seeds. On this line the weighing interval, not the correction, is the decision that releases or rejects the tail.

9.6.7 Runtime

A whole challenge batch of 2353 s of line time runs headless in 49 ms, and the suggested batch of 2612 s in 56 ms. In the browser the longest case, 4800 vials in 4187 s, plays in 34.9 s at the top speed of 120 times real time, and every batch takes 12 to 35 s at that speed.

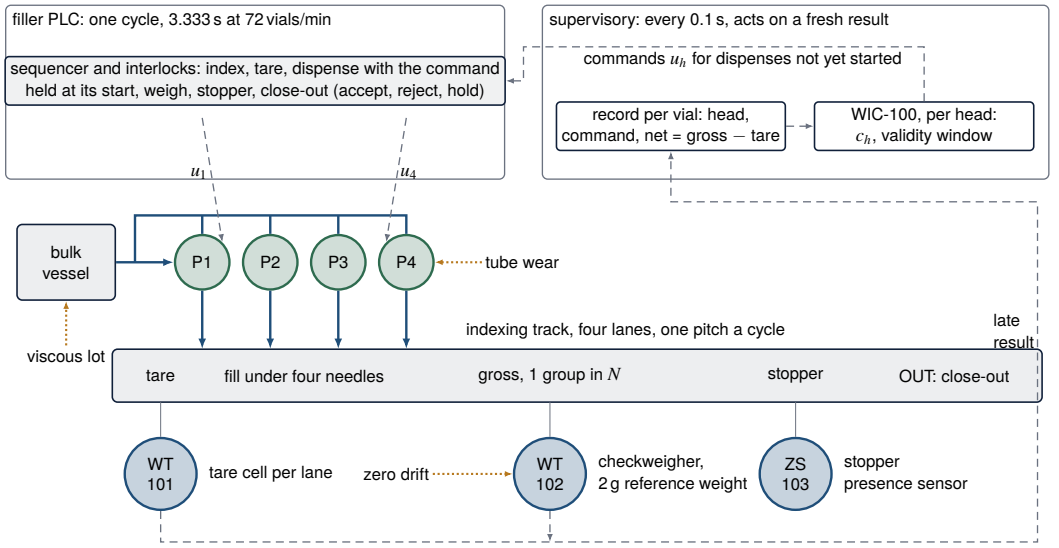


Figure 9.5 The chapter's loop on plant hardware. The filler's sequencer indexes the track, tares each vial, starts each dispense with the command its head holds, weighs one group in N, checks the stopper and closes each vial out. WIC-100 reads only the reported records, each net of its vial's own tare, and moves the commands of dispenses not yet started. Dotted arrows mark disturbances. Tags other than WIC-100 are the book's.

9.7 Technology in Practice

What Measures It

The line of this chapter weighs every vial empty on its lane's tare cell and weighs a sampled group gross on the checkweigher two pitches later, and a presence sensor looks for each stopper. The net of a vial's own gross and tare is the only fill result the controller receives, and it never reads a true fill. Figure 9.5 draws the loop on that hardware. The model's scales add 0.5 mg of noise, round to 1 mg and report each result a latency after the weighing dwell of 3.0 s.

> **TECHNOLOGY Sensor: Tare Cells and a Gross Checkweigher**
>
> A weighing cell turns force into an electrical signal. In a load cell, one or more strain gauges are bonded to a metal structure of precisely known elastic properties, and the strain changes their resistance. Vibration from machinery around a weighed mass puts oscillating forces on the cell [18]. The balance behind the dose records resolved 0.002 g, repeated within ± 0.0005 g and settled in under 120 ms [6], and it cancelled vibration in real time [4]. The model's scale noise of 0.5 mg is that repeatability bound used as a standard deviation. Its zero b_G may drift, and a correction that trusts the scale then turns the bias into an underfill while the reported fills stay on target, which only the reference weight exposes.

What Moves It

Each lane has its own head, tube and needle, so each head needs its own command. In the dose study the pump was calibrated by the pump controller's own function, which computes pump parameters from weight measurements to minimize the dosing error [4]. In the model that calibration leaves each head with the offset $G_h(0)$, and the dispense multiplies the command by the head's gain and adds a replayed deviation, so the rollers, the occlusion and the tube geometry are absent (Section 9.3).

TECHNOLOGY Final Element: Four Peristaltic Heads

A peristaltic pump compresses sterile, disposable tubing with rotating rollers, so the liquid touches only the inside of the tube. The same tubing brings variability through tube wear and temperature [6]. In the dose study, new tubing and a newly assembled circuit introduced calibration drift, so the pump was calibrated five times at the start of each experiment, after a 30 s purge that removed air bubbles [4]. At the needle, a fast suck-back drove product down into a plug at the tip that dried quickly, and a slow one prevented clogging [3]. In the model, wear w_h lowers each head's gain G_h, and purge shots prime the heads after a stop.

Where the Controller Runs

The dose study ran its rig from one controller that served both as a programmable logic controller (PLC) and as a PC-based platform [4], and its compensation worked through the pump controller's standard interface rather than through the mechanics of one pump [6]. A PLC can be the primary controller of a discrete process, and control loops in industry run with cycle times from milliseconds to minutes [19]. WIC-100 executes every 0.1 s but changes a command only at a fresh result, which in case 3 arrives 50.2 s after its fill. Read as plant hardware, the sequence of index, tare, dispense, weighing and close-out, with its interlocks, belongs to the filler's PLC, and the per-head correction from the checkweigher is a supervisory function above the pump controllers.

IN PRACTICE Interventions Go in the Batch Record

The European rules for sterile products ask for an authorized list of qualified interventions, both inherent ones such as aseptic assembly and corrective ones such as clearing a component jam. Each line stoppage or intervention goes into the batch record with its time, its duration and the operators involved, and an intervention that was not qualified is weighed in batch disposition. The filling zone and the stopper bowl lie in grade A, the critical zone [20]. In case 8 of this chapter a glove intervention puts 16 vials on aseptic hold, whatever they weigh (Table 9.5).

What Is Modernizing It

Three developments touch this line. The first takes people out of the critical zone. The European rules ask manufacturers to consider barrier systems, isolators, robotic systems, rapid methods and continuous monitoring [20], and the 2004 guidance notes that robotics can further reduce the risk to the product [1]. The second is the pump and its needle. A mechanical design with eight rollers in place of the industry-standard six makes smaller, more frequent compression points and less pulsation [6], and a nozzle tip with a hydrophobic face and a pressure-resistant barrier gave consistent suck-back heights and about 90 percent of fills within 3 percent of target [2]. The third is software. Compensation learned from the pump's own dose records matched or exceeded that mechanical design, and an ensemble of autoregressive models ran in under 1 s per cycle against about 11 s for a recurrent network [6]. The cycle of this chapter's line lasts 3.333 s at 72 vials/min.

9.8 Limitations

The dose deviations are those of purified water on one pump with one container, so the data cannot validate protein viscosity, temperature, sterility, a multihead line, stoppages or scale bias, and every mechanical law beyond the data is a teaching assumption. The density, the limits, the throughput and the giveaway target are teaching values. Actual limits are set by the product's specification, the pharmacopeial requirements for the content of the container and the site's control strategy. The controller comparison rests on the declared action-response surrogate and on the three independence assumptions of Section 9.3, which the records cannot test, and it is a simulation result, not a prospective experiment. The seeds share replay segments, so their ranges understate the variation of a new noise record, and the accepted-underfill counts in particular are default-seed values with the spread reported in Section 9.6. The dispositions are teaching rules, the aseptic holds are discrete states and nothing contaminates a vial, so a simulated accepted vial is not a released vial and a batch with

no rejects is not evidence of sterility. The protection the guard band gives depends on the declared scale noise, a repeatability bound used as a standard deviation. The validity window, the guard band and the restart check are additions to the plan's correction law, each kept because a measurement on this line justified it, and a line with different noise would need them set again.

OPEN CHALLENGE **Filling Real Formulations, Not Water**

Compensation learned from dose records has been validated on purified water only. Whether it adapts to formulations of other viscosities and rheology, and whether it carries over to other pump architectures, control systems and environments, are open questions that the study itself calls critical for future research [6]. The data are scarce, because the technical details and critical parameters of peristaltic filling are rarely published, and those of viscous, high-concentration formulations least of all [2]. Progress would be open dose records of real formulations on several pumps, against which a record-based correction could be tested beyond its water surrogate.

9.9 Conclusion

The simulated filler fills four lanes from one bulk vessel, replays recorded water doses stroke by stroke per head, and keeps one record per vial that ties its weights to its head and its command, so that a correction sees only what a checkweigher reports and only when it reports it. A pooled correction keeps the average right while one wearing head fills 10.4 mg under the target. A correction that cancels each whole error widens the heads' spread from 5.0–7.1 to 6.1–9.2 mg, as the records' own prediction errors of 4.95, 6.12 and 7.98 mg anticipate. With about six results per head in flight the plain per-head law oscillates to a 48 mg spread, while the record-based estimate holds 4.6 to 6.8 mg there and a mean head spread of 5.8 to 6.0 mg at every latency tested. A plain law detuned to 0.10 mL/g matches it on the challenge line, 2324 against 2325 accepted, and the weighing interval, not the correction, decides whether underfills are released. The comparison is a simulation under declared surrogate and independence assumptions. A fuller model would add a pump map with a measured command response, a viscous formulation's own dose records, weighing after closure with stopper-mass compensation, and a prospective test of the corrections on a filler. The simulator shows that a correction has to know which vial a result describes before it can act on it.

KEY POINTS

- Each weight describes a vial filled tens of seconds earlier, by a pump whose command later corrections have already changed.
- A pooled correction keeps the average right while one wearing head fills 10.4 mg under the target.
- With about six results per head in flight the plain per-head law oscillates to a 48 mg spread, while the record-based estimate holds 4.6 to 6.8 mg there.
- A plain law detuned to 0.10 mL/g matches it on the challenge line, 2324 against 2325 accepted.
- The weighing interval, not the correction, decides whether underfills are released.

Where else this applies. Volume 5, Chapter 6 holds tablet weight on a rotary press, and Chapter 7 applies the same integral form to pack pressure from each weighed part (Section 1.4.5).

References

[1] U.S. Food and Drug Administration, "Guidance for industry: Sterile drug products produced by aseptic processing — current good manufacturing practice," Docket FDA-2003-D-0145, Oct. 2004. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/sterile-drug-products-produced-aseptic-processing-current-good-manufacturing-practice

[2] W. Shieu, D. Lamar, O. B. Stauch, and Y.-F. Maa, "Filling of high-concentration monoclonal antibody formulations: Investigating underlying mechanisms that affect precision of low-volume fill by peristaltic pump," *PDA Journal of Pharmaceutical Science and Technology*, vol. 70, no. 2, pp. 143–156, 2016.

[3] S. Hanslip, K. G. Desai, M. Palmer, S. Bell, P. Schofield, P. Varma, F. Roche, J. D. Colandene, T. M. Smith, and D. P. Nesta, "Syringe filling of high-concentration mAb formulation: Slow suck-back pump speed prevented filling needle clogging," *Journal of Pharmaceutical Sciences*, vol. 106, no. 12, pp. 3651–3653, 2017.

[4] D. Privitera, A. Mecocci, and S. Bartolini, "Extensive dataset for peristaltic pump accuracy enhancement in pharmaceutical environments," *Scientific Data*, vol. 12, no. 1, p. 1618, 2025.

[5] ——, "DosePI: A comprehensive dataset for peristaltic pump accuracy enhancement in pharmaceutical environments," Dataset, figshare, licensed CC BY 4.0, 2025.

[6] ——, "Ensemble methods for peristaltic pump accuracy enhancement," *IEEE Access*, vol. 13, pp. 125 157–125 179, 2025.

[7] L. de Mare, S. Velut, P. Hagander, and M. Akesson, "Feedback control of flow rate from a peristaltic pump using balance measurements," in *2001 European Control Conference (ECC)*, 2001, pp. 2164–2168.

[8] M. Meller, M. Niedźwiecki, and P. Pietrzak, "Adaptive filtering approach to dynamic weighing: A checkweigher case study," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 5927–5932, 2014.

[9] J. F. MacGregor, "A different view of the funnel experiment," *Journal of Quality Technology*, vol. 22, no. 4, pp. 255–259, 1990.

[10] G. Box and T. Kramer, "Statistical process monitoring and feedback adjustment—a discussion," *Technometrics*, vol. 34, no. 3, pp. 251–267, 1992.

[11] R. P. Good and S. J. Qin, "On the stability of MIMO EWMA run-to-run controllers with metrology delay," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 1, pp. 78–86, 2006.

[12] A. Ingolfsson and E. Sachs, "Stability and sensitivity of an EWMA controller," *Journal of Quality Technology*, vol. 25, no. 4, pp. 271–287, 1993.

[13] E. Sachs, A. Hu, and A. Ingolfsson, "Run by run process control: Combining SPC and feedback control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 8, no. 1, pp. 26–43, 1995.

[14] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.

[15] G. E. P. Box and G. M. Jenkins, "Some statistical aspects of adaptive optimization and control," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 24, no. 2, pp. 297–331, 1962.

[16] E. Del Castillo and A. M. Hurwitz, "Run-to-run process control: Literature review and extensions," *Journal of Quality Technology*, vol. 29, no. 2, pp. 184–196, 1997.

[17] W. G. Hunter and C. P. Kartha, "Determining the most profitable target value for a production process," *Journal of Quality Technology*, vol. 9, no. 4, pp. 176–181, 1977.

[18] T. R. Kuphaldt, *Lessons in Industrial Instrumentation*, version 2.33 ed. Open textbook, Creative Commons Attribution 4.0 license, ibiblio.org/kuphaldt/socratic/sinst, 2024.

[19] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[20] European Commission, "EudraLex, volume 4, EU guidelines for good manufacturing practice for medicinal products for human and veterinary use, annex 1: Manufacture of sterile medicinal products," C(2022) 5938 final, Brussels, Aug. 2022. [Online]. Available: https://health.ec.europa.eu/system/files/2022-08/20220825_gmp-an1_en_0.pdf

9.10 Exercises

9.1 ⌕ Concept check · 2 min

In Scenario 1 the pooled correction moves all four heads by the filtered error of each weighed group's average, while head 3's tube wears. What do the true fills of the four heads show over each head's last 100 vials?

(a) All four heads fill within about a milligram of the target, since the group average is held there

(b) Head 3 fills under the target, and the other three heads are pushed over it

(c) Head 3 is brought back to the target, and the other three heads stay where they were calibrated

(d) All four heads fall under the target, because only one group in four is weighed

9.2 Concept check · 2 min

In Scenario 3 each result arrives 40 s after its vial is weighed, and about six results per head are always in flight. The plain per-head law oscillates, and the heads' true fills spread 47 to 48 mg. What is the cause?

(a) The late results carry more scale noise, so the correction chases noise it would otherwise filter out

(b) The validity window rejects most late results, so the heads run without correction between them

(c) Each result's error is added to a command that earlier results have already corrected for the same error

(d) The step limit of one percent of the nominal command is too small for the correction to keep up

9.3 Concept check · 1 min

In Scenario 7 the line stops twice for 4 min and restarts with the commands it had. The first fill of each head after a restart is 10.2 to 23.7 mg under the target. What restores those first fills?

(a) The per-head correction, as soon as the results of the restart group are reported

(b) A higher correction gain, so that the first results move the commands further

(c) A reject guard band, which keeps the short vials inside the fill limits

(d) Purge shots before the first vial, which replace the fallen drop and work the tube

9.4 Theory and calculation · 6 min

The line fills four vials per cycle. Each cycle at rate R (vials per minute) lasts $T_c = 60 \cdot 4/R$, of which $t_{idx} = 0.9\,\mathrm{s}$ is the index move and $t_{set} = 0.3\,\mathrm{s}$ a settle, and the dose statistics are referred to $R_{ref} = 60\,\mathrm{vials/min}$ (Equations (9.1) and (9.3), Table 9.2).

(a) At 72 vials/min, find T_c, the fill window T_f, the pump-speed ratio ω/ω_{ref} and the spread factor f_σ on water. *Checkpoint:* 3.333 s, 2.133 s, 1.3125 and 1.146.

(b) In a weighing cycle the dwell after the index lasts at least $t_w = 3.0\,\mathrm{s}$. Find the line rate before stops when every group is weighed. *Checkpoint:* 61.5 vials/min.

(c) A lot at 6 mPa s replaces water at 1.0 mPa s. Find the change in every head's delivery and the factor on the spread of its doses. *Checkpoint:* −2.15 % and 1.565.

9.5 Theory and calculation · 6 min

Equation (9.4) models one head's dose deviation as $e_k = p_k + r_{k-1} - r_k$, with independent pump terms p_k of standard deviation $\sigma_p = 2.85\,\mu\mathrm{L}$ and drops of standard deviation $\sigma_r = 3.17\,\mu\mathrm{L}$.

(a) Derive the variance of e_k and its lag-1 autocorrelation, and evaluate the standard deviation and ρ_1. *Checkpoint:* 5.31 µL and −0.356.

(b) Convert the standard deviation to milligrams of water at $998.2\,\mathrm{kg/m^3}$ and compare it with Table 9.3. *Checkpoint:* 5.30 mg against 5.31 mg on training.

(c) A predictor that always predicts the target meets both the spread within a segment and the spread of the segment means. Treat the two as independent and estimate its RMSE on the training third. *Checkpoint:* 6.86 mg against 6.82 mg in the table.

9.6 Theory and calculation · 8 min

With the report filter at 1.00 the plain per-head law adds $K_h(m_{SP} - \hat{m}_r)$ to the head's command after each result r. A head delivers $\xi\rho$ grams per milliliter of command, with $\rho = 1.030\,\mathrm{g/mL}$, and each result reaches the controller after D later results of the same head have already moved the command.

(a) Show that the loop gain per result, $\lambda\xi$ in Equation (1.9), is $K_h\rho\xi$, and evaluate $K_h\rho$ for K_h of 0.97, 0.50, 0.25 and 0.10 mL/g. *Checkpoint:* 0.999, 0.515, 0.258 and 0.103.

(b) Put $z = e^{i\omega}$ into Equation (1.9) and show that the loop is stable for a loop gain below $g^* = 2\sin\big(\pi/(4D+2)\big)$. Check $D = 0$ to 3 against Section 1.4.5. *Checkpoint:* 2, 1, 0.618 and 0.445.

(c) With $D = 6$ and $\xi = 1$, which of the four gains are stable? *Checkpoint:* $g^* = 0.241$: only 0.10 mL/g.

9.7 App challenge · 8 min

Open Scenario 1, "1 · One needle drifting, a pooled correction", at the default seed, and run it to the end at 120× (120 s of line time per wall second, about 12 s) on the Heads view.

(a) Read each head's true mean over its last 100 vials on the "last 100" line of its panel, and the tile Accepted outside the limits (true). Tolerance 0.1 mg. *Checkpoint:* Head 3 −10.4 mg, heads 1, 2 and 4 +6.7, +3.4 and +0.9 mg, 32 under.

(b) Press Reset, set Correction on the WIC-100 card to Per-head correction, and run to the end. Read head 3's true mean, the underfills accepted and the fill-bracket holds on the tile Rejected · held. *Checkpoint:* −0.7 mg, 3 under, 18 held against 59.

(c) Read the four commands on the head panels at the end of each run and explain the difference. *Checkpoint:* Pooled: 1.2186 mL on every head. Per head: 1.2084 to 1.2333 mL.

9.8 App challenge · 8 min

Open Scenario 3, "3 · A late checkweigher result", at the default seed, on the Heads view at 120× (about 19 s per batch).

(a) Run it to the end as shipped. Read the SD on each head's panel and the lead tile Accepted vials · giveaway. Tolerance 0.5 mg. *Checkpoint:* SD 46.7 to 48.1 mg, 648 of 2400 accepted.

(b) Press Reset, set Correction on the WIC-100 card to Per-head, record-based, and run to the end. Read the same values. *Checkpoint:* SD 4.6 to 6.8 mg, 2394 accepted.

(c) Press the reload button beside the scenario list, set the correction gain on all four Head cards to 0.10 mL/g, and run to the end. Compare with part (b), and say what the lower gain costs. *Checkpoint:* 2394 accepted, SD 4.7 to 6.8 mg.

9.9 App challenge · 9 min

Open the challenge, "Challenge · Least giveaway, nothing out of limits", at the default seed, and run at 120× (about 20 s per batch).

(a) Run it to the end as shipped. Read the lead tile, Accepted outside the limits (true) and Throughput. *Checkpoint:* 1598 of 2400 at 20.9 mg, 92 under and 18 over, 62.0 vials/min.

(b) Press Reveal suggested tuning, then Reset, and run to the end. Read the same tiles. Tolerance 0.1 mg on the giveaway. *Checkpoint:* 2329 of 2400 at 13.3 mg, none outside, 55.8 vials/min, target met.

(c) Open Advanced on the WIC-100 card, set Reject guard band to 0 mg, press Reset and run again. Read the same tiles and explain the result. *Checkpoint:* 2393 accepted at 12.9 mg, 6 under, target not met.

9.10 Application area · 4 min

The label states 1.20 mL of a product at 1.030 g/mL, and the upper fill limit is 104 % of the label minimum.

(a) Find the label minimum, the upper limit, and the margin from the 1.250 g target to the minimum in standard deviations of a replayed dose (5.7 mg). *Checkpoint:* 1.236 g, 1.2854 g, 14 mg or 2.5 standard deviations.

(b) The shipped challenge gives away 20.88 mg per accepted vial and the suggested settings 13.25 mg. Express the difference for a batch of 2400 vials, and name what the line pays for it. *Checkpoint:* 7.63 mg per vial, about 18.3 g per batch, for 55.8 instead of 62.0 vials/min and about 10 g of purge.

9.11 Application area · 3 min

A quality engineer proposes dropping the reference-weight check, because the per-head correction holds the reported mean of the last 100 vials within 1 mg of the target. In the chapter's run of Scenario 1, with head 3's wear removed, the gross scale's zero drifts by 30 mg/h over the 1407 s batch.

(a) Find the zero offset at the end of the batch and the command change a correction that trusts the scale would make. *Checkpoint:* 11.7 mg and 11.4 µL lower.

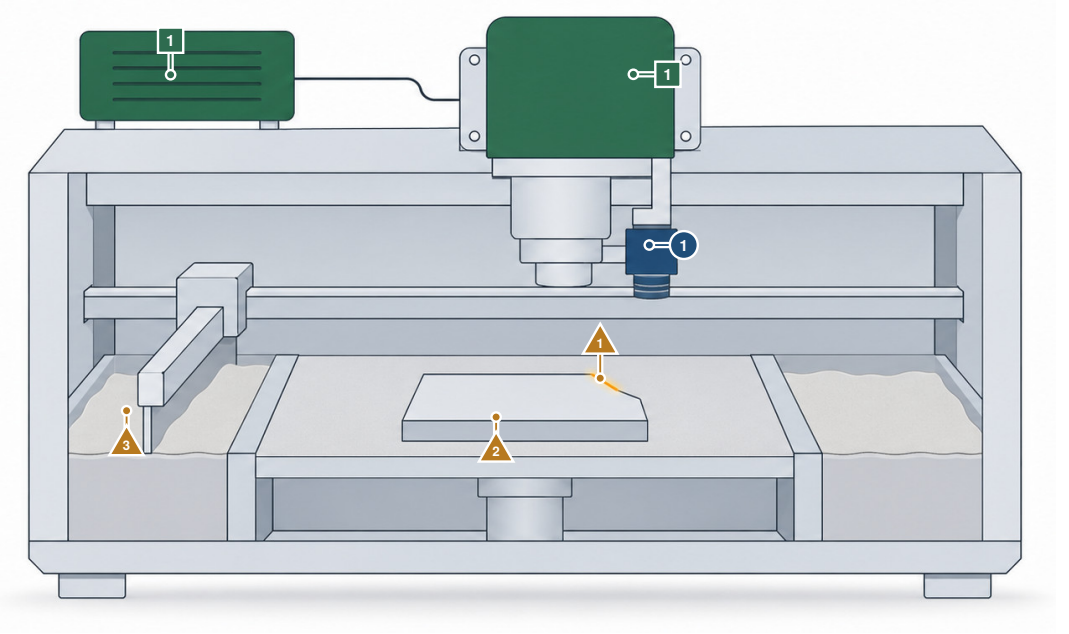
(b) Using the chapter's result, say what the proposal would hide. *Checkpoint:* Reported +0.6 mg against a true −9.7 mg; accepted underfills from 2 to 158.

9.12 Review the agent · 5 min

An AI assistant reviews the challenge line and writes: "Weighing every group costs 6 vials per minute. Keep one group in four, the record-based estimate, the 3 mg guard band and the two purge shots, and raise the target from 1.249 to 1.256 g. The giveaway rises to about 20 mg, which the plant accepts for the throughput. A mean 7 mg higher is more than twice the guard band, so the tail clears the label minimum and no underfilled vial is released." Find the flaw, and say how you would check it in the app.

CHAPTER 10

Learning From the Last Layer: A Powder Bed Fusion Simulator for Iterative Learning Control



The controller moves 1 the laser and the scan head that set the power along the path

The controller reads 1 the photodiode that watches the melt pool

What disturbs it 1 a sharp corner of the outline 2 heat that builds up in the part 3 a layer spread short of powder

The laser traces the same outline on every layer, so the pool overheats at the same corner each time and a controller can learn where. Illustration, not to scale.

IN PLAIN TERMS **A Laser That Learns From the Last Layer**

The main idea. A laser melts metal powder one thin layer at a time. It traces the same outline on every layer. At each sharp corner the small pool of molten metal runs too hot and too wide, in the same places every time.

The challenge. A controller that watches the pool and turns the power down reacts too late, because a corner is over in about a thousandth of a second. Remembering where the corners were on the last layer lets it lower the power just before each one.

Why it matters. A pool that is too wide or too narrow leaves tiny holes in metal parts made for aircraft and medical use.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history and the first fact in Section 10.2, the second fact and both try-its in Section 10.5, and the key points
Study	30 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 10.5), Technology in Practice (Section 10.6), Limitations (Section 10.7) and the two open challenges

The control question A powder bed fusion laser overheats the melt pool at the same corners on every layer, faster than a loop with sensor and laser latency can react. What can feedback do about it, what does learning from the last layer add, and what goes wrong when the layers stop repeating?

Learning objectives

- Explain why a melt pool loop with 300 µs of latency removes the slow heat drift but not the corner overheating, and why more gain makes the build worse.
- Compute the melt pool width the model supports at a corner and the heat that accumulates in a part from layer to layer.
- Explain how the learning update with a lead and a Q-filter converges, and why it diverges without the filter.
- Compare keeping and forgetting a learned profile when the outline changes, and judge how much feedback to keep under the learning.

Prerequisites Sections 1.4.6 (the learning update and its convergence factor), 1.6.2 (the PID of the feedback loop) and 1.6.1 (seed, speed and Reset), and Chapter 7 for a correction made between cycles.

The app https://apmonitor.com/apps/laser-powder-bed/

See also Volume 3, Chapter 6 for feedback that acts after the disturbance has passed and the feedforward that replaces it, and Volume 3, Chapter 2 for a loop structure that decides performance before any tuning.

10.1 Introduction

Feedback is taught as the remedy for disturbances, and for most disturbances it is. A disturbance that arrives faster than the loop can respond is the exception, and a disturbance that arrives at the same place in every repetition of a task is the exception that can be exploited. Iterative learning control was formulated for exactly that case: a robot repeating the same motion corrects its input from the error it saw on the previous repetition, and the error that repeats shrinks from one trial to the next [1, 2]. The theory that decides whether the learning converges, and what a low-pass Q-filter does to that answer, is mature [3]. Section 1.4.6 sets out the update and its convergence factor on the batch kettle.

Laser powder bed fusion is repetitive by construction. Each layer the laser traces the part cross-section, and features such as corners, thin walls and edges built over loose powder heat the melt pool at the same position on every layer. Too narrow a pool leaves unfused powder between tracks, too wide a pool becomes a vapor keyhole that collapses into pores, and both defects stay in the part [4, 5]. Coaxial photodiodes and cameras watch the melt pool, and closed-loop power control on their signal was demonstrated more than a decade ago [6]. A corner, however, lasts about a millisecond, which is comparable to the latency of the sensor, the filter and the laser modulation together.

A common assumption, which the simulator is designed to test, is that feedback fixes every repeatable error if tuned hard enough. The broader class is control of repetitive processes: batch reactors run the same recipe, wafer scanners repeat the same exposure, rolling mills pass the same product, and additive machines repeat nearly the same path, and in each the choice between reacting and remembering decides what is achievable [2]. Chapter 7 corrects the pack pressure of a molding machine from each weighed part, the same choice made one number per cycle. Spatial learning controllers indexed by position rather than time have been proposed specifically for additive manufacturing [7], and layer-to-layer control of interlayer temperature in powder bed fusion has been demonstrated more recently [8].

This chapter sets out a powder bed fusion simulator in which feedback, learning, the learning gain, the Q-filter and the response to a change in part geometry can be compared on the same build. Iterative learning control, its Q-filter and melt pool feedback are established methods. What the simulator adds is a contour-per-layer model whose geometry effects repeat exactly along arc length, with a learning controller indexed the same way. The scenarios move from a fixed-power build through a learned one to a build that changes shape.

Section 10.2 reviews melt pool monitoring and why it led to control. Section 10.3 sets down the melt pool, the geometry along the path, the heat buildup and the photodiode, with the assumptions. Section 10.4 gives the feedback loop on the photodiode and the learning controller with its lead and Q-filter. Section 10.5 reports the measured builds: what feedback can and cannot do, what learning does, why the Q-filter matters, overhangs, heat and broken repetition, and the graded challenge. Section 10.6 sets the loops on machine hardware, Section 10.7 states the limitations, Section 10.8 summarizes, and the exercises follow in Section 10.9.

10.2 Melt Pool Monitoring and Control

Powder bed fusion began as a prototyping process and became a production one for metal parts over the 2000s, and with that change the question of process quality moved from appearance to porosity, microstructure and properties [5]. The melt pool is where those are decided. Its width and depth follow the energy delivered per unit length of track, and above a threshold the pool changes from conduction mode to keyhole mode, with a sharp increase in depth and in porosity [4].

Monitoring grew first. How a coaxial photodiode and camera came to close a loop on laser power is told in the history box below. Quality control schemes for powder bed machines built on the same sensors [9], and reviews of in-situ monitoring later cataloged photodiodes, pyrometers, cameras, acoustic sensors and layer imaging, together with the difficulty of relating a sensor signal to a defect [10]. The same signals define the practical limit of melt pool feedback, because corners and overhangs change the pool within a time comparable to the sensing, filtering and actuation delay.

> ⧖ HISTORY **2010 to 2012: A Photodiode Finds the Hot Corners**
>
> A university group at Leuven built a control system for laser powder bed fusion around the melt pool. A camera and a photodiode looked down the laser's own optical path. The camera gave the pool's shape, the photodiode signal followed the pool's area, and that signal was fed back to the laser power, which improved the surface roughness of the parts. The group later saw the photodiode reading rise where the pool overheated at acute corners of the scan path. By mapping the signal against the beam position on each layer, it found deformation from thermal stress and hot zones over overhangs. A 2015 review of what real-time control of the process still needs tells this sequence first among the control schemes it surveyed [11].

Qualification drives the interest in control. Metal parts for aerospace and medical use must be shown to be free of critical defects, and doing so part by part through inspection is costly, which has made process monitoring and control central to qualification strategies [12]. Iterative learning control fits the problem because the path repeats, and because a learned correction is a record of what the machine did to each feature. Spatial formulations indexed by position for additive processes [7] and layer-to-layer feedback on part temperature [8] are two recent answers, and the simulator described here uses the same idea on melt pool width along a contour.

> ✦ DID YOU KNOW? **Powder Holds the Heat In**
>
> Loose metal powder conducts heat far worse than solid metal, potentially less than 1 percent as well. Near an overhang or a fine feature the powder keeps heat in the melt pool, which overheats, turns unstable and grows [13]. The model's downskin edge over loose powder adds 35 percent to the width the pool supports (Table 10.2).

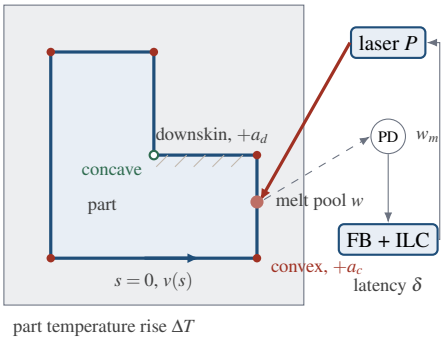


Figure 10.1 One contour pass of laser powder bed fusion. The laser of power P traces the part outline from $s = 0$ at scan speed $v(s)$, slowing at each corner. Convex corners (filled dots) add heat a_c, the concave corner (open dot) removes it, and the downskin edge over loose powder (hatched) adds a_d. A coaxial photodiode PD reads the melt pool width w as w_m, and the feedback and learning controllers set the power after a latency δ. The part warms by ΔT as layers accumulate.

Table 10.1 Variables at the Nominal Operating Point

Symbol	Role	Nominal	Range
w	CV, melt pool width	100 µm	SP 70–130
P	MV, laser power	200 W	0–370
$P_{ff}(s)$	MV, learned correction	0 W	−200 to 170
w, s, E	states	—	—
ΔT	per-layer state, part	0 K	0–150
$g(s)$	DV, geometry gain	1	0.93–1.44
$v(s)$	DV, scan speed	800 mm/s	400–800
σ_{sp}	DV, spatter (1σ)	3 %	0–20
ν	DV, photodiode noise	2 µm	0–20
f_{pow}	DV, powder layer	1	0.85 once

10.3 Process Model

The laser traces the outline of a 10 mm square part at 800 mm/s and a nominal 200 W, one pass per layer, with the melt pool width targeted at 100 µm. A pass takes about 52 ms. The recoat between layers is not simulated in time, so the clock counts scanning time only. The arrangement is shown in Figure 10.1.

The variables are listed in Table 10.1. Three states are integrated: the melt pool width w, the arc position s along the outline and the energy E delivered on the current layer. The part temperature rise, the learned profile and the layer scores change once per layer.

10.3.1 Melt Pool Width

The width follows a first-order lag behind the width the local conditions support, Equation (10.1).

$$\tau_p \frac{\mathrm{d}w}{\mathrm{d}t} = w_{ss} - w$$
$$w_{ss} = w_0\, g(s)\,(1 + \alpha \Delta T)\, f_{pow} \left(\frac{P}{P_0}\right)^n \left(\frac{v_0}{v(s)}\right)^m \tag{10.1}$$

with $\tau_p = 0.3$ ms, $w_0 = 100$ µm at $P_0 = 200$ W on bulk metal, $n = 0.7$, $m = 0.35$ and $\alpha = 1.5 \times 10^{-3}\,\mathrm{K}^{-1}$. The plant gain at nominal is $\hat{G} = n w_0 / P_0 = 0.35$ µm/W. Each pass starts with the pool already formed, as if the laser were switched on over a lead-in outside the part.

Table 10.2 Process Parameters, With Hidden Ones Marked

Symbol	Meaning	Value	Class
τ_p	melt pool time constant	0.30 ms	visible
w_0, P_0	nominal width, power	100 µm, 200 W	visible
n, m	power, speed exponents	0.70, 0.35	hidden
v_0, ρ	scan speed, corner fraction	800 mm/s, 0.50	visible
a (convex, concave)	corner heat, chill	15, 12 %	visible
a_d	downskin heat	35 %	visible
α	width per kelvin	1.5×10^{-3} K^{-1}	hidden
c_E, κ	heat per joule, kept	1.36 K/J, 0.90	visible
P_{max}	laser power limit	370 W	visible
δ	sensor and control latency	300 µs	visible
—	defect window	85–125 µm	visible

10.3.2 Geometry Along the Path

The outline is parameterized by arc length s and the effects of geometry are fixed functions of s, identical on every layer of the same outline, Equations (10.2) and (10.3).

$$g(s) = 1 + \sum_c \pm a\, b(s - s_c) + a_d \mathbb{1}_d(s) \tag{10.2}$$

$$v(s) = v_0 \left[1 - \tfrac{1-\rho}{2} \left(1 + \cos \frac{\pi (s - s_c)}{d_c} \right) \right] \tag{10.3}$$

The corner bump b rises as a Gaussian of width 0.3 mm into the corner and decays exponentially over 1.2 mm after it. Convex corners add $a = 15\,\%$ and concave corners remove 12 %, the downskin edge adds $a_d = 35\,\%$ with 0.3 mm smooth edges, and the mirrors slow the spot to $\rho = 0.5$ of full speed within $d_c = 0.5$ mm of every corner, Equation (10.3) applying for $|s - s_c| < d_c$. The square has four convex corners, and the L-shaped bracket has five convex corners, one concave corner and a 5 mm downskin edge, both with a 40 mm perimeter.

10.3.3 Heat Buildup and the Photodiode

Between passes the part keeps a fraction $\kappa = 0.9$ of its temperature rise and gains $c_E = 1.36\,\mathrm{K/J}$ per joule delivered, Equation (10.4).

$$\Delta T_{k+1} = \kappa \Delta T_k + c_E E_k \tag{10.4}$$

About 10 J per pass at 200 W settles a small part between 100 and 140 K warmer depending on the power. The photodiode reads the width with multiplicative spatter and additive noise, $w_m = w(1 + \sigma_{sp}\xi) + \nu$, where ξ is colored noise with a 0.2 ms time constant. Width outside the window from 85 to 125 µm is counted as defective, lack of fusion below and keyhole above. That defect share is a proxy for porosity and not a measurement of it.

Table 10.2 collects the parameter values.

10.3.4 Assumptions

One contour is scanned per layer. The interior hatch, which takes most of the energy and has its own repeating turnarounds, is not simulated.

The melt pool is a first-order lag with a power-law steady state. The sharp transition to keyhole mode is represented only by the upper width threshold [4].

Geometry and heat effects are fixed profiles along the path rather than the result of three-dimensional conduction, so they repeat exactly, which real layers only approximately do [5].

The photodiode is a width sensor with noise. Real signals depend on emissivity, plume and optics and need calibration before they can be read as a width [10].

The recoat takes no time and the part cools by a fixed fraction per layer.

10.4 Implemented Control

The engine of Section 1.6.1 integrates the melt pool with a fixed 20 µs Runge-Kutta step of fourth order, and the controllers execute at 10 kHz.

10.4.1 Feedback on the Photodiode

The feedback loop is a PI controller on the filtered photodiode width, executed every $\Delta t_c = 0.1\,\text{ms}$ in the velocity form of Section 1.6.2, Equation (10.5).

$$\begin{aligned} e_k &= w_{SP} - \mathscr{F}_{\tau_f}[w_m]_k \\ \Sigma_k &= \Sigma_{k-1} + K_c\,(e_k - e_{k-1}) \\ c_k &= \Sigma_k + I_{k-1} \\ u_k &= \text{clamp}(c_k, -P_b, P_{max} - P_b) \\ I_k &= I_{k-1} + \frac{K_c}{\tau_I}\,e_k\,\Delta t_c + \frac{\Delta t_c}{\tau_t}\,(u_k - c_k) \end{aligned} \tag{10.5}$$

where $P_b = P_0 + P_{ff}(s)$ is the power the learned profile asks for and the commanded power is $P = P_b + u_k$. The running sum Σ_k, the demanded output c_k and the integral I_k are the S_k, v_k and I_k of Equations (1.15) to (1.17), with no derivative and no bias. The loop acts to cut power when the pool is too wide. The defaults are $K_c = 1.0\,\text{W}/\mu\text{m}$, $\tau_I = 1\,\text{ms}$ and a photodiode filter $\tau_f = 0.2\,\text{ms}$. The clamp keeps the total power between zero and P_{max} around whatever the learned profile asks for, and while it acts the back-calculation, with tracking time $\tau_t = \tau_I$, removes a tenth of the clipped excess from I each scan. The command is held for the control period and reaches the melt pool after the 300 µs latency. In manual mode the operator sets the total power, and transfer in either direction is bumpless. Each run starts with the photodiode filter at the initial width and the delay line holding the nominal power, so the first scan commands 200 W and a power chosen before the run reaches the pool only after the latency.

10.4.2 Learning Along the Outline

The learned profile P_{ff} is stored in 0.2 mm bins along the outline. During a pass the controller averages the measured width and the feedback output in each bin. After the pass the profile is updated by Equation (10.6), the update of Equation (1.10) indexed by arc length instead of time, with the feedback effort added.

$$P_{ff}^{k+1}(s) = Q\left[P_{ff}^{k}(s) + \bar{u}_{fb}^{k}(s) + \frac{\gamma}{\hat{G}}\,e_k(s+\lambda)\right] \tag{10.6}$$

where e_k is the error of the bin-averaged measured width, $\gamma = 0.5$ the learning gain, $\lambda = 0.6\,\text{mm}$ the lead that times each correction ahead of the melt pool lag, the latency and the filter, and $\bar{u}_{fb}$ the average feedback output in each bin, included only when feedback is running. The lead is a whole number of bins, and its setting steps in 0.2 mm. Q is a zero-phase first-order low-pass applied forward and backward around the closed outline with length $\ell_Q = 0.2\,\text{mm}$, so that $a = \exp(-0.2\,\text{mm}/\ell_Q)$ per bin. The profile is clamped so that the feedforward power stays between zero and P_{max}. When the outline changes between layers, the profile is either kept or reset to zero. Figure 10.2 draws the update as it runs between two layers.

The lead of 0.6 mm is the sum of the melt pool time constant, the latency and the filter at full scan speed, as if all three were pure delays. The condition for monotone convergence at a spatial frequency ω is Equation (10.7), the factor of Equation (1.11) written along the path.

$$\left|Q(\omega)\left[1 - \frac{\gamma}{\hat{G}}\,G(\omega)\,e^{j\omega\lambda}\right]\right| < 1 \tag{10.7}$$

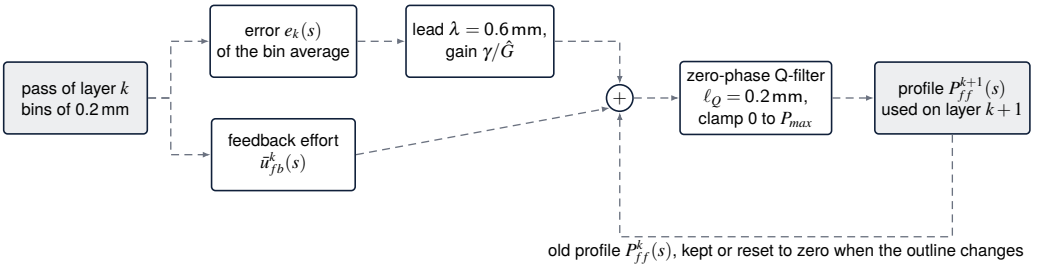


Figure 10.2 The learning update of Equation (10.6) as it runs between two layers. The error of each 0.2 mm bin is read a lead ahead and scaled by the learning gain, the feedback effort of the bin is folded in, and the sum with the old profile is filtered and clamped before the next layer uses it. The Q-filter is the block that keeps short wavelengths, where the fixed lead is wrong, out of the learning.

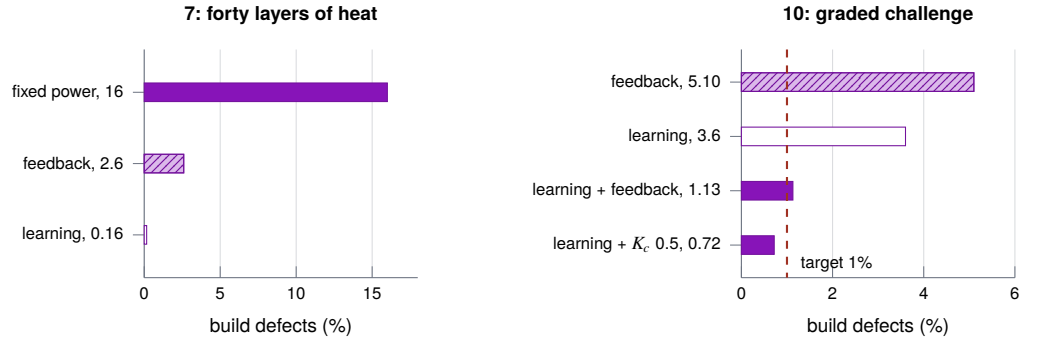


Figure 10.3 Build defects of two cases of Table 10.4, with each value printed beside its label. Over forty layers of heat buildup, feedback removes the slow drift but not the corners, and learning removes both. In the graded challenge, learning alone leaves most of its defects on the first bracket layer and the short-fed layer, and learning with feedback underneath passes the 1 percent target (dashed) only with the feedback gain lowered to 0.5 W/µm.

At short wavelengths a first-order lag contributes much less phase than a delay of the same time constant, so a fixed lead overcompensates there, and the bracket in Equation (10.7) can exceed one [3]. The Q-filter removes those wavelengths from the learning.

🕮 **Pause.** Try Exercise 10.2 before reading on.

10.5 Results

Table 10.3 lists the ten cases, and Table 10.4 collects the outcomes measured at the app's default seed (Section 1.6.1). Each case was run from its loaded settings, with any change made before Run. Figure 10.3 draws the build defects of cases 7 and 10 from that table.

10.5.1 What Feedback Can and Cannot Do

At a fixed 200 W the width reaches 140 µm at every corner on the first layer and 4.7 % of the path is in the keyhole regime. By the fifth layer the part is about 50 K warmer, the corners reach 150 µm and 7.2 % of the path is defective.

Feedback at 1 W/µm takes the corners to about 135 µm and the keyhole share to about 2.5 % on every layer, with an rms error near 6.7 µm, and it removes the slow heat drift completely. It cannot do more, because by the time the photodiode sees the pool widen, the filter, the 300 µs latency and the melt pool lag have used most of the corner. Volume 3, Chapter 6 meets the same limit on a

Table 10.3 Scenario Sequence

#	Scenario	Teaching point
1	Fixed power	corners and heat
2	Feedback on the melt pool	what feedback cannot catch
3	More feedback gain	latency and spatter
4	Learn from the last layer	iterative learning
5	Without the Q-filter	divergence
6	A bracket with an overhang	learning with feedback
7	Forty layers of heat	slow drift
8	The part changes shape	broken repetition
9	Learning with feedback	a layer that does not repeat
10	Challenge	graded build

Table 10.4 Measured Results

Case and condition	Result
1, layer 1 / layer 5	corners 140 / 150 µm, defects 4.7 / 7.2%
2, feedback $K_c = 1$, 10 layers	corners ≈ 135 µm, ≈2.5%, 6.7 µm rms
3, K_c 1 / 2 / 3 / 5	build defects 2.6 / 2.9 / 4.1 / 36.4%
3, spatter 10%, K_c 3 / 1	build defects 8.2 / 2.7%
4, ILC $\gamma = 0.5$	4.7, 1.8, 0% from layer 3, 3.0 µm at layer 10
4, $\gamma = 1.0$ / 1.5	3.4 µm at layer 3 / 1.0–3.1% on layers 6–10
5, no Q, 25 layers	low 5.3 µm at layer 4, then 17 µm, 24.8%
5, ℓ_Q 0.2 / 0.6 mm	2.5–3.1 / 4.7 µm
5, no Q, lead 0.2 mm	converges to 1.6 µm
6, bracket, feedback	6.3% (3.6 keyhole, 2.8 lack of fusion)
6, feedback and ILC	0.7% at layer 2, then ≤0.4%
6, $K_c = 0.5$ with ILC	0% from layer 3, 4.2 µm rms
7, 40 layers, fixed power	+140 K, 4.7 to 23.6% per layer, build 16%
7, feedback / ILC	build 2.6 / 0.16%
8, shape change, keep / reset	13.2 / 23.2% on first new layer
8, keep, with feedback	5.5%
9, short feed, ILC only	12.5% then peak 127 µm, build 1.7%
9, ILC and feedback $K_c = 0.5$	0, 0.4, 0.7%, build 0.40%
10, FB / ILC / ILC+FB / K_c 0.5	build 5.10 / 3.6 / 1.13 / 0.72%

slower scale, where a coagulant loop on the settled water acts on water that has already left the plant. Raising the gain makes things worse. The defect share over ten layers is 2.6 % at 1, 2.9 % at 2 and 4.1 % at 3 W/µm, because the extra gain turns spatter into power and pushes the width below the lack-of-fusion threshold. Over eight other noise seeds the share is 2.6 to 2.7 % at 1 and 2.7 to 3.5 % at 2, higher at 2 for seven of the eight, so the first step is small. At 5 the latency makes the loop oscillate, growing layer by layer until 74.0 % of the tenth layer and 36.4 % of the build are defective. With 10 % spatter the gap widens to 8.2 % against 2.7 %.

🕮 **Pause.** Try Exercise 10.1 before reading on.

✦ DID YOU KNOW? **More Gain, More Defects**

From the simulator: raising the feedback gain from 1 to 2, 3 and 5 W/µm raises the defective share of a ten-layer build from 2.6 to 2.9, 4.1 and 36.4 percent. The extra gain turns spatter into power, and at the highest gain the latency makes the loop oscillate (Table 10.4).

10.5.2 What Learning Does

With feedback off and learning on at $\gamma = 0.5$, the first layer is the fixed-power case at 4.7 %, the second is 1.8 %, and from the third there are no defects, while the rms error continues down to 3.0 µm by layer ten, less than half of what feedback achieved (Table 10.4). Figure 10.4 shows the defects

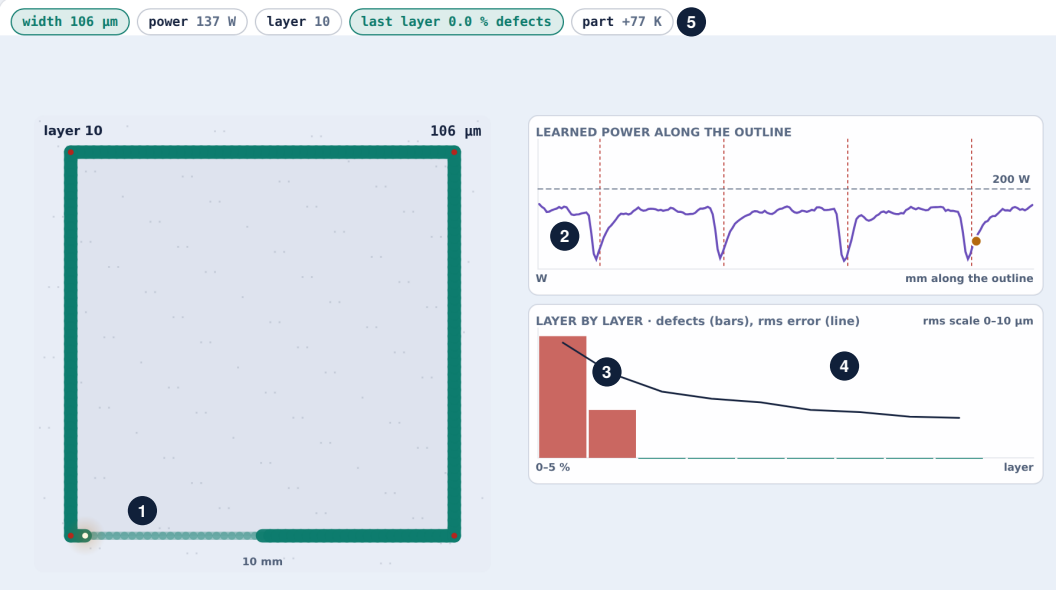


Figure 10.4 The powder bed application running scenario 4, learning without feedback on the square outline, cropped to the plant view. The run is at 0.515 s, 35 mm into the tenth and last layer. (1) The melt pool moves along the outline, and the melted track is colored by width over the faint track of the previous layer, with the four corners marked. (2) The learned power profile dips a fraction of a millimeter before each corner, each marked by a dashed line. (3) The layer-by-layer bars show defects on the first two layers only. (4) The rms error of each layer falls from layer to layer. (5) The status chips give the width, the power, the layer and the part 77 K warmer.

confined to the first two layers and the rms error falling. A learning gain of 1 reaches 3.4 µm by the third layer rather than the eighth, at the cost of 2.1 % lack of fusion on the second. A gain of 1.5 overcorrects, and because each update also copies the noise of the previous layer it leaves 1.0 to 3.1 % defective on layers 6 to 10.

> **TRY IT IN THE APP** **4 · Learn From the Last Layer**
>
> Load the scenario and press Run. Feedback is off and learning is on for ten layers. Watch the layer-by-layer plot: 4.7 percent of the first layer is defective, 1.8 percent of the second, and none from the third on, while the rms error falls to 3.0 µm by layer ten (Table 10.4).

10.5.3 Why the Q-Filter Matters

Without the Q-filter the learned profile improves for a few layers, with the rms error lowest at 5.3 µm on layer four, and then short-wavelength ripple grows layer after layer until 24.8 % of the path is defective at layer 25 with an rms error of 17 µm. With a 0.2 mm filter the error settles at 2.5 to 3.1 µm with no defects after layer two, and with a 0.6 mm filter it converges more slowly, with defects until layer eight, and settles near 4.7 µm because the filter also blurs the corners. A lead of 0.2 mm, one bin and closer to the true lag, converges to 1.6 µm even without a filter, while a lead of 0.4 mm already drifts back up to 4.4 µm by layer 25. The filter is insurance against the model error in the lead, which a real machine never knows exactly, and choosing it is most of the design effort in a practical learning controller [2, 3].

TRY IT IN THE APP **5 · Without the Q-Filter**

Load the scenario, press Run and let all 25 layers run. The rms error is lowest at 5.3 µm on layer four, then climbs to 17 µm, with 24.8 percent of layer 25 defective. Press Reset, set Q-filter length to 0.2 mm and run again: the error settles at 2.5 to 3.1 µm with no defects after layer two.

10.5.4 Overhangs, Heat and Broken Repetition

The bracket adds an overhang edge that runs hot and an inside corner that runs cold. Feedback alone leaves about 6.3 % of every layer defective, 3.6 % keyhole and 2.8 % lack of fusion. With learning added to feedback the defects fall to 0.7 % on the second layer and stay at or below 0.4 %, and with the feedback gain lowered to 0.5 W/µm they are zero from the third layer at 4.2 µm rms, because feedback no longer has the repeating part to chase and only adds noise. Over forty layers at fixed power the part warms by about 140 K, the defect share rises from 4.7 to 23.6 % per layer and 16 % of the build is defective. Feedback removes the drift and leaves 2.6 %, all at the corners, and learning alone leaves 0.16 %, all on the first two layers.

When the square becomes a bracket at layer nine, the learned profile lowers power where a convex corner has become a cold concave one and has nothing for the two new corners and the overhang, so 13.2 % of that layer is in the keyhole regime, and one layer later the new shape has been learned. Forgetting the profile is worse, 23.2 %, because three corners and the heat buildup were still right. With feedback running underneath, the first bracket layer is 5.5 % defective. A layer fed short of powder is the opposite failure. Learning alone leaves that layer 12.5 % lack of fusion (12 to 22 % over eight other noise seeds) and then applies the extra power it learned to the next layer, where the width peaks at 127 µm, for 1.7 % of the build. With feedback at 0.5 W/µm underneath, the short layer has no defects, the next is 0.4 % defective, the one after 0.7 % with lack of fusion, and the build 0.40 %.

Pause. Try Exercise 10.3 before reading on.

10.5.5 The Graded Challenge

The challenge builds twenty layers, six of the square and then the bracket, with a short powder feed on layer fifteen and the part heating throughout, starting with feedback only at 1 W/µm. It is met with less than 1 % of the scanned path outside the width window and the width within 30 µm of setpoint over the last third. As loaded, feedback leaves 5.1 % of the build defective, about 6 % of every bracket layer, and the width 48.6 µm from setpoint. Learning alone leaves 3.6 %, most of it on the first bracket layer and on the short-fed layer, where 50 % of the path is lack of fusion. Learning with the default feedback gain leaves 1.13 %, just over the target, with the width 34.5 µm from setpoint, and learning with the feedback gain lowered to 0.5 W/µm leaves 0.72 % and 26.0 µm, with the first bracket layer the worst at 8.7 % and the short-fed layer at 0.6 %. That run passes with 0.28 percentage points and 4.0 µm to spare, and the same settings carried through Reset give the same result. The margins are narrow against the noise. Over eight other seeds the default gain passes once and the lower gain fails once, on the width.

10.6 Technology in Practice

What Measures It

The loop of this chapter reads one signal, the melt pool width seen by a coaxial photodiode and filtered with a 0.2 ms time constant, and the learning controller averages the same signal in 0.2 mm bins along the outline. Figure 10.5 draws both loops on machine hardware. A research testbed carries more than one view of the process: a coaxial high-speed camera behind a band-pass filter at 850 nm, which records up to 10 000 frames a second, and a layer camera above the bed that images every layer before and after the powder is spread [14]. The chapter treats its photodiode as a calibrated width sensor, which Section 10.7 lists as a limitation.

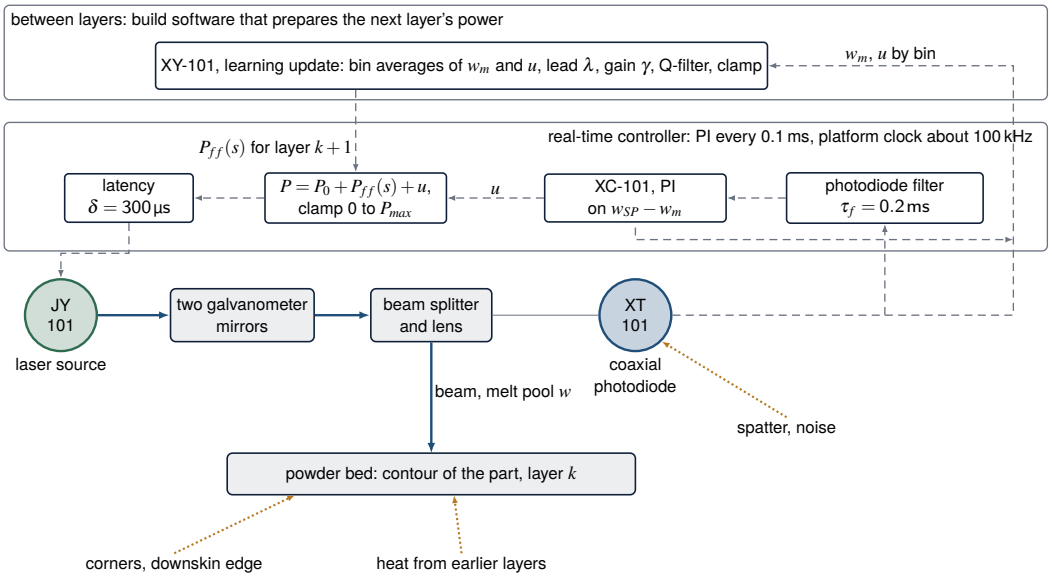


Figure 10.5 The chapter's two loops on machine hardware. Inside a pass, the PI loop on the filtered photodiode width trims the power that the learned profile asks for, and the command reaches the melt pool after the latency. Between passes, the learning update turns the binned width error and feedback effort of layer k into the power profile of layer $k+1$. Dotted arrows mark disturbances. Tags are the book's; the chapter names none.

TECHNOLOGY **Sensor: A Coaxial Photodiode**

A coaxial camera looks down the laser's optical path, so the melt pool stays still in its view wherever the mirrors point [14]. A photodiode placed the same way gives one signal that follows the pool's area [11]. It integrates the thermal emission continuously and has no spatial resolution, but its high sampling rate and its sensitivity to thermal transients complement a camera [15]. One research machine sampled its coaxial photodiode at 50 kHz in a band of 700 to 1000 nm [16]. On a trend by layer, sharp rises in intensity marked hot spots, a surface raised above the build plane and swelling [15]. The model's photodiode reads the width itself, with 3 % spatter and 2 µm of noise (Table 10.1).

What Moves It

Laser power is the only manipulated variable of the chapter, and the path and the speed along it are fixed by the outline. On a real machine the scan path and the power are generated from the sliced part before the build and interpolated into digital position and power commands. A power adjustment module can sit between those commands and the controller and change the power before it is sent [13]. That is where a learned profile such as $P_{ff}(s)$ would enter, as a power written into the next layer's commands.

TECHNOLOGY Final Element: Laser and Galvanometer Mirrors

Two mirrors turned by a galvanometer system steer the beam, a lens focuses it to a spot 0.05 to 0.5 mm wide, and scan speeds reach 7 m/s [11]. The controller sends two digital streams: positions to the galvanometer drive, which closes its own PID loop, and a power command to the laser, usually as an analog voltage [13]. Command and result differ. On one testbed the true power was slightly less than commanded, and a galvanometer position error built the parts slightly contracted [14]. In the model the power reaches the pool 300 µs after its command, and the spot slows to half speed near every corner (Table 10.2).

Where the Controller Runs

A typical controller of the process works on a clock of 100 kHz [13], and one testbed issues each position and power command on a 10 µs step [14]. The PI loop of this chapter executes every 0.1 ms, ten of those steps, and its learning update runs once per layer, after the pass. A control loop may act discretely between the completion of build layers rather than continuously [11], and commercial systems already adjust the energy input at every layer from layer-wise measurements [16]. Read as machine hardware, the PI loop belongs in the real-time controller beside the mirror and power streams, and the learning update with its lead, Q-filter and clamp belongs in the software that prepares the next layer between passes.

IN PRACTICE A Qualified Process Is a Locked Process

The space agency standard NASA-STD-6030, *Additive Manufacturing Requirements for Spaceflight Systems*, released in 2021, sets qualification and certification strategies for additive parts. A qualified and locked process defines the flaw state a part is expected to carry, and each part must be shown to represent that process. In-situ monitoring today looks for the conditions that lead to flaws rather than for the flaws themselves, and a monitoring baseline of one component can serve as a fingerprint for its later builds [17]. The defect share of this chapter is likewise a proxy for porosity, not a measurement of it (Section 10.3).

What Is Modernizing It

Three developments touch this loop. The first is feedforward from geometry. A geometric conductance factor computed from the share of solid and powder near the melt pool lowers the power near edges and overhangs, and the parts built with the strongest adjustment showed the lowest variation of melt pool intensity [13]. The second adds a model. A vector-level feedforward controller that couples a fast thermal model to a reduced melt pool model reduced geometric inaccuracy in key dimensions by 62 percent, porosity by 16.5 percent and photodiode variation by 6.8 percent on average [16]. Both act before the pool widens, as the learned profile of this chapter does, but from a model rather than from the last layer. The third is richer monitoring: photodiode traces mapped layer by layer beside camera images to classify anomalies, as groundwork for machine learning [15], and monitoring baselines used as a fingerprint of a qualified build [17].

10.7 Limitations

Only the contour is scanned. The hatch that fills the interior dominates the energy input and has its own features, such as vector turnarounds, that a learning controller would also address.

The melt pool model is phenomenological. A real pool changes mode abruptly at the keyhole threshold, and its width is only one of the quantities that determine porosity [4].

Geometry repeats exactly on every layer of the same outline. Real cross-sections change continuously with height, which blurs the distinction between the repeating error and the new one [7].

OPEN CHALLENGE **Which Loop, on Which Time Scale**

Most reported control of powder bed fusion uses melt pool temperature or size to set the laser power or speed. Loops could also act once between layers rather than continuously, but it is not yet clear which signatures are best measured or modeled, and which inputs are best controlled on which time scale. A review of the field called for a control architecture that names every possible loop and shows which one suits which link between parameter, signature and part quality [11]. This chapter compares two such loops on one contour. Progress would be that comparison made on whole parts, with the hatch and evolving cross-sections.

The photodiode is treated as a calibrated width sensor. Relating a real photodiode signal to a width is itself a modeling problem [10].

The part cools by a fixed fraction per layer, whereas real interlayer temperature depends on part geometry, scan strategy and recoat time, which is the subject of layer-to-layer temperature control [8].

OPEN CHALLENGE **Learning Under a Locked Recipe**

Feedback on the melt pool stays reactive and cannot act ahead of a known overhang, while changing process parameters in real time, layer by layer or faster, complicates certifying parts for critical applications [16]. Qualification rests on a locked process [17], and a learned profile changes the power on every layer until it converges. Progress would be a way to qualify the learning controller itself, with bounds on what it may change and each learned profile kept as part of the build record.

10.8 Conclusion

A melt pool that overheats at the same corners on every layer is a disturbance feedback cannot catch and learning can. In the simulator described here, fixed power leaves 4.7 % of the first layer in the keyhole regime, feedback with 300 µs of latency trims it to about 2.5 %, and more gain raises the defect share to 4.1 % at 3 W/µm. Learning from the previous layer removes the defects from the third layer on and more than halves the rms error to 3.0 µm. Without its Q-filter the learning diverges to 24.8 % defects by layer 25, because a fixed lead overcompensates the melt pool lag at short wavelengths. When the repetition breaks, keeping the learned profile (13.2 % on the first new layer) beats discarding it (23.2 %), and feedback underneath the learning limits the damage of a layer that does not repeat, taking the build from 1.7 % to 0.40 % defective. The model scans one contour, uses a phenomenological melt pool and counts a width proxy rather than porosity. Further work would add the hatch, a keyhole transition, geometry that evolves with height and a sensor model. The app is a way to practice control of a process that repeats itself.

KEY POINTS

- A melt pool that overheats at the same corners on every layer is a disturbance feedback cannot catch and learning can.
- Feedback with 300 µs of latency trims the keyhole share to about 2.5 %, and more gain makes the build worse.
- Learning from the previous layer removes the defects from the third layer on and more than halves the rms error to 3.0 µm.
- Without its Q-filter the learning diverges, because a fixed lead overcompensates the melt pool lag at short wavelengths.
- When the outline changes, keeping the learned profile beats discarding it, and feedback underneath the learning limits the damage of a layer that does not repeat.

Where else this applies. Chapter 7 makes the same choice between reacting and remembering with one number per cycle (Section 1.4.5), and Section 1.4.6 sets out the learning update on the batch kettle.

References

[1] S. Arimoto, S. Kawamura, and F. Miyazaki, "Bettering operation of Robots by learning," *Journal of Robotic Systems*, vol. 1, no. 2, pp. 123–140, 1984.

[2] R. W. Longman, "Iterative learning control and repetitive control for engineering practice," *International Journal of Control*, vol. 73, no. 10, pp. 930–954, 2000.

[3] M. Norrlöf and S. Gunnarsson, "Time and frequency domain convergence properties in iterative learning control," *International Journal of Control*, vol. 75, no. 14, pp. 1114–1126, 2002.

[4] W. E. King, H. D. Barth, V. M. Castillo, G. F. Gallegos, J. W. Gibbs, D. E. Hahn, C. Kamath, and A. M. Rubenchik, "Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing," *Journal of Materials Processing Technology*, vol. 214, no. 12, pp. 2915–2925, 2014.

[5] T. DebRoy, H. Wei, J. Zuback, T. Mukherjee, J. Elmer, J. Milewski, A. Beese, A. Wilson-Heid, A. De, and W. Zhang, "Additive manufacturing of metallic components – Process, structure and properties," *Progress in Materials Science*, vol. 92, pp. 112–224, 2018.

[6] T. Craeghs, F. Bechmann, S. Berumen, and J.-P. Kruth, "Feedback control of Layerwise Laser Melting using optical sensors," *Physics Procedia*, vol. 5, pp. 505–514, 2010.

[7] D. J. Hoelzle and K. L. Barton, "A new spatial Iterative Learning Control approach for improved micro-Additive Manufacturing," in *2014 American Control Conference*, 2014, pp. 1805–1810.

[8] B. Kavas, E. C. Balta, M. Tucker, A. Rupenyan, J. Lygeros, and M. Bambach, "Layer-to-layer closed-loop feedback control application for inter-layer temperature stabilization in laser powder bed fusion," *Additive Manufacturing*, vol. 78, p. 103847, 2023.

[9] S. Berumen, F. Bechmann, S. Lindner, J.-P. Kruth, and T. Craeghs, "Quality control of laser- and powder bed-based Additive Manufacturing (AM) technologies," *Physics Procedia*, vol. 5, pp. 617–622, 2010.

[10] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, "Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing," *Materials & Design*, vol. 95, pp. 431–445, 2016.

[11] M. Mani, B. Lane, A. Donmez, S. Feng, S. Moylan, and R. Fesperman, "Measurement science needs for real-time control of additive manufacturing powder bed fusion processes," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. NISTIR 8036, 2015.

[12] M. Seifi, A. Salem, J. Beuth, O. Harrysson, and J. J. Lewandowski, "Overview of Materials Qualification Needs for Metal Additive Manufacturing," *JOM*, vol. 68, no. 3, pp. 747–764, 2016.

[13] H. Yeung, B. Lane, and J. Fox, "Part geometry and conduction-based laser power control for powder bed fusion additive manufacturing," *Additive Manufacturing*, vol. 30, p. 100844, 2019.

[14] B. Lane and H. Yeung, "Process monitoring dataset from the additive manufacturing metrology testbed (AMMT): Overhang part X4," *Journal of Research of the National Institute of Standards and Technology*, vol. 125, p. 125027, 2020.

[15] H. Hyer, L. Scime, Z. Snow, C. Joslin, T. McDonald, C. Leach, J. Kendall, B. Schreiber, and C. Petrie, "Demonstrating advanced sensors for in situ monitoring toward qualification of nuclear-relevant components," Oak Ridge National Laboratory, Oak Ridge, TN, Tech. Rep. ORNL/TM-2025/4056, 2025.

[16] N. Kirschbaum, N. Wood, C.-E. Kim, T. U. Tumkur, and C. Okwudire, "Vector-level feedforward control of LPBF melt pool area using a physics-based thermal model," *Additive Manufacturing*, vol. 112, p. 104981, 2025.

[17] W. Tilson, D. Wells, E. Lanigan, B. West, R. Russell, A. Park, B. McEnerney, and A. Glendening, "NASA's certification and qualification challenges for additively manufactured hardware: What is next?" in *Proceedings of the 17th European Conference on Spacecraft Structures, Materials and Environmental Testing (ECSSMET)*, Toulouse, France, 2023, nASA Technical Reports Server 20230003172.

10.9 Exercises

10.1 Concept check · 1 min

In scenario 2 a PI loop on the photodiode, running every 0.1 ms at $K_c = 1\,\mathrm{W}/\mu\mathrm{m}$, still lets every corner reach about 135 µm. Which reason does the chapter give?

(a) The filter, the latency and the melt pool lag use up most of a corner that lasts about a millisecond

(b) The power limit of 370 W stops the loop from moving the power as far as a corner needs

(c) The integral time of 1 ms is too long for the integral to build up before each corner

(d) The 3 % spatter on the photodiode hides the widening at a corner from the loop

10.2 Concept check · 1 min

Without its Q-filter the learning controller of scenario 5 improves for four layers and then gets worse until a quarter of the path is defective. Why?

(a) The learning gain of 0.5 is too high for the nominal plant gain, so every wavelength overshoots

(b) The fixed 0.6 mm lead overcompensates the lag at short wavelengths, so those components grow

(c) The part heats from layer to layer, so the profile learned on one layer no longer fits the next

(d) The clamp on the learned power cuts the corner corrections and leaves an error that accumulates

10.3 Concept check · 2 min

In scenario 8 the square outline becomes a bracket at layer 9. Keeping the learned profile leaves 13.2 % of that layer defective and forgetting it leaves 23.2 %. Why is keeping it better?

(a) The kept profile already holds a correction for the overhang, learned from the bracket on earlier layers

(b) Choosing Forget also switches the feedback loop off until the new outline has been scanned once

(c) Three corners and the heat buildup are unchanged, so most of what was learned is still right

(d) The kept profile is filtered again at the change, which removes the corner it no longer needs

10.4 Theory and calculation · 6 min

The scan speed falls to $\rho = 0.5$ of $v_0 = 800\,\mathrm{mm/s}$ within $d_c = 0.5\,\mathrm{mm}$ of every corner, by Equation (10.3). At the apex of a convex corner $g = 1.15$. Use $n = 0.7$, $m = 0.35$, $w_0 = 100\,\mu\mathrm{m}$ at $P_0 = 200\,\mathrm{W}$ and a cold part, and $\int_{-\pi}^{\pi} \mathrm{d}x/(A - B\cos x) = 2\pi/\sqrt{A^2 - B^2}$.

(a) Find the time the spot spends in the slowed zone of one corner, and the time of one pass around the 40 mm square with its four corners. *Checkpoint:* 1.77 ms and 52.1 ms.

(b) Find the width the melt pool would settle to at the apex at 200 W, and say why the first layer of scenario 1 peaks lower, at 140 µm. *Checkpoint:* 146.6 µm.

(c) Find the power that holds 100 µm at the apex, and the share of the time in the slowed zone that the photodiode filter, the latency and the melt pool time constant take together. *Checkpoint:* 115.8 W; 0.8 ms, about 45 %.

10.5 Theory and calculation · 6 min

At a fixed 200 W a pass of the square takes 52.06 ms. The part gains $c_E = 1.36\,\mathrm{K/J}$ per joule delivered and keeps $\kappa = 0.9$ of its rise through each recoat, by Equation (10.4), and the width rises by $\alpha = 1.5 \times 10^{-3}\,\mathrm{K^{-1}}$ per kelvin.

(a) Find the energy of one pass and the temperature rise the part settles to after many layers. *Checkpoint:* 10.41 J and 141.6 K.

(b) Find the rise during layer 5, after four passes, and the factor by which it widens the pool. *Checkpoint:* 48.7 K and 1.073.

(c) The corners of layer 1 peak at 139.6 µm. Predict the peak on layer 5, and the width on a straight edge, 100 µm on a cold part, once the part has settled. *Checkpoint:* 149.8 µm and 121 µm.

10.6 Theory and calculation · 7 min

The Q-filter of Section 10.4 runs $y_i = a\,y_{i-1} + (1-a)\,x_i$ over the 0.2 mm bins once forward and once backward, with $a = \exp(-0.2\,\mathrm{mm}/\ell_Q)$. A component that turns by θ radians per bin is multiplied by $(1-a)/|1-a\,\mathrm{e}^{-\mathrm{j}\theta}|$ on each pass.

(a) Find a for $\ell_Q = 0.2$ and 0.6 mm, and the combined gain of the two passes as a function of θ. *Checkpoint:* 0.368 and 0.717; $(1-a)^2/(1-2a\cos\theta+a^2)$.

(b) Evaluate that gain for a constant, for a pattern that alternates from bin to bin (a 0.4 mm wavelength) and for a 2 mm wavelength, for both filters. *Checkpoint:* 1, 0.214, 0.740 at 0.2 mm; 1, 0.027, 0.227 at 0.6 mm.

(c) The filter multiplies the bracket of Equation (10.7). Find the largest bracket magnitude the 0.2 mm filter still turns into convergence at the two wavelengths, and say what the 0.6 mm filter costs at a corner. *Checkpoint:* 4.7 and 1.35.

10.7 App challenge · 8 min

Load scenario 3, More feedback gain, which opens with Feedback gain at 3 W per µm. Use the default seed and a speed of 0.1, which plays 0.1 s of scanning each second, so a ten-layer build takes about 5 s. The run stops itself at the end of the build. Make every change before pressing Run.

(a) Run as loaded and read Build defects. *Checkpoint:* 4.10 %.

(b) Press the reload button, set Feedback gain to 1 and run. Repeat at 2. *Checkpoint:* 2.57 % and 2.93 %.

(c) Press the reload button, set Spatter on the photodiode (1σ) to 10 %, and run at a gain of 3. Then press Reset, set the gain to 1 and run again. Explain the gap between the two. *Checkpoint:* 8.25 % and 2.66 %.

Accept Build defects within 0.02 percentage points.

10.8 App challenge · 9 min

Load scenario 4, Learn from the last layer, which runs ten layers with Feedback off and Learning (ILC) on at a learning gain of 0.5. Use the default seed and a speed of 0.1. After each run export the CSV file: a layer's scores sit in the columns `layer_rms_um`, `layer_defect_pct` and `layer_lof_pct` of the first row whose `layer` column shows the next layer, and those of layer 10 in the last row.

(a) Run as loaded. Read the defects of layers 1, 2 and 3 and the rms error of layer 10. *Checkpoint:* 4.69, 1.84 and 0.00 %; 2.98 µm.

(b) Press Reset, set Learning gain to 1 before Run and run. Read the lack of fusion on layer 2 and the rms error of layer 3. Which layer of the first run first had a lower rms error than that? *Checkpoint:* 2.10 %, 3.44 µm; layer 8.

(c) Press Reset, set Learning gain to 1.5 and run. Read the defects of layers 6 to 10. Then set Spatter on the photodiode and Photodiode noise to zero, press Reset and run again. *Checkpoint:* 0.97 to 3.10 %; no defects from layer 3.

Accept rms errors within 0.02 µm and shares within 0.02 percentage points.

10.9 App challenge · 8 min

Load scenario 5, Without the Q-filter, which runs 25 layers of learning alone with Q-filter length at 0. Use the default seed and a speed of 0.25, so a build takes about 5 s. Read each layer's scores from the exported CSV file, as in the previous exercise.

(a) Run as loaded. Find the layer with the lowest rms error and its value, and the rms error and defect share of layer 25. *Checkpoint:* Layer 4, 5.34 µm; 17.03 µm and 24.79 %.

(b) Press Reset, set Q-filter length to 0.2 mm before Run and run. Repeat at 0.6 mm. Read the rms error of layer 25 and the layers that still have defects. *Checkpoint:* 2.83 µm, defects on layers 1 and 2 only; 4.70 µm, defects through layer 8 and on layers 19 and 24.

(c) Press Reset, set Q-filter length back to 0 and Learning lead (under Advanced) to 0.2 mm, and run. Read the rms error of layer 25 and explain why the shorter lead converges without a filter. *Checkpoint:* 1.65 µm.

Accept rms errors within 0.02 µm and shares within 0.02 percentage points.

10.10 Application area · 3 min

A quality manager proposes recording the defect share the controller reports for each layer as the porosity of the part, for its qualification file. Give two reasons from the chapter why that number cannot serve as porosity, and one thing that would have to be added.

10.11 Application area · 3 min

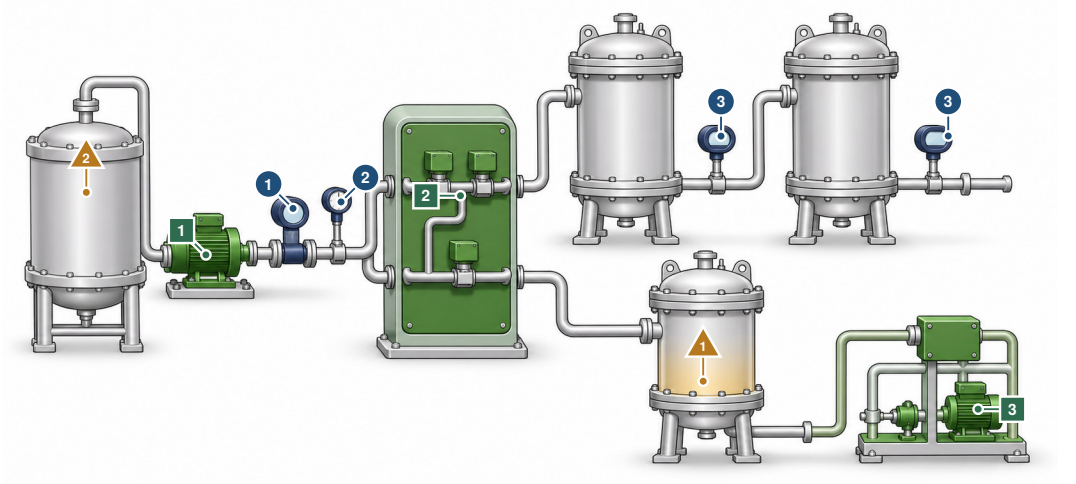
A machine builder plans to ship iterative learning on its production machines. From scenarios 6, 8 and 9 and the limitations, state three rules for the design and the reason for each.

10.12 Review the agent · 5 min

An AI assistant is asked how to pass the challenge, A build that changes, and writes: “Switch Learning (ILC) on and leave the rest as loaded. Learning removes the corners, the overhang and the heat on every repeated layer, and the default feedback gain of 1 catches the short powder feed. The build will come in under 1 %.” Find the flaw and say how you would check it.

CHAPTER 11

Switching a Three-Column Antibody Capture Train on Resin of Unknown Capacity



The controller moves 1 the feed pump 2 the valve block that switches the columns 3 the pump that regenerates the offline column

The controller reads 1 the feed flow 2 the pressure at the train inlet 3 a detector at each column outlet

What disturbs it 1 the capacity the resin has left, which nothing measures 2 what the harvest brings

Notice that the two columns in the upper row run in series, so what passes the first is caught by the second, while the third is regenerated on its own pump. Illustration, not to scale.

IN PLAIN TERMS Catching a Medicine on Columns That Wear Out

The main idea. A protein medicine is caught from a broth by pumping it through columns of tiny beads that hold it. Two columns run in a row so the second catches what spills past the first, while a third is washed out.

The challenge. Nobody can measure how much a column can still hold, and it shrinks with every use. Switch too early and beads and wash liquid are wasted, too late and medicine leaves the second column.

Why it matters. The rule for when to switch sets how much medicine each column makes per hour and how much is lost.

THREE WAYS TO READ THIS CHAPTER

Explore	13 min	the first page, the history in Section 11.2, the two try-its and the fact in Section 11.6, the fact in Section 11.7, and the key points
Study	40 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Sections 11.6 and 11.7), Technology in Practice (Section 11.8), Limitations (Section 11.9) and the open challenge

The control question A three-column Protein A train takes its lead column off while the next one is still being regenerated, and the resin capacity that decides when to switch is never measured. Which switching rule puts the most antibody through each liter of resin per hour without losing it, and what changes when the resin ages or the regeneration cannot keep up?

Learning objectives

- Explain why breakthrough the lag column catches is not loss, and read both on the scorecard.
- Check a three-column schedule for feasibility from the regeneration recipe and the time to load one column.
- Compute the capacity an aged column offers and the rate at which three columns can take antibody.
- Explain how a filtered UV estimate and a capacity learned from a two-minute average move the breakthrough a policy realizes.
- Choose a switching rule and a response to a late column for resin of uncertain capacity, and justify the choice with the measured trade-offs.

Prerequisites Sections 1.4.2 (guards that end a phase), 1.4.5 (EWMA), 1.4.7 (hybrid automata), 1.4.9 (cyclic steady state) and 1.6.2 (the PID of the two loops), and Chapter 8 for phases ended on a clock and on an integrated estimate.

The app https://apmonitor.com/apps/mab-capture/

See also Volume 19, Chapter 8 for the culture that secretes the antibody, fed on daily samples, and Volume 19, Chapter 9 for a recovery train whose bottleneck vessel limits the harvest it can take. Volume 5, Chapter 3 for control through a slow analyzer and Volume 5, Chapter 7 for a continuous line judged on a delayed at-line measurement.

11.1 Introduction

The lead column of a capture train can break through by 90 % without losing antibody. In the simulated train of this chapter, three Protein A columns of 0.503 L take a clarified harvest at 3.0 g/L. Two of them are in series, so the lag column catches what the lead lets through, while the third is washed, eluted, cleaned and re-equilibrated in 23.2 min. Switching when the lead's outlet reaches 40 % of the feed loses 0.06 % of the antibody, and switching at 90 % loses 0.09 % while the lag catches 78.3 instead of 21.8 g over the campaign. What a deep switch costs is time and capacity downstream, and the limit on it is the resin's capacity, the one number no instrument in the plant reads.

The simulator is built around a misconception that breakthrough measures yield loss. A batch column is loaded to a fraction of its dynamic capacity because whatever leaves it is lost, and the dynamic capacity of Protein A media depends strongly on residence time [1]. Periodic counter-current chromatography removes that penalty by loading the lead into its breakthrough [2], and the decision moves from how far to load to when to switch, under three uncertainties at once. The capacity falls with cycles of use, the titer varies from lot to lot, and the 280 nm signal that reports breakthrough also carries host-cell proteins and media components [3]. A schedule must also be feasible, because with three columns the regeneration has to finish within the loading time [4]. Fixed beds of ion exchanger follow the same flowing-exchange kinetics [5], and any train of regenerable beds run lead and lag shares the decision of when a loaded bed comes off while the next one is still being restored.

Continuous capture is now an established option for antibodies. A four-column system has run uninterrupted for 30 days coupled to a perfusion culture [6], optimal design has compared batch and multi-column processes [7], and a mechanistic model has been used for online control under changing upstream conditions and resin aging [8]. These studies were carried out on plants whose bound antibody, true breakthrough and remaining capacity are known only through samples, so a reader cannot load the same column twice with one decision changed and see where every gram went. This chapter ranks fixed-time, integrated-load, breakthrough-threshold and adaptive-capacity switching by recovery, resin utilization, buffer use, productivity and switching losses when each runs a simulated three-column Protein A train with the same resin, feed, regeneration recipe and information. It then

follows how the ranking moves as the resin's capacity departs from what its recipe assumes. The culture that secretes the antibody upstream of this train is the subject of Volume 19, Chapter 8.

The kinetic Langmuir law, the finite-volume transport, the switching rules and the learned capacity all have antecedents in [2, 3, 5, 8]. What the chapter adds are measurements on one synthetic train. Among them, a 25-minute clock that passes the challenge on new resin loses 8.86 % of the antibody when the resin holds 85 % of its capacity, and at 77.5 % no policy tested meets both the loss limit and the productivity target because the regeneration sets the pace. The simulator is a 380-state model whose antibody balance closes to 1.2×10^{-15} of the gross flux in each of 110 valve states. It is built on a column transport module shared with the peptide purification simulator of Volume 6, Chapter 9, and ten cases each credit one decision.

Section 11.2 follows the domain from the first immunosorbent to a rotating train. Section 11.3 sets out the train: transport and binding in each column, the capacity and its aging, the valves, the surge tank, the pressure and the instruments. Section 11.4 describes the two loops, the breakthrough estimate, the switching triggers and the supervisor that checks feasibility. Section 11.5 reports the code verification and the agreement with the literature, Section 11.6 the ten cases, and Section 11.7 the policies compared under equal resin and feed and over the resin's age. Section 11.8 describes the real detectors, valves and controllers, Section 11.9 states what the model leaves out, and the exercises of Section 11.11 work through the cases.

11.2 From an Immunosorbent to a Rotating Train

Affinity chromatography was introduced under that name in 1968, as a selective purification of enzymes on a ligand bound to the support [9]. Four years later Protein A from *Staphylococcus aureus* was itself isolated by affinity chromatography and used as an immunosorbent for the isolation of immunoglobulins [10]. The molecule became one of the most employed affinity purification systems, and a domain derived from it was improved to withstand the harsh environment of industrial purification [11]. By 2007 a platform process for monoclonal antibodies and Fc fusion proteins, developed on 14 molecules, rested on Protein A as its highly selective capture step [12].

Batch capture made dynamic capacity the design number. A comparison of 15 commercially available Protein A media measured equilibrium and dynamic capacities for human immunoglobulin G and found the dynamic capacity strongly dependent on residence time, with agarose media binding more at residence times above 3 min, and described the breakthrough profiles with the Thomas solution [1]. That solution goes back to a 1944 analysis of heterogeneous ion exchange in a flowing system [5], whose reversible second-order rate is the kinetic Langmuir law that this simulator integrates on a finite-volume grid.

Continuous operation reached capture next. A design method for periodic counter-current chromatography derived the residence time and number of columns from an empirical breakthrough model, verified them on two enzymes and an antibody, and attributed the economics to higher resin capacity utilization and lower buffer consumption [2]. The same group integrated a four-column system with a perfusion bioreactor and reported 30 days without a time-based decline in performance [6]. A decisional tool predicted direct-cost savings near 30 % for early clinical manufacture but near 5 % at commercial scale, and flagged the loss of binding capacity over long campaigns as a facility-fit risk [13]. A review of continuous downstream processing placed chromatography as the most advanced unit operation in continuous mode [14], and an optimal-design study found three-, four- and two-column processes with similar maximum capacity utilization, all well above batch [7].

Control of the load followed the measurements. A mechanistic model of affinity capture with transport into the beads was used both to design a two-column process and to change its operating conditions online, holding yield under changing upstream conditions and resin aging [8]. The measurements that next ended the load, and the guideline that followed, are told in the history box below.

Table 11.1 Variables by Role, With Values From the Reference Configuration

Variable	Role	Reference	Range
Q_f (FT-101)	CV of FIC-101	0.30 L/min	SP 0.10–0.80
u_P	MV, feed pump	31 %	0–100 %
P_{in} (PT-101)	CV of PIC-102	1.4 bar	limit 1.0–3.0 bar
Q_{lim}	MV of PIC-102	inactive	0.02–1.0 L/min
switch instant	discrete MV	policy	5 rules
Q_r	regeneration flow	0.40 L/min	0.1–0.8
$c_{P,h}$	DV, harvest titer	3.0 g/L	1–7
$c_{H,h}$	DV, UV background	1.2 g/L	0–4
r	DV, background drift	0	−50–150 %/h
a_0	DV, resin capacity left	1.0	0.4–1, hidden
$\widehat{BT}$	estimate, lead breakthrough	—	0–100 %
$\hat{L}, \hat{C}$	estimates, load, capacity	—	g/L of bed
c_i, q_i, φ	states, per cell	0 at $t = 0$	300
line tanks	states, 4 lines	feed line full	36
V_s, f_k, pumps, ledgers	scalar states	V_s 25 L	44
P	outcome, g/L/h	—	lead tile

> **HISTORY** **2017 to 2022: The Load Ends on a Measurement**
>
> A Protein A load was long fixed before it began. Its volume came from a titer measured offline and from a conservative capacity found in resin-lifetime studies, and an automated chromatograph that ended a load at 1 % breakthrough still carried a sample's delay [3]. A study published in 2017 used a partial least squares model on UV and visible spectra of the effluent to end a Protein A load in real time at a chosen breakthrough, after subtracting an impurity background averaged once the signal had stopped changing [3]. A 2018 study wrote the three-column condition, a regeneration shorter than the loading time, and found the yield lower in the first cycle, whose preload is recovered only after the run [4]. The guideline on continuous manufacturing, adopted in November 2022, illustrates a therapeutic-protein process with continuous capture chromatography, and lists resin lifetime among its run-time considerations [15].

11.3 The Capture Train as Modeled

The simulated route starts at clarified harvest and ends at the eluate pool, as shown in Figure 11.1. Harvest from the clarification arrives in a surge tank, the feed pump P-101 under FIC-101 sends it through FT-101, PT-101 and the feed detector AT-101 to the lead column, the lead's outlet runs into the lag, and the lag's outlet is the flow-through, a loss. The offline column runs the regeneration recipe on its own pump, P-201, and each column's outlet line ends in a valve block that sends it to the next column, the flow-through, the wash waste, the eluate pool or the cleaning and equilibration waste. Low-pH virus inactivation, polishing and formulation are downstream context only. The route is a teaching route proposed from published evidence and does not reproduce any commercial process.

The variables appear by role in Table 11.1. Each column has 20 cells of five states (antibody and background in the liquid and on the resin, and the elution strength), each outlet line and the feed line three well-mixed tanks of three states, and 44 scalars carry the pumps, the surge tank, the destinations' inventories, the buffers and the per-column fouling, for 380 states in all. Engine time is in seconds and the display shows minutes and hours.

11.3.1 Transport in Each Column

Each column is the shared finite-volume module of the series, a packed bed with axial dispersion and a flux inlet, written for a peptide purification simulator and moved into the framework when this simulator became its second user, unchanged in arithmetic. The liquid balance for species i (antibody

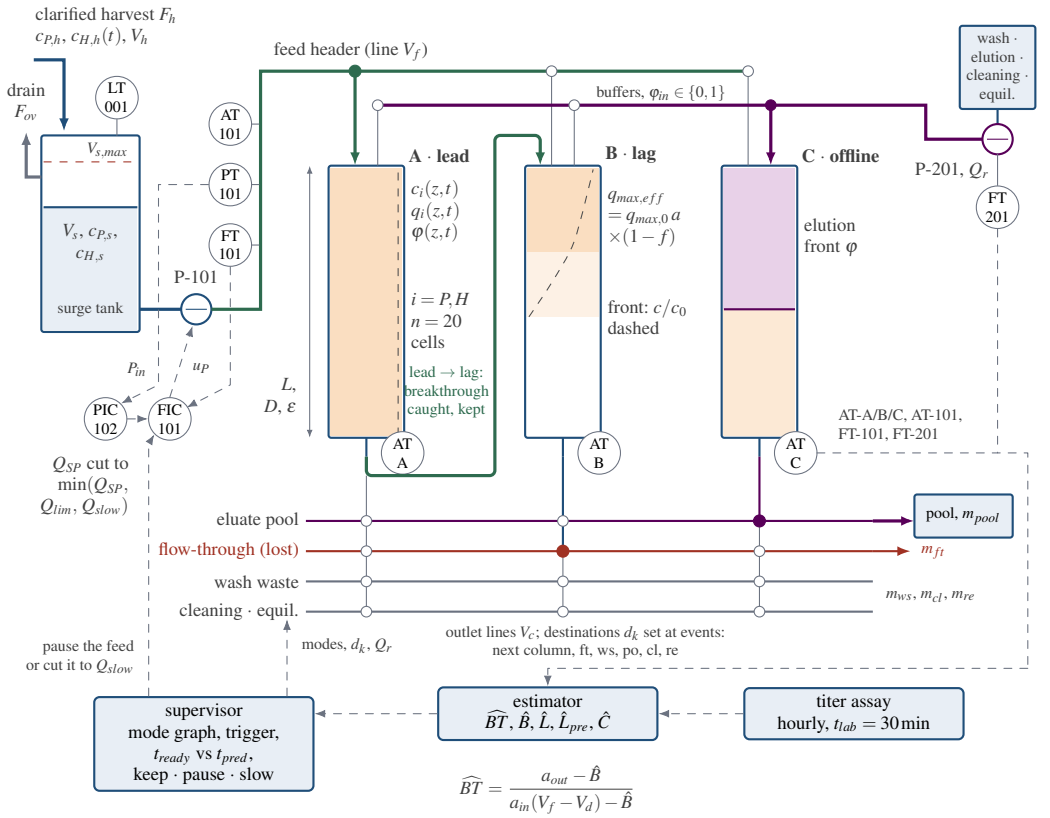


Figure 11.1 Three-column Protein A capture train as modeled, drawn at a switch. Harvest F_h fills the surge tank V_s, P-101 feeds the lead at Q_f under FIC-101 and the pressure override PIC-102, and each column of length L, diameter D and interstitial fraction ε carries $c_i(z,t)$, $q_i(z,t)$ and the elution strength $\varphi(z,t)$ on 20 cells, with a capacity $q_{max,eff}$ set by aging a and fouling f. The lead's outlet enters the lag, so its breakthrough is kept, while the lag's outlet joins the flow-through and is lost. P-201 drives the offline column through the buffers at Q_r, and the outlet manifold sends each line to the destination d_k set at events. The estimator and the supervisor read only AT-101, AT-A, AT-B, AT-C, FT-101, FT-201, PT-101, LT-001 and the titer assay. Process model only, not a piping diagram.

P or background H) and the elution strength φ are given in Equation (11.1).

$$\begin{aligned} \frac{\partial c_i}{\partial t} + \frac{1-\varepsilon}{\varepsilon}\frac{\partial q_i}{\partial t} + v\frac{\partial c_i}{\partial z} &= D_{ax}\frac{\partial^2 c_i}{\partial z^2}, \qquad v = \frac{Q}{\varepsilon A} \\ \frac{\partial \varphi}{\partial t} + v\frac{\partial \varphi}{\partial z} &= D_{ax}\frac{\partial^2 \varphi}{\partial z^2} \\ Qc_{i,in} = Qc_i(0) - \varepsilon A D_{ax}\frac{\partial c_i}{\partial z}\Big|_0, &\qquad \frac{\partial c_i}{\partial z}\Big|_L = 0 \end{aligned} \tag{11.1}$$

where q_i is mass per volume of stationary phase, $A = \pi D^2/4$, and $D_{ax} = D_{ax,ref}\, Q/Q_{ref}$ with $Q_{ref} = 0.30\,\mathrm{L/min}$. The inlet is the Danckwerts flux condition and the outlet a zero gradient [16]. The bed is

cut into $n = 20$ equal cells, and the flux through the face between cells j and $j+1$ is

$$\begin{aligned} F_{j+\frac{1}{2}} &= Q\Big(c_j + \tfrac{1}{2}\psi(c_j - c_{j-1},\, c_{j+1} - c_j)\Big) \\ &\quad - \varepsilon A D_{ax} \frac{c_{j+1} - c_j}{\Delta z} \\ \psi(a,b) &= \frac{2ab}{a+b} \text{ if } ab > 0, \text{ else } 0 \end{aligned} \tag{11.2}$$

the van Leer limited upwind value [17], with $F = Qc_{in}$ at the inlet face and no dispersive term at the outlet face. A cell's liquid changes by the difference of its face fluxes less what its resin takes up, $\varepsilon A \Delta z\, \dot{c}_j = F_{j-\frac{1}{2}} - F_{j+\frac{1}{2}} - (1-\varepsilon) A \Delta z\, \dot{q}_j$, so the column inventory closes to round-off.

11.3.2 Binding, Elution and Capacity

Adsorption follows a kinetic Langmuir law whose constants move with the elution strength, which is 0 in the load, wash and equilibration buffers and 1 in the elution and cleaning buffers, as shown in Equation (11.3).

$$\begin{aligned} \frac{\partial q_P}{\partial t} &= k_{on,0} \mathrm{e}^{-a\varphi} c_P (q_{max,eff} - q_P) - k_{off,0} \Big(\frac{k_{off,el}}{k_{off,0}}\Big)^{\varphi} q_P \\ \frac{\partial q_H}{\partial t} &= k_{on,H} \mathrm{e}^{-a\varphi} c_H (q_{max,H} - q_H) - k_{off,H} 10^{\varphi} q_H \end{aligned} \tag{11.3}$$

where φ is clipped to $[0,1]$ and a negative c is read as zero in the rate. Elution is therefore desorption where the elution front has arrived and transport of what desorbs, never $q = 0$ at a valve. In every elution of eleven audited runs the largest one-step fall of bound antibody is 1.17 % of the bound mass, and at most 0.020 % is left for the cleaning. The background H is one lumped species of host-cell protein and media components, weakly and reversibly bound.

> **✦ DID YOU KNOW? One Hundred Cycles on a Cleaner Feed**
>
> A four-column Protein A capture train was run for 100 cycles on two harvests. The harvest whose clarification also removed host-cell proteins and DNA gave 49 % more productivity than the depth-filtered one, and its eluate held 11 times less host-cell protein [4].

The capacity a column offers is its new capacity times two factors, as shown in Equation (11.4).

$$\begin{aligned} q_{max,eff} &= q_{max,0}\, a_k \left(1 - \min(f_k, 0.9)\right) \\ a_k &\leftarrow a_k (1-\delta) \quad \text{at the end of each cleaning} \\ \frac{\mathrm{d}f_k}{\mathrm{d}t} &= k_f \frac{\dot{m}_{H,k}}{V_{col}}\Big|_{\text{lead, lag}} - \frac{f_k}{\tau_{CIP}}\Big|_{\text{cleaning, hold}} \end{aligned} \tag{11.4}$$

where a_k is the resin's remaining life, starting at a_0, and f_k is fouling grown by the background mass rate $\dot{m}_{H,k}$ that enters column k while it loads. Cleaning removes fouling only while its solution is in contact and never restores a_k. The static capacity of new resin is $(1-\varepsilon) q_{max,0} = 65\,\mathrm{g/L}$ of packed bed.

11.3.3 Valves, Lines, Surge Tank and Pressure

A column's mode decides what feeds it, and its outlet destination d_k is set only at an event boundary. The flows follow the chain from the pumps, as shown in Equation (11.5).

$$\begin{aligned} Q_k &= Q_f[k \text{ lead}] + Q_r[k \text{ regenerating}] + \textstyle\sum_{j:\, d_j = k} Q_j \\ Q_k c_{i,in,k} &= Q_f c_{i,f}[k \text{ lead}] + \textstyle\sum_{j:\, d_j = k} Q_j c_{i,j}^{out} \\ \frac{\mathrm{d}c_{i,m}}{\mathrm{d}t} &= \frac{3Q_k}{V_c}\left(c_{i,m-1} - c_{i,m}\right), \quad m = 1,2,3 \end{aligned} \tag{11.5}$$

where the last line is the outlet line of volume V_c as three tanks, and $c_{i,f}$ leaves the feed line of volume V_f, built the same way. A switch moves valves and nothing else, and the liquid standing in a bed or a line goes wherever its valve now sends it.

The surge tank is well mixed, as shown in Equation (11.6).

$$\begin{aligned} \frac{\mathrm{d}V_s}{\mathrm{d}t} &= F_h - Q_f - F_{ov} \\ \frac{\mathrm{d}(V_s c_{i,s})}{\mathrm{d}t} &= F_h c_{i,h} - (Q_f + F_{ov}) c_{i,s} \\ c_{H,h}(t) &= c_{H,h,0}\,(1 + r t) \\ F_{ov} &= \max(0,\, F_h - Q_f) \end{aligned} \tag{11.6}$$

where F_h is the clarification's delivery until V_h has arrived and the overflow F_{ov} applies only while the tank overflows. A full tank either holds the clarification until it falls to 90 % or overflows to the drain, the learner's choice, and the feed pump stops at 1 L and restarts at 3 L. The pumps and the pressure at the train inlet are given in Equation (11.7).

$$\begin{aligned} \tau_f \frac{\mathrm{d}Q_f}{\mathrm{d}t} &= Q_{pump} \frac{u_P}{100}\left(1 - 0.04\, P_{in}\right) - Q_f \\ \Delta P_k &= \Delta P_{ref} \frac{Q_k}{Q_{ref}} (1 + k_p f_k) \\ P_{in} &= \Delta P_{line} \frac{Q_{lead}}{Q_{ref}} + \Delta P_{lead} + \Delta P_{lag} \end{aligned} \tag{11.7}$$

where P_{in} is in bar, $\tau_f = 4$ s, and P-201 is a stiff positive-displacement pump that delivers its command. The feed stops while P-201 pushes into the loading zone, so the train never carries both flows.

11.3.4 Instruments and the Outcome

Every UV detector reads the Beer-Lambert sum of both species, as shown in Equation (11.8).

$$a = \ell\,(\varepsilon_P c_P + \varepsilon_H c_H) + a_0^{UV} \tag{11.8}$$

through a 2 s lag, first-order colored noise of 0.0025 AU and saturation at 3 AU. FT-101 and FT-201 have 2 s lags and 0.5 % noise, PT-101 0.01 bar of noise and LT-001 0.3 %. The titer assay samples the feed every hour, reports 30 min later with 0.03 g/L of noise and holds its value until the next result. The manufacturing outcome is the antibody in the pool per liter of installed resin per hour of the whole campaign,

$$\begin{aligned} P &= \frac{m_{pool}}{3 V_{col}\, T} \\ m_{in} &= m_{held} + m_{pool} + \textstyle\sum_x m_x \end{aligned} \tag{11.9}$$

where T runs from the checkpoint to the last re-equilibration, m_{held} is the surge tank, lines and columns, and x runs over the flow-through, wash, cleaning, equilibration and drain exits. The ledger of the second line is checked at every step. El-Sabbahy et al. define productivity on the bound mass over the whole run time [4], and this simulator uses the pooled mass instead.

11.3.5 Parameters, Evidence and Initial State

Every parameter appears in Table 11.2 with the label of its source. Two values were read in the anchor papers, the antibody absorptivity at 280 nm [4] and the 2 mm UV path [3]. One, the nominal breakthrough shape of the adaptive policy, was fitted to this model's own fresh column. Everything else is a teaching assumption, and the kinetic constants were chosen only so that the breakthrough moves earlier and flattens as the residence time falls, the trend of El-Sabbahy et al.'s Figure 4, read

Table 11.2 Parameters. R Read in a Cited Source, F Fitted to This Model's Own Column, T Teaching Assumption, H Hidden From the Panel

Symbol	Meaning	Value	Unit	Src
L, D	bed length, diameter	0.10, 0.08	m	T
ε	interstitial fraction	0.35	—	T
$q_{max,0}$	new capacity (solid)	100	g/L	T
$k_{on,0}$	adsorption rate	1.5×10^{-3}	L/(g s)	T
$k_{off,0}$, $k_{off,el}$	desorption, load / elution	2×10^{-6}, 0.03	1/s	T
a	suppression by φ	12	—	T, H
$q_{max,H}$	background capacity	4	g/L	T, H
$k_{on,H}$	background adsorption	3×10^{-3}	L/(g s)	T, H
$k_{off,H}$	background desorption	3×10^{-3}	1/s	T, H
a_0	capacity left at start	1.0 (0.70, 0.85)	—	T
δ	capacity lost per cleaning	0.2 (0.4)	%	T
k_f	fouling gain	2×10^{-3}	L/g	T
τ_{CIP}	cleaning removal time	150	s	T
$D_{ax,ref}$	dispersion at 0.30 L/min	1.5×10^{-6}	m^2/s	T
V_c, V_f	outlet, feed line	0.02 (0.15), 0.05	L	T
ΔP_{ref}, ΔP_{line}	column, lines, 0.30 L/min	0.60, 0.20	bar	T
k_p	fouling on ΔP	2	—	T, H
Q_{pump}	pump capacity	1.0	L/min	T, H
ℓ	UV path length	0.2	cm	R
ε_P	antibody absorptivity	1.46	L/(g cm)	R
ε_H	background absorptivity	1.0	L/(g cm)	T, H
a_0^{UV}	detector zero	0.01	AU	T, H
$c_{P,h}$, $c_{H,h,0}$	titer, background	3.0, 1.2	g/L	T
F_h	harvest delivery	0.40	L/min	T
V_h	harvest still to come	90	L	T
$V_{s,0}$, $V_{s,max}$	surge start, working	25, 80	L	T
recipe	1.5, 3, 4, 2 CV, hold, 4 CV	5 min	0.40 L/min	T
m, s	nominal shape	1.02, 0.11	—	F
limits	loss, inlet pressure	1, 3	%, bar	T

qualitatively with no point digitized [4]. Each case starts from a documented checkpoint, three columns equilibrated in binding buffer with nothing bound, empty outlet lines, the feed line full of harvest, 25 L in the surge tank and the feed pump at its setpoint, with A as lead, B as lag and C ready.

The engine of Section 1.6.1 advances the state by fourth-order Runge-Kutta integration on a fixed step of 0.5 s. At 0.60 L/min the Courant number of a cell is 0.57, and the fastest binding relaxation, $k_{on,0} q_{max,0} (1 - \varepsilon)/\varepsilon$, is about 0.3 s^{-1}, both inside the method's stable range.

11.3.6 Assumptions

Film and pore diffusion in the beads are lumped into the kinetic rates, so the residence-time dependence of the breakthrough is a property of $k_{on,0}$ and not of a pore model, and it fails where intraparticle diffusion dominates. The buffers are one scalar φ, so pH, conductivity and the cleaning chemistry are not resolved, and the cleaning strips antibody exactly as elution does. Aging is a uniform loss of sites in every cell, although real aging also changes the kinetics and varies along the bed. The background is one species with no displacement by the antibody, which would bend the baseline near breakthrough. The outlet lines are three well-mixed tanks, whereas tubing is closer to plug flow. One regeneration pump serves one offline column at a time.

11.4 Loops, Switching Rules and the Supervisor

The controller executes every 2 s on the measurement packet of the instruments above, and the historian logs every 20 s (Section 1.6.1). No loop, estimator or live decision reads a model state.

11.4.1 FIC-101 and the PIC-102 Override

Both loops use the framework's velocity-form PID of Section 1.6.2, with back-calculation at a tracking time equal to τ_I. FIC-101 drives the feed pump from the applied flow setpoint, $\min(Q_{SP}, Q_{lim}, Q_{slow})$, with $K_c = 25\,\%$ per L/min, $\tau_I = 20\,\text{s}$, no derivative and an output of 0 to 100 %. PIC-102 is an override whose output is the flow limit Q_{lim}, with $K_c = 0.15$ L/min per bar, $\tau_I = 15\,\text{s}$ and a range of 0.02 to 1.0 L/min. While the inlet pressure is below its limit of 2.5 bar it tracks the request plus 0.02 L/min, so it cannot wind up. It takes over at 0.1 bar below the limit and lets go once its output exceeds the request by 0.02 L/min with the pressure 0.15 bar under the limit. Manual mode drives the pump by hand while the loop tracks it, so the return to automatic is bumpless (Section 1.6.2).

11.4.2 The Breakthrough Estimate

Total UV is not a product measurement, so the lead's breakthrough is estimated as an aligned ratio of product signals, as shown in Equation (11.10).

$$\begin{aligned} r_k &= \frac{a_{lead}(t) - \hat{B}}{\max\left(0.02,\; a_{in}(V_f - V_d) - \hat{B}\right)} \\ V_d &= \varepsilon V_{col} + V_c \\ \widehat{BT}_k &= \widehat{BT}_{k-1} + \left(1 - \mathrm{e}^{-\Delta t_c/\tau_{bt}}\right)\left(r_k - \widehat{BT}_{k-1}\right) \end{aligned} \tag{11.10}$$

where $a_{in}(V_f - V_d)$ is the feed UV that entered the lead one transit volume earlier ($V_d = 0.196\,\text{L}$), read from a buffer keyed by the totalized FT-101 volume, and $\tau_{bt} = 20\,\text{s}$. The background $\hat{B}$ is the lead outlet's UV captured once that outlet has risen by less than 0.004 AU/min for a minute (or after 10 min). It can be captured once for the campaign, the single-load practice of [3], at each new lead, or tracked through a 60 s filter on the lag's outlet. A plateau that leaves less than half the feed's product signal is breakthrough, not background, and the baseline the column showed as a fresh lag is kept instead. The lag's own outlet ratio against its baseline gives the end-of-train estimate.

11.4.3 Five Triggers and a Bound

The integrated load of the lead since it became lead is $\hat{L} = \sum Q_{f,m} c_{assay} \Delta t_c / V_{col}$, from the flow meter and the last reported titer. The triggers are a fixed time $t \geq t_{sw}$, an integrated load $\hat{L} \geq L^*$, a breakthrough threshold $\widehat{BT} \geq BT^*$ after 2 min of loading with the baseline captured, and the adaptive capacity of Equation (11.11). The fifth is a switch by hand, from a button on the FIC-101 card or the S key. The clock and the integrated load are the clock and estimate guards of Section 1.4.2, which Chapter 8 uses to end its coating layers.

$$\begin{aligned} x(b) &= m + s \ln \frac{b}{1-b} \\ \hat{C} &\leftarrow \hat{C} + \lambda\left(\frac{\hat{L}_{tot}}{x(\overline{BT}_{sw})} - \hat{C}\right) \\ \text{switch when } & \;\; \hat{L}_{pre} + \hat{L} \geq \hat{C}\, x(BT^*) \end{aligned} \tag{11.11}$$

where $\hat{L}_{pre}$ is the load the column caught as lag, integrated from the previous lead's $\widehat{BT}$ times the measured feed, $\hat{L}_{tot} = \hat{L}_{pre} + \hat{L}$ at the switch, $\overline{BT}_{sw}$ is the mean of $\widehat{BT}$ over the last 2 min before the switch, $\lambda = 0.5$ and $\hat{C}$ starts at 65 g/L. The capacity update is an exponentially weighted moving average of the observed capacities, updated once per switch, the estimator of the run-to-run controller of Section 1.4.5. An observation at either end of the shape's range, below 2 % or above 98 %, is used only as a bound. Because $\overline{BT}_{sw}$ is a two-minute average on a rising front, it reads below the breakthrough at the trigger, so the learned capacity is slightly high and the policy switches deeper than its target, as Section 11.7 measures. The true-breakthrough trigger reads the outlet line's concentration over the feed one transit volume earlier, which no instrument measures. It is a teaching bound, not a policy.

Table 11.3 Code Verification on the Shipped Engine

Check	Result
derivative identity, 11 runs, 110 valve states	worst 1.2×10^{-15} of gross flux
ledger of Eq. (11.9) at the end of each run	$\leq 7.9 \times 10^{-13}$ relative
antibody moved by a switch	0 in every switch
aging after n cycles against $a_0(1-\delta)^n$	within 2.2×10^{-16}
$t_{offline} \leq t_{available}$, hand against measured, 48 cells	agree in 48 of 48
fresh column, 10 % load at 10, 20, 40, 80 cells	49.63, 50.68, 50.94, 51.00 g/L
challenge, suggested, Δt against $\Delta t/2$	38.32 / 38.33 g/L/h, 0.706 / 0.705 %
challenge, suggested, 20 against 40 cells	38.32 / 38.34 g/L/h, 0.706 / 0.693 %
challenge ranking at Δt, $\Delta t/2$, 40 cells	identical
cycle to cycle, last three elutions against three before	0.01–0.69 % in 13 runs
campaign of 11.68 h, headless	1.7 s on one core

11.4.4 The Supervisor and Feasibility

Each column follows the mode graph ready, lag, lead, wash into the loading zone, wash, elution, cleaning, contact hold, re-equilibration and ready again, with a standby state used only at the end of the campaign. The graph is a hybrid automaton in the sense of Section 1.4.7. A step ends on the P-201 totalizer volume or on the hold clock, and a switch happens once, at an event boundary, when the trigger has fired and the third column is ready. The schedule is feasible only if $t_{offline} = \sum V_{step}/Q_r + t_{hold} \leq t_{available}$, 23.2 min for the shipped recipe and 23.6 min with the 20 to 22 s the wash into the train waits for the feed to stop. When the trigger fires early, the supervisor keeps loading the lead (its default), pauses the feed, or slows it ahead of time to $Q_{slow} = V_{rem}/(t_{ready} + 60\,\mathrm{s})$, where V_{rem} is the feed the policy predicts before its trigger and t_{ready} the regeneration still to run. The feed stops at the surge tank's low level, and when the harvest is through, the last two columns are eluted and cleaned in turn.

11.5 Code Verification and Agreement With the Literature

Code verification, agreement with the cited studies and educational evaluation are kept apart here, and only the first is quantitative. No learning study has been carried out, so no gain in learning is claimed. The checks in Table 11.3 run against the engine as shipped. An independent audit computed, at every step boundary of eleven disturbed runs, the rate of change of everything held plus every exit against the harvest inflow, from the raw state slots and the declared geometry. In each of the 110 valve states those runs exercised, the worst residual was 1.2×10^{-15} of the gross flux for the antibody. No antibody entered a column outside lead or lag, the pool received antibody only from an eluting column, and the count of elution steps equaled the cycles and the pool records in every run.

The feasibility condition was tested across titers of 2 to 5 g/L, flows of 0.2 to 0.5 L/min and resin at 1.0, 0.85 and 0.70 of new under integrated-load switching at 45 g/L. A hand prediction, the interval being the 1.88 min wash into the train plus the time the feed takes to bring 45 g/L of bed at the titer, agreed with the measurement on feasibility in all 48 cells, the measured margin lying 0.33 to 1.39 min above the hand value. At the reference flow of 0.30 L/min every column is ready with 4.7 min to spare. Cyclic operation, the cyclic steady state of Section 1.4.9, is shown rather than assumed. In each of 13 runs the mean of the last three elutions lies within 0.01 to 0.69 % of the three before, and successive elutions agree within 1 % from the second to the sixth cycle, after a first cycle that differs because the first lead starts fresh.

Agreement with the literature is qualitative, as shown in Figure 11.2. A fresh column at 0.125 L/min, 0.30 and 0.50, residence times of 4.0, 1.7 and 1.0 min, passes 10 % of its feed at 58.8, 50.7 and 41.1 g/L of bed, while the 50 % point stays at 66.1 g/L. The front moves earlier and flattens as the residence time falls, the trend that El-Sabbahy et al. report for residence times of 0.25 to 4 min [4] and that Hahn et al. report as a strong dependence of dynamic capacity on residence

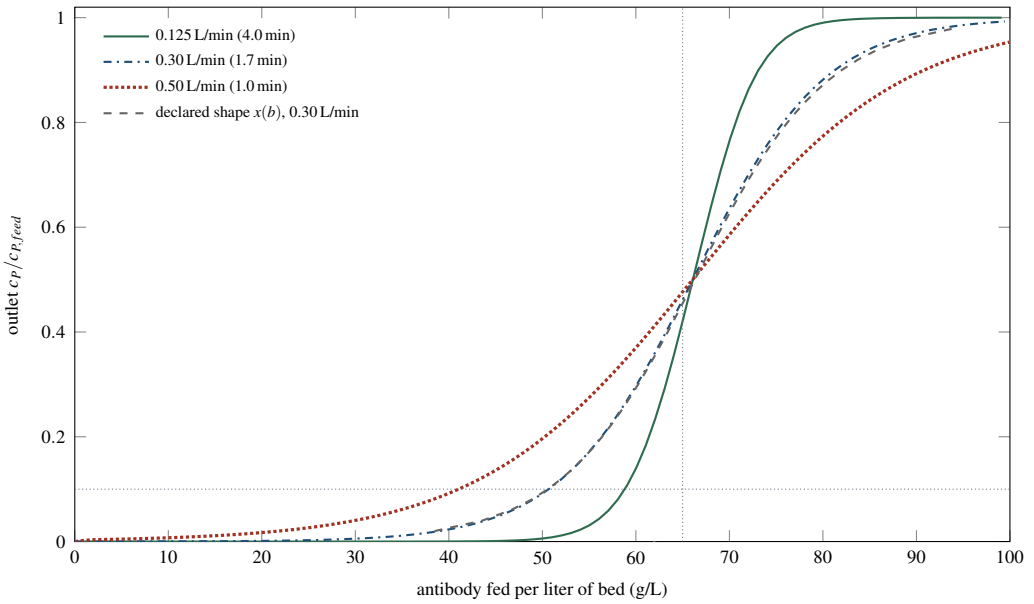


Figure 11.2 Breakthrough of one fresh column at three load flows, 0.125 L/min (solid), 0.30 L/min (dash-dotted) and 0.50 L/min (dotted). The front moves earlier and flattens as the residence time falls, while the 50 % point stays near the static capacity of 65 g/L. The dashed curve is the adaptive policy's declared shape, fitted to this model at 0.30 L/min.

time [1]. No breakthrough data were fitted, so the figure supports the direction of the effect and not its size for any resin.

11.6 Ten Cases, Measured

The cases appear in Table 11.4. All start from the reference configuration (integrated load at 45 g/L, 0.30 L/min, keep loading through a late column, the shipped recipe) and change only what their rows name. Every number in this section was measured headless at the default seed on the engine stepped exactly as in the browser, with any setting a row names entered before Run. Loaded at startup, selected after another case, or replayed after Reset, a case yields the same trajectory to the last logged digit. All 200 numbers that the app's own texts quote reproduce on the shipped engine. At the top playback speed of 900, the longest case, 11.68 h, plays in about 47 s. Each case is credited by a completion rule that requires both the learner's action and the resulting state of the train, so opening a case earns nothing, and a move made by the script never counts as the learner's. The challenge's verdict uses the same function in its scorecard tile, its completion rule and its tests.

11.6.1 Breakthrough Is Not Loss

Case 1 shows a conservative clock. At 0.20 L/min a 30-minute clock loads 31.9 g/L per switch, half the static capacity, and the campaign takes 11.68 h for 19.4 g/L/h, as listed in Table 11.5. The same clock at 0.30 L/min loads 48.1 g/L and makes 27.7 g/L/h, and at 0.40 L/min the lead takes 65.1 g/L and 1.52 % of the antibody leaves past the lag. Case 3 then separates breakthrough from loss. At a threshold of 90 % each lead comes off with 60.4 g/L instead of 54.6, the train switches 11 times instead of 12 and uses 0.278 instead of 0.299 L of buffer per gram, at a cost of 0.03 points of loss, as shown in Figure 11.3. The UV estimate runs about 2 points behind the truth at the trigger, because the 20 s filter sits on a front that climbs about 7 points a minute.

Table 11.4 The Ten Cases and the Decision Each Credits

#	Setup as shipped	Credited decision
1	fixed time 30 min, 0.20 L/min	raise the load flow without loss
2	fixed time 30 min, 0.35 L/min, titer +50 % at 1.5 h	switch on integrated load
3	breakthrough threshold 40 %	raise it to 85 to 95 %
4	load 45 g/L at 0.50 L/min	compare with 0.30 L/min
5	load 45 g/L on resin at $a_0 = 0.70$	switch on the column
6	threshold 20 %, background once, drift 60 %/h	keep the background current
7	0.40 L/min, 15 min cleaning hold, keep loading	pause or slow the feed
8	threshold 90 %, $V_c = 0.15$ L, no wash into the train	wash into the loading zone
9	150 L at 3.5 g/L arriving at 0.60 L/min, $a_0 = 0.85$	keep the antibody, hold upstream
C	120 L at 3.0 g/L, $a_0 = 0.85$, load 55 g/L at 0.45 L/min	meet the target

Table 11.5 Measured Outcomes, Default Seed, Whole Campaigns

#	Run	g/L/h	Lost (%)	Also
1	shipped, 0.20 L/min	19.4	0.03	31.9 g/L per switch, 11.68 h
	0.30 L/min	27.7	0.05	48.1 g/L, 8.18 h
	0.40 L/min	—	1.52	65.1 g/L per switch
2	shipped	35.6	8.57	up to 77.9 g/L per lead
	integrated load 45 g/L	39.1	0.10	waits 29.3 min
3	threshold 40 %	28.8	0.06	lag caught 21.8 g
	threshold 90 %	28.9	0.09	lag caught 78.3 g
4	shipped, 0.50 L/min	42.2	0.40	waits 49.1 min
	0.30 L/min	27.4	0.04	margin 4.7 min
5	shipped	26.8	2.10	lag caught 191.5 g
	threshold 60 %	26.8	0.10	17 switches, not 15
6	shipped (once)	23.1	0.04	20 switches, 0.47 L/g
	each lead	24.4	0.04	14 switches, 0.34 L/g
7	shipped	31.9	9.05	waits 90.4 min, 6.46 h
	pause / slow	23.3 / 22.6	0.05 / 0.04	9.75 / 10.01 h
8	shipped	30.0	3.01	2.95 in the wash
	wash 1.5 CV into the train	28.8	0.12	feed stopped 24.4 min
9	shipped	—	15.5	8.48 h
	pause / slow	—	0.17 / 0.09	held 91 / 121 min
C	shipped	38.4	4.62	fails on loss
	suggested	38.3	0.71	passes, 7.42 h

TRY IT IN THE APP 3 · The Breakthrough Threshold

Load the scenario and press Run. Watch the lead breakthrough climb to its 40 % target on the first trend strip, and the lead come off each time. At the end the scorecard shows 0.06 % of the antibody lost while the lag caught 21.8 g. Press Reset, set Breakthrough target to 90 % and run again: the lag catches 78.3 g and the loss is 0.09 %.

Pause. Try Exercise 11.1 before reading on.

11.6.2 Titer, Residence Time and Resin Age

A clock loads by time. In case 2 the richer lot puts up to 77.9 g/L on each lead against 56 before it and 8.57 % of the antibody is lost. Integrated load at 45 g/L follows the titer once the assay reports it, the first report coming at 2.5 h from the sample drawn at 2.0 h, and loses 0.10 %, but its trigger then comes in 21.6 min, faster than a column turns round, and the rotation waits 29.3 min. In case 4 a load flow of 0.50 L/min reaches the 45 g/L trigger in 15.1 min. With keep loading the leads run on to 58.2 g/L, the lag catches 96.9 g, and the campaign makes 42.2 g/L/h with 0.40 % lost. With the feed paused so that every lead carries only its 45 g/L, the lag still catches 21.6 g at 0.50 L/min against 5.3 g at 0.30, the per-pass cost of a shorter residence time. Case 5 runs the 45 g/L recipe on resin with 70 % of its capacity, 45.5 g/L of static capacity. The lag catches 191.5 g and still passes 1.96 % of the



Figure 11.3 Case 3 with the breakthrough target raised to 90 %, on the Columns view at $t = 2.50$ h, the step before the fourth switch, a composite of the app's header, its plant view and its first two trend strips. (1) Lead A is saturated (bound 58 of 58 g/L, true) and its last cell holds 92 % of the feed concentration, flowing to (2) lag B, whose front is a quarter of the way down (bound 14 of 61 g/L) while its outlet carries nothing. (5) The estimate reads 90 % against a true 91 %, and (3) C has been ready since the last regeneration. AT-101 reads 1.12 AU, AT-A 1.05 AU, AT-B 0.25 AU against a background estimate of 0.26 AU. The lag has caught 28.4 g, the pool holds 88.7 g in 6.0 L, and (4) 0.08 g has gone to the wastes and the flow-through. The trends show three earlier leads climbing to the target with the lag outlet near zero.

antibody, while a 60 % threshold loses 0.10 % at the same 26.8 g/L/h, because the feed flow sets the rate.

11.6.3 Background, Feasibility and Dead Volume

In case 6 the harvest background rises 60 % of its start per hour, reaching 2.2 times its start in the feed, and a baseline captured once reads 26 % for late leads whose true breakthrough is 1 %. The train switches 20 times instead of 14 at 33.8 g/L instead of 47.3, and uses 0.47 instead of 0.34 L/g of buffer, while losing only 0.04 %, because an estimate that reads high errs toward wasted resin. In case 7 a 15 min cleaning hold makes the recipe 33.2 min while the 45 g/L trigger comes after 18.9 min. Keep loading loses 9.05 %, pausing loses 0.05 % in 9.75 h and slowing 0.04 % in 10.01 h. In case 8 the liquid standing in a lead switched at 90 % is nearly feed, and with 0.15 L outlet lines and no wash into the train 3.01 % is lost, 2.95 % of it in the wash. Washing 1.5 column volumes into the new lead cuts that to 0.12 % at the cost of 24.4 min of paused feed.

🕮 **Pause.** Try Exercise 11.3 before reading on.

Table 11.6 Switching Policies on the Challenge Plant With the Feed Paused for a Late Column, Mean Over Six Paired Seeds

Policy	P g/L/h	Rec. %	FT %	Sw. %	Util. %	BT_{sw} %	Buffer L/g	Wait min
time 25 min	37.23	90.42	8.61	0.24	100.2	101.4	0.297	0
load 55 g/L	38.28	94.52	4.53	0.23	99.1	96.4	0.302	0
load 45 g/L	34.74	99.09	0.11	0.07	88.1	48.0	0.347	55.5
threshold 60 %	35.16	99.04	0.18	0.06	92.8	62.4	0.339	43.4
threshold 85 %	38.32	98.57	0.54	0.16	98.2	87.5	0.306	23.6
adaptive 60 %	36.59	98.94	0.26	0.08	94.8	70.9	0.322	38.2
adaptive 85 %	37.98	98.38	0.76	0.14	98.8	91.1	0.307	24.3
bound, true 85 %	36.54	98.79	0.42	0.07	97.4	83.0	0.323	28.1
bound, true 90 %	38.25	98.55	0.57	0.16	98.3	88.5	0.306	23.6

P captured per liter of resin per hour, Rec. recovery, FT flow-through loss, Sw. switching losses, Util. bound over static capacity at the switch (teaching truth), BT_{sw} true breakthrough at the switch, Wait the feed paused for a late column. Across seeds the standard deviation of P is at most 0.01 g/L/h for the clock and the UV-based rules and 0.25 and 0.89 g/L/h for the two load rules, which read the noisy assay.

> **TRY IT IN THE APP 6 · A UV Background That Drifts**
>
> Load the scenario and press Run. Late in the campaign the breakthrough estimate on the first trend strip reads far above the dashed truth, and the train ends with 20 switches and 0.47 L of buffer per gram while losing 0.04 %. Press Reset, set UV background (Advanced) to Each lead and run again: 14 switches and 0.34 L per gram.

11.6.4 A Harvest the Train Cannot Take

Case 9 delivers 150 L at 3.5 g/L at 0.60 L/min, 2.1 g of antibody a minute, onto resin at 85 %. Three columns that each turn round in 23.6 min with about 55 g/L of capacity can take about 1.2 g a minute. As shipped the train keeps loading and loses 15.5 %. Pausing keeps the antibody (0.17 % lost) but holds the clarification for 91 min, slowing holds it 121 min, and overflowing to the drain loses 15.47 %. A logged sweep of 60 settings (load flows from 0.30 to 0.54 L/min, integrated load, adaptive capacity, thresholds of 85 and 95 % and the true-breakthrough bound, with each of the three responses) finds none with the loss within 1 % and the surge tank never full. The closest, an 85 % threshold with the feed paused, loses 0.71 % and still holds the clarification 57 min. With this regeneration recipe the harvest rate is infeasible, and the case credits keeping the antibody and holding upstream, not a rate. The insulin train of Volume 19, Chapter 9 also counts an infeasible harvest among its cases, set by a refolding vessel that is the bottleneck of its train.

11.7 Policies Under Equal Resin and Feed

The policy comparison uses the challenge plant, a 120 L harvest at 3.0 g/L arriving at 0.33 L/min onto resin at $a_0 = 0.85$ that loses 0.4 % per cleaning, fed at 0.45 L/min. Every policy has the same instruments, recipe, feed and resin, and the same response to a late column, Pause the feed, so only the switching rule differs. Each row is the mean over six paired seeds (the default and 1 to 5), whose noise realizations every policy shares. Recovery is the pool over the antibody delivered. Utilization is the bound antibody in the outgoing lead at each switch over its static capacity at that moment, averaged over switches. Switching losses are the wash, cleaning and equilibration wastes, apart from the flow-through.

The measured policies appear in Table 11.6. The clock and the 55 g/L load run the lead past saturation, 100.2 and 99.1 % bound, and lose 8.61 and 4.53 % in the flow-through, so their productivity is high only because the antibody they lose is not slowing them down. The 45 g/L load is safe at 99.09 % recovery but leaves 12 % of the capacity unused, waits 55.5 min for regeneration, and uses 0.347 L of buffer per gram against 0.306 for the 85 % threshold. The 85 % threshold makes 38.32 g/L/h with 98.57 % recovery and 98.2 % utilization, and is the only measured rule that meets the challenge at all six seeds besides the adaptive 85 % rule (37.98 g/L/h, 98.38 %) and the 90 %

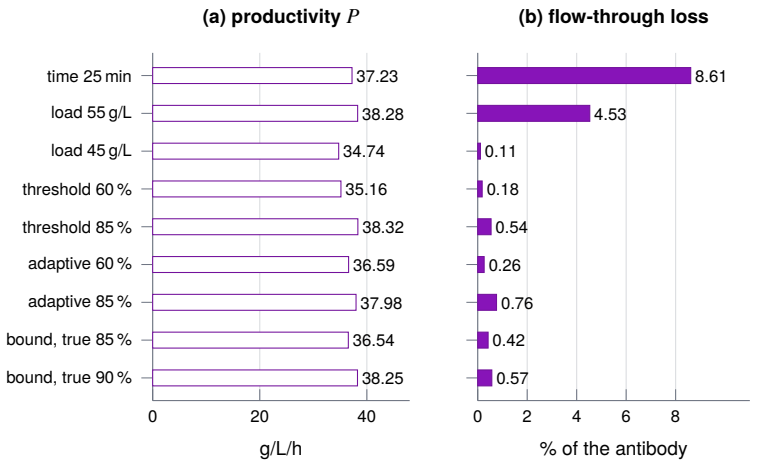


Figure 11.4 The switching policies of Table 11.6 on the challenge plant with the feed paused for a late column, mean over six paired seeds: (a) antibody captured per liter of resin per hour, (b) the share lost in the flow-through. Notice that the clock and the 55 g/L load are not more productive than the 85 % threshold, although they load the lead past saturation and lose far more.

bound. Switching losses stay between 0.06 and 0.24 % for every rule, so the flow-through, not the switch, carries the loss. Figure 11.4 draws the productivity and the flow-through loss of every policy side by side.

> ✦ DID YOU KNOW? **A Saturated Lead Does Not Pay**
>
> From the simulator: on the challenge plant a 25-minute clock takes each lead to 100.2 % of its static capacity and loses 8.61 % in the flow-through, yet makes 37.23 g/L/h, less than the 38.32 g/L/h of an 85 % breakthrough threshold that loses 0.54 % (Table 11.6).

Two departures from intuition follow from the information each rule has. The true-breakthrough bound at 85 % does worse than the UV estimate at 85 % (36.54 g/L/h against 38.32), because the estimate trails the front and so realizes 87.5 %. The bound is a bound only at a matched realized breakthrough, and at 90 % it realizes 88.5 % and matches the estimate within 0.07 g/L/h. The adaptive rule switches deeper than its target, at a true 91.1 % for a target of 85 %, because it learns from the two-minute average described in Section 11.4. Two ablations isolate the parts. With the learning gain at zero the adaptive rule keeps its starting guess of 65 g/L and loses 2.56 % at an 85 % target, against 0.90 % with learning. Dividing by the current rather than the aligned inlet changes nothing on this plant (0.70 % lost either way), because its titer is constant and the surge tank smooths the feed.

🕮 **Pause.** Try Exercise 11.2 before reading on.

The resin-age sweep shows what uncertain capacity changes, as shown in Figure 11.5. On new resin the 25-minute clock passes the challenge with 40.76 g/L/h and 0.28 % lost, and the 55 g/L load passes with 40.19 g/L/h and 0.13 %. At $a_0 = 0.925$ the clock already loses 2.43 %, and at 0.85 only the breakthrough, adaptive and bound rules pass. At $a_0 = 0.775$ none does. The 85 % threshold loses 1.05 % and makes 34.68 g/L/h, the 45 g/L load loses 0.65 % at 33.77 g/L/h, and the threshold's paused time grows from 0 on new resin to 50.7 min because the switches come faster than the regeneration as the capacity shrinks. The best rule therefore changes from a clock on new resin to a measured breakthrough on aged resin, and on still older resin no switching rule recovers the target with this recipe.

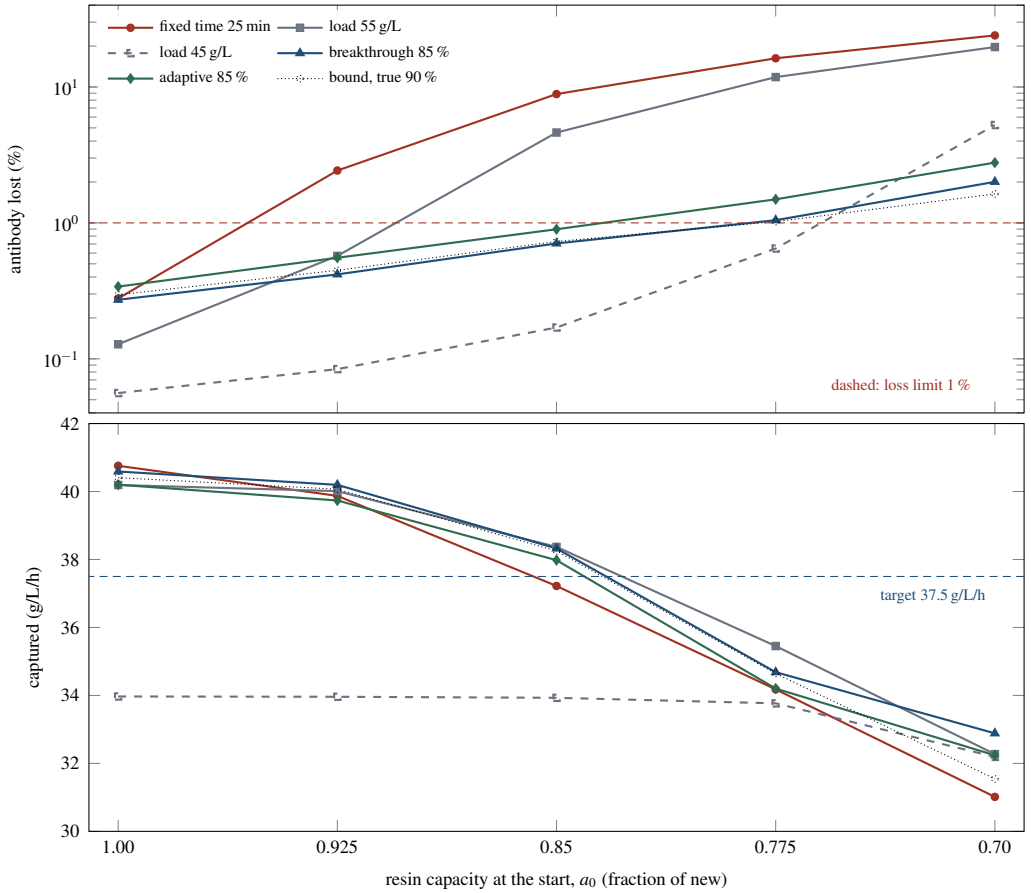


Figure 11.5 Loss and productivity against the resin's starting capacity, the challenge harvest with the feed paused for a late column, default seed. The fixed clock and the 55 g/L load pass only on resin near new. The breakthrough and adaptive rules keep the loss within 1 % down to $a_0 = 0.85$, and at 0.775 the regeneration waits, not the rule, decide the productivity.

11.8 Technology in Practice

What Measures It

The switching decision reads two kinds of measurement, in-line UV detectors around the columns and a titer assay. In-line UV flow cells that monitor therapeutic protein concentration are one of the examples the ICH guideline on continuous manufacturing gives of process analytical technology [15]. The titer was long measured offline for every lot to set the load volume. At-line monitoring carries the risk of human error, contamination, time delays and missing data, and an automated analytical chromatograph is expensive and error-prone, and its delay between sample and result risks a late reaction or forces a slower process [3]. The assay of this chapter, an hourly sample reported 30 min later, is that delay, and case 2 shows the integrated-load rule waiting for it.

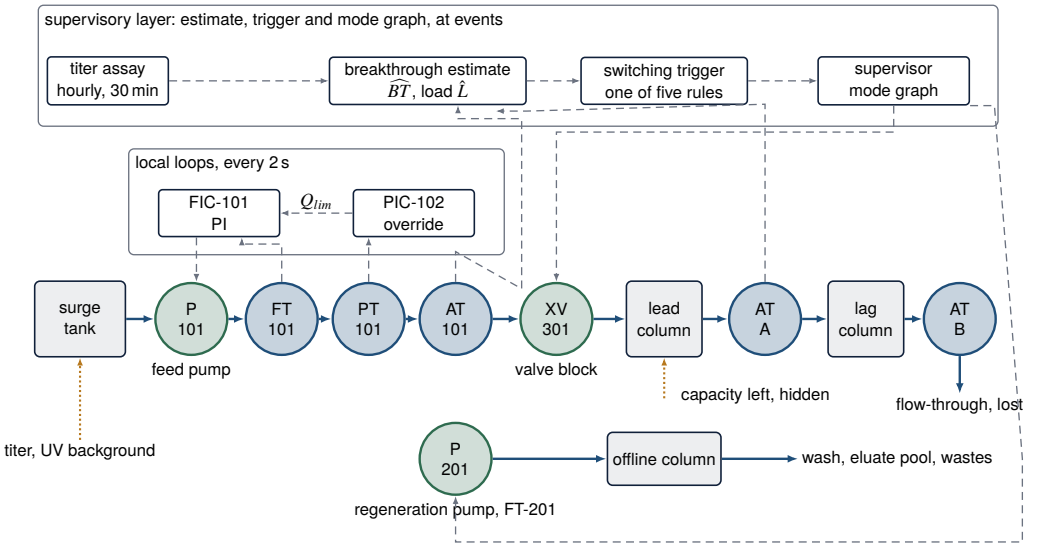


Figure 11.6 The chapter's loops on capture hardware. In the local layer FIC-101 drives the feed pump P-101 toward its flow setpoint, and the PIC-102 override lowers the flow limit when PT-101 nears its limit. In the supervisory layer the UV detectors before and after the lead and the titer assay feed the breakthrough and load estimates, a trigger decides the switch, and the supervisor moves the valve block and runs the regeneration pump P-201 on the offline column. Dotted arrows mark disturbances. XV-301 is the book's tag for the outlet manifold; the others are the chapter's.

TECHNOLOGY **Sensor: Two UV Detectors Around the Lead Column**

The antibody and the impurities of a harvest both absorb at 280 nm, so one wavelength cannot tell them apart. A UV method for continuous Protein A capture places one detector at the column inlet and one at its outlet. While the column loads, the outlet signal settles at an impurity baseline, and once the antibody breaks through it rises above that baseline by an amount that corresponds to the breakthrough level. Switching on that difference loads every column to the same percentage breakthrough whatever the titer or the capacity left. Two detectors can drift unequally, and impurities displaced by the antibody can keep the baseline from settling [3]. One laboratory setup carried a UV monitor with a 2 mm path and a diode array flow cell with a 0.4 mm path [3]. AT-101 and AT-A play the two detectors here.

What Moves It

Three final elements act on the train: the feed pump P-101, the regeneration pump P-201 and the valve block that sets where each column's inlet and outlet lead. Figure 11.6 draws them with the loops that move them. The model gives the feed pump a 4 s lag and a delivery that falls with the inlet pressure, and treats P-201 as a stiff positive-displacement pump that delivers its command.

> **TECHNOLOGY** **Final Element: The Valve Block of the Train**
>
> In a multi-column capture system the pumps feed the inlet ports of a valve block, and the block's software sets the inlets and outlets of each column, including two columns run in series. In one four-column laboratory system five pumps fed four inlet ports. Its steady-state cycle was split into four switches, each moving the feed on to the next column, and each switch into four steps that changed the wash, elution, strip or equilibration positions of the two columns not loading [4]. In this chapter the outlet manifold moves only at event boundaries, so a switch moves valves and no liquid, and case 8 measures what the liquid standing in the bed then costs.

Where the Controller Runs

In one laboratory study the chromatography system's own control software ran the method, and a script on a separate computer read the detector spectra and sent a signal back to end a phase once a set condition was met [3]. At plant scale a distributed control system holds a supervisory level that oversees localized controllers, PLCs serve as local controllers within such a scheme, and the loops of both levels run with cycle times from milliseconds to minutes [18]. The chapter's controller executes every 2 s and its historian logs every 20 s. Read as plant hardware, FIC-101 and PIC-102 belong in the local controller beside the pump, and the breakthrough estimate, the five triggers and the mode graph of the supervisor belong in the supervisory layer, which acts at events.

> **IN PRACTICE** **A State of Control That Is Not a Steady State**
>
> The ICH guideline on continuous manufacturing lets the size of a batch be defined by the quantity of output, the quantity of input or the run time at a defined mass flow rate. A continuous process can be in a state of control while its parameters change within specified ranges without being at steady state, and the plant should notice when they drift or trend inside those ranges and find the root cause. The guideline's own example is a change in the elution profile of a therapeutic-protein process that may be due to resin aging [15]. The cyclic steady state of Section 1.4.9 and the aging of Equation (11.4) are this chapter's versions of both ideas.

What Is Modernizing It

Three developments touch this train. The first is the selective soft sensor: a partial least squares model on UV and visible spectra predicted the antibody in the effluent with a root mean square error of 0.01 mg/mL and ended a load at 0.15 mg/mL, about 5 % breakthrough [3]. The second is new optics. A multi-pass Raman flow cell has monitored the breakthrough of a Protein A column, although advanced chemometrics and a large computational effort were needed to reach a usable model [19]. The third is model-based control: model predictive control has been developed for multi-column solvent-gradient purification, a continuous chromatography process [19], and process analytical technology that detects disturbances in real time makes continuous processes amenable to feedforward and feedback control [15]. Each would give the switching rule of this chapter a better reading than total UV, or a model to act on it.

11.9 Limitations

Every coefficient except the antibody absorptivity and the UV path is a teaching value, no breakthrough or elution data were fitted, and the agreement with the literature is the direction of the residence-time effect only. Productivity, utilization and buffer per gram are properties of this synthetic resin, recipe and column and do not predict any commercial medium. The pore and film resistances are lumped into the kinetic rates, the buffers are one scalar, and aging is uniform in the bed. The disturbance steps of the cases are larger than a plant would show (a 50 % titer step, 60 % more background per hour), but each isolates one cause. The background is one species, so a selective spectral estimate [3] would face a different and smaller error than the total-UV estimate here.

Capture alone is not a purification. Host-cell protein and DNA clearance, aggregates, leached ligand, low-pH virus inactivation and the polishing steps are outside the model, and viral safety of a continuous process follows the principles of ICH Q5A as Q13 restates [15]. The 1 % loss limit, the 3 bar pressure limit and the surge volume are teaching values. Real limits come from the product's specification, the column hardware and the site's validation, and passing in simulation does not release a batch. The adaptive rule's learning on a two-minute average makes it switch about 6 points deeper than its target, and the supervisor's surge-tank hold reads the true volume rather than LT-001. Each result holds for one plant and its seeds, and the six seeds move the UV-based rules by at most 0.01 g/L/h.

OPEN CHALLENGE **Reading the Capacity a Column Has Left**

A Protein A column loses capacity over its cycles, through leaching and degradation of the ligand and through pores and ligands blocked by impurities or product left behind, so batch plants load to a conservative capacity found in resin-lifetime studies [3]. Fluorescence through the bed has shown the fouling of Protein A resin, but a direct correlation with it is difficult [19]. In this chapter the capacity left is hidden from every instrument (Table 11.1), and the adaptive rule learns it from the breakthrough of earlier leads. Progress would be an in-line reading of the capacity a column has left, so that a switching rule could learn from a measurement of the resin itself.

11.10 Conclusion

This chapter has set out a three-column Protein A train as a 380-state simulator whose antibody balance closes to 1.2×10^{-15} of the gross flux in every valve state tested, and whose regeneration must finish within one loading interval. Its central lesson is that breakthrough the lag catches is not loss. Switching at 90 % instead of 40 % raises what the lag catches from 21.8 to 78.3 g for 0.03 points of loss. Under equal resin and feed on resin at 85 % of its capacity, an 85 % UV threshold makes 38.32 g/L/h with 98.57 % recovery, the learned capacity 37.98 g/L/h, a safe load 34.74 g/L/h with 12 % of the capacity unused, and a 25-minute clock loses 8.86 %. On resin at 77.5 % of its capacity no rule tested meets the target, because the regeneration sets the pace. The kinetics are qualitative and every limit is a teaching value. Further work would fit the binding law to measured breakthrough at several flows, add pore diffusion, a selective spectral estimate, a fourth column and the low-pH hold. The simulator shows that a capture schedule is as good as what it knows about the resin and the time the regeneration needs.

KEY POINTS

- Antibody that breaks through the lead column is caught by the lag, so breakthrough the lag catches is not loss.
- Switching at 90 % instead of 40 % raises what the lag catches from 21.8 to 78.3 g for 0.03 points of loss.
- On resin at 85 % of its capacity an 85 % UV threshold makes 38.32 g/L/h with 98.57 % recovery, and a 25-minute clock loses 8.86 %.
- On resin at 77.5 % of its capacity no rule tested meets the target, because the regeneration sets the pace.

Where else this applies. The clock and integrated-load triggers that end a lead here are the guards that end the coating layers of Chapter 8 (Section 11.4).

References

[1] R. Hahn, R. Schlegel, and A. Jungbauer, "Comparison of protein A affinity sorbents," *Journal of Chromatography B*, vol. 790, no. 1–2, pp. 35–51, 2003.

[2] R. Godawat, K. Brower, S. Jain, K. Konstantinov, F. Riske, and V. Warikoo, "Periodic counter-current chromatography – design and operational considerations for integrated and continuous purification of proteins," *Biotechnology Journal*, vol. 7, no. 12, pp. 1496–1508, 2012.

[3] M. Rüdt, N. Brestrich, L. Rolinger, and J. Hubbuch, "Real-time monitoring and control of the load phase of a protein A capture step," *Biotechnology and Bioengineering*, vol. 114, no. 2, pp. 368–373, 2017.

[4] H. El-Sabbahy, D. Ward, O. Ogonah, L. Deakin, G. M. Jellum, and D. G. Bracewell, "The effect of feed quality due to clarification strategy on the design and performance of protein A periodic counter-current chromatography," *Biotechnology Progress*, vol. 34, no. 6, pp. 1380–1392, 2018.

[5] H. C. Thomas, "Heterogeneous ion exchange in a flowing system," *Journal of the American Chemical Society*, vol. 66, no. 10, pp. 1664–1666, 1944.

[6] V. Warikoo, R. Godawat, K. Brower, S. Jain, D. Cummings, E. Simons, T. Johnson, J. Walther, M. Yu, B. Wright, J. McLarty, K. P. Karey, C. Hwang, W. Zhou, F. Riske, and K. Konstantinov, "Integrated continuous production of recombinant therapeutic proteins," *Biotechnology and Bioengineering*, vol. 109, no. 12, pp. 3018–3029, 2012.

[7] D. Baur, M. Angarita, T. Müller-Späth, F. Steinebach, and M. Morbidelli, "Comparison of batch and continuous multi-column protein A capture processes by optimal design," *Biotechnology Journal*, vol. 11, no. 7, pp. 920–931, 2016.

[8] F. Steinebach, M. Angarita, D. J. Karst, T. Müller-Späth, and M. Morbidelli, "Model based adaptive control of a continuous capture process for monoclonal antibodies production," *Journal of Chromatography A*, vol. 1444, pp. 50–56, 2016.

[9] P. Cuatrecasas, M. Wilchek, and C. B. Anfinsen, "Selective enzyme purification by affinity chromatography," *Proceedings of the National Academy of Sciences*, vol. 61, no. 2, pp. 636–643, 1968.

[10] H. Hjelm, K. Hjelm, and J. Sjöquist, "Protein A from *Staphylococcus aureus*. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins," *FEBS Letters*, vol. 28, no. 1, pp. 73–76, 1972.

[11] S. Hober, K. Nord, and M. Linhult, "Protein A chromatography for antibody purification," *Journal of Chromatography B*, vol. 848, no. 1, pp. 40–47, 2007.

[12] A. A. Shukla, B. Hubbard, T. Tressel, S. Guhan, and D. Low, "Downstream processing of monoclonal antibodies—Application of platform approaches," *Journal of Chromatography B*, vol. 848, no. 1, pp. 28–39, 2007.

[13] J. Pollock, G. Bolton, J. Coffman, S. V. Ho, D. G. Bracewell, and S. S. Farid, "Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture," *Journal of Chromatography A*, vol. 1284, pp. 17–27, 2013.

[14] A. Jungbauer, "Continuous downstream processing of biopharmaceuticals," *Trends in Biotechnology*, vol. 31, no. 8, pp. 479–492, 2013.

[15] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, "ICH harmonised guideline Q13: Continuous manufacturing of drug substances and drug products," Adopted 16 November 2022, 2022. [Online]. Available: https://database.ich.org/sites/default/files/ICH_Q13_Step4_Guideline_2022_1116.pdf

[16] P. V. Danckwerts, "Continuous flow systems," *Chemical Engineering Science*, vol. 2, no. 1, pp. 1–13, 1953.

[17] B. van Leer, "Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme," *Journal of Computational Physics*, vol. 14, no. 4, pp. 361–370, 1974.

[18] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

[19] L. Rolinger, M. Rüdt, and J. Hubbuch, "A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing," *Analytical and Bioanalytical Chemistry*, vol. 412, no. 9, pp. 2047–2064, 2020.

11.11 Exercises

11.1 Concept check · 2 min

In case 3 the breakthrough target of the lead column is raised from 40 % to 90 % of the feed, with nothing else changed. What happens to the share of the antibody the train loses over the campaign?

(a) It rises about in proportion to the extra breakthrough, because breakthrough is antibody leaving the lead

(b) It falls, because a deeper switch means fewer switches and so less antibody sent to the wash

(c) It rises by several points, because the lag saturates once the lead passes most of its feed

(d) It barely changes, because the lag catches what the lead lets through on capacity still fresh

11.2 Concept check · 2 min

On the challenge plant, with the feed paused for a late column, the adaptive capacity rule at a breakthrough target of 85 % takes its leads off at a true breakthrough of 91.1 % on average. What makes it switch deeper than its target?

(a) It learns from a two-minute average of the estimate, which reads below the breakthrough at the trigger

(b) Its starting guess of 65 g/L is the capacity of new resin, and a learning gain of 0.5 never forgets it

(c) The titer assay reports 30 min late, so the integrated load it compares with the capacity runs high

(d) The UV background drifts upward through the campaign, so the estimate reads low on the late leads

11.3 Concept check · 1 min

In case 6 the UV background of the harvest rises through the campaign, and the breakthrough estimate subtracts a background captured once from the first lead. What does the train do?

(a) It switches later than it should and loses antibody in the flow-through

(b) It switches earlier than it should and wastes resin and buffer, with little loss

(c) It switches as before, because the aligned feed UV in the ratio carries the same background

(d) It switches as before, because the end-of-train estimate on the lag corrects the lead's estimate

11.4 Theory and calculation · 6 min

The shipped regeneration recipe pushes 1.5 column volumes into the loading zone, then 3 of wash, 4 of elution, 2 of cleaning, a 5 min contact hold and 4 of re-equilibration, at 0.40 L/min. A column holds 0.5027 L. The feed pauses during the first step, and the integrated-load trigger is set at 45 g/L with a titer of 3.0 g/L.

(a) Find the offline time $t_{offline}$ of one column. *Checkpoint:* 23.2 min.

(b) Find the time FT-101 takes to feed 45 g/L of bed at 0.30, 0.40 and 0.50 L/min. *Checkpoint:* 25.13, 18.85 and 15.08 min.

(c) The time a column has offline is the wash into the train plus the feed time of the next lead. Find the margin at each flow by this hand rule, and the highest load flow it calls feasible. *Checkpoint:* 3.8, −2.5 and −6.3 min, and about 0.35 L/min.

11.5 Theory and calculation · 8 min

New resin offers a static capacity $(1-\varepsilon)q_{max,0} = 65\,\mathrm{g/L}$ of bed, and Equation (11.4) scales it by the remaining life a. A column holds 0.503 L and turns round offline in 23.6 min.

(a) Find the static capacity of resin at $a_0 = 0.85$ and at $a_0 = 0.70$, and the antibody one column holds at $a_0 = 0.85$. *Checkpoint:* 55.25 and 45.5 g/L, and 27.8 g.

(b) Case 9 delivers 150 L of harvest at 3.5 g/L at 0.60 L/min onto resin at $a_0 = 0.85$, and the surge tank starts at 25 L with a working volume of 80 L. Bound the antibody per minute the train can take, and estimate when the surge tank fills if the train runs at that bound. *Checkpoint:* 1.18 g/min against 2.1 g/min, full after about 209 min of the 250 min the harvest takes to arrive.

(c) The adaptive rule starts at $\hat{C} = 65\,\mathrm{g/L}$ with $\lambda = 0.5$. If every observation were 55.25 g/L, after how many switches would $\hat{C}$ be within 1 g/L of it? *Checkpoint:* 4 switches.

11.6 Theory and calculation · 7 min

The breakthrough estimate of Equation (11.10) is filtered with $\tau_{bt} = 20\,\mathrm{s}$ at the 2 s scan, and the adaptive rule learns from its mean over the last two minutes. Assume the front passes the target

climbing steadily at 7 points a minute, as in case 3, and that the nominal shape $x(b) = m + s\ln[b/(1-b)]$ with $m = 1.02$ and $s = 0.11$ holds.

(a) Find the filter coefficient and how far the filtered estimate trails the true breakthrough on the ramp. *Checkpoint:* 0.0952, and about 2.3 points.

(b) The trigger fires when the filtered estimate reaches 85 %. Find the true breakthrough at that instant and the two-minute mean the rule learns from. *Checkpoint:* About 87.3 and 78 %.

(c) Find the factor by which the capacity learned from that switch is too high, taking x at the true breakthrough as right, and the true breakthrough at which a switch at $\hat{C}\,x(0.85)$ then comes. *Checkpoint:* 1.063, and about 92 %.

11.7 App challenge · 8 min

Load scenario 3, The breakthrough threshold, and change no setting. Set the speed to 900, at which one campaign of about 7.9 h plays in about 31 s, and use the default seed. Each run ends when the campaign finishes and the run pauses itself.

(a) Run the scenario as shipped, with the target at 40 %. Read the tiles Antibody lost, Breakthrough the lag caught, the switch count on Regeneration margin, worst switch, and Buffer per gram captured. *Checkpoint:* 0.06 %, 21.8 g, 12 switches, 0.30 L/g.

(b) Press Reset, set the Breakthrough target to 90 % and run again. Read the same four tiles. *Checkpoint:* 0.09 %, 78.3 g, 11 switches, 0.28 L/g.

Accept the tiles exactly as stated. Find how many grams more the lag caught and how many grams more were lost out of the 345 g the harvest delivered, and say what the deeper switch bought.

11.8 App challenge · 8 min

Load scenario 7, The next column is not ready, in which a 15 min cleaning hold makes the offline recipe 33.2 min while the 45 g/L trigger comes after 18.9 min. Set the speed to 900 and use the default seed. Run each case until the campaign finishes and the run pauses itself. For each run read Antibody lost and Captured per liter of resin per hour, and the campaign hours in the tooltip of the Captured tile.

(a) Run the scenario as shipped, with If the next column is not ready on Keep loading. *Checkpoint:* 9.05 %, 31.9 g/L/h, 6.46 h.

(b) Press Reset, set If the next column is not ready to Pause the feed and run again. *Checkpoint:* 0.05 %, 23.3 g/L/h, 9.75 h.

(c) Press Reset, set it to Slow the feed and run again. *Checkpoint:* 0.04 %, 22.6 g/L/h, 10.01 h.

Accept the tiles exactly as stated. Find the antibody that pausing saves against keeping on loading, out of the 345 g delivered, and the hours it costs.

11.9 App challenge · 9 min

Load the challenge, The most antibody per liter of resin per hour. Its target is at least 37.5 g/L/h with at most 1 % lost, the inlet under 3 bar and the surge tank never full. Set the speed to 900, use the default seed and do not change the model or the disturbances. Run each case until the campaign finishes and the run pauses itself, and read Captured per liter of resin per hour, Antibody lost and Challenge target.

(a) Run the challenge as shipped. *Checkpoint:* 38.4 g/L/h, 4.62 %, not met.

(b) Press Reset, set the Switching strategy to Breakthrough threshold, the Breakthrough target to 85 % and If the next column is not ready to Pause the feed, and run again. *Checkpoint:* 38.3 g/L/h, 0.71 %, met.

(c) Press Reset, set If the next column is not ready back to Keep loading, leaving the threshold at 85 %, and run again. *Checkpoint:* 38.7 g/L/h, 3.95 %, not met.

Accept the tiles exactly as stated. Say which decision fixed the loss.

11.10 Application area · 3 min

A plant manager prefers the integrated load at 45 g/L as the safe recipe for the challenge plant, against the 85 % breakthrough threshold. Use the policy comparison of Table 11.6.

(a) Find the extra buffer per kilogram of captured antibody that the 45 g/L load uses, as liters and as a share. *Checkpoint:* 41 L, 13 % more.

(b) Find the resin-hours each rule needs per kilogram captured, and the share of extra resin the 45 g/L load needs for the same output. *Checkpoint:* 28.8 against 26.1 liter-hours, about 10 % more.

11.11 Application area · 3 min

In case 8 the train switches 11 times at a 90 % breakthrough target, with outlet lines of 0.15 L and no wash into the loading zone, and 345 g of antibody enters the train. A column of 0.503 L has an interstitial fraction of 0.35, and the liquid in a lead at such a switch is close to the feed's 3.0 g/L.

(a) Estimate the antibody standing in the bed and in its outlet line at one switch. *Checkpoint:* About 0.53 and 0.45 g.

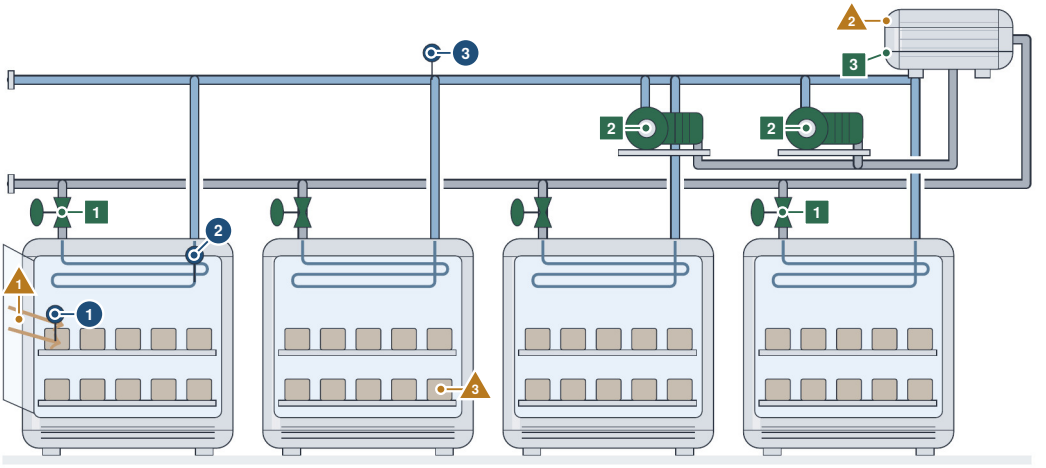
(b) Estimate the share of the campaign's antibody that goes to the wash over all switches, compare it with the measured loss, and say how plants recover it. *Checkpoint:* About 3.1 % against 2.95 % in the wash.

11.12 Review the agent · 5 min

An AI assistant comparing switching rules on the challenge plant writes: "The true-breakthrough trigger knows the lead's real breakthrough, which no instrument measures, so it is the best any rule can do. Set it at 85 % with the feed paused for a late column to find the ceiling. A UV threshold at the same 85 % only estimates the breakthrough, so it must make less antibody per liter of resin per hour." Find the flaw and say how you would check it.

CHAPTER 12

Thermostats That Fall Into Step on a Simulated Supermarket Refrigeration Pack



The controller moves 1 the valve into each case, opened and closed by its thermostat and its superheat loop 2 the compressors started and stopped on the rack 3 the speed of the condenser fans

The controller reads 1 the air in each case 2 the superheat at each coil outlet 3 the shared suction pressure

What disturbs it 1 a door held open 2 a hot afternoon at the condenser 3 warm goods restocked

Four display cases share one suction line and one rack of compressors (two of the chapter's three are drawn), so notice that the four valves open and close against one pressure. Illustration, not to scale.

IN PLAIN TERMS Display Cases That Fall Into Step

The main idea. A supermarket keeps four display cases cold with one shared set of compressors. Each case has a thermostat. It is a switch that lets cold liquid into the case when its air is too warm and shuts it off when the air is cold enough.

The challenge. The cases send their gas back through one shared pipe, so each case that switches slows or hurries the others. Within about an hour the four switch together, and the compressors must start and stop for all of them at once.

Why it matters. Every start wears a compressor, and a simple rule that keeps the cases apart cut three hours of starts from 45 to 8.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 12.2, the fact in Section 12.3, the fact in Section 12.5, the two try-its in Section 12.6, and the key points
Study	50 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 12.6), Technology in Practice (Section 12.7), Limitations (Section 12.8) and the open challenges

The control question Four display cases boil their refrigerant at one suction pressure, so thermostats that never exchange a signal still act on one another, and an expansion valve can hunt with its tuning unchanged. What pulls the cases into step, what keeps them apart, and what setpoint keeps a superheat loop stable?

Learning objectives

- Explain how a shared suction pressure pulls independent hysteresis thermostats into step, and why a larger suction volume does not prevent it.
- Compute the order parameter R from the thermostats' phases.
- Derive how much steeper the superheat answers the valve below the minimum stable superheat, and explain why a smaller gain is not the cure.
- Compare offsets, a scheduler and a profile MPC by compressor starts and goods temperature.
- Judge what floating pressures, a staggered defrost and head-pressure unloading can and cannot buy on a hot day.

Prerequisites Sections 1.4.7 (thermostats with hysteresis and dwell), 1.4.8 (integer actuators), 1.4.9 (cycle-to-cycle maps), 1.6.2 (the PI law as implemented) and 1.6.4 (scenarios and data). Section 1.6.3 is needed only for the predictive scheduler.

The app https://apmonitor.com/apps/refrigeration-pack/

See also Volume 7, Chapter 3 for a setpoint that chooses the plant, Volume 7, Chapter 14 for loops that settle alone and amplify together, and Volume 17, Chapter 8 for devices that hunt on one shared voltage.

12.1 Introduction

The major retail food outlets of the United Kingdom account for about 4.0 Mt of carbon dioxide a year through their energy use, according to a survey of 2570 stores that covered about 30 % of the stores there with more than 280 m^2 of sales area. The survey found a wide spread of electrical intensity, even within one chain, and estimated that bringing every store above the average down to it would save about 840 GWh a year, or 10 % [1]. The refrigeration and environmental control systems a store uses are among the factors it names for that spread. In the pack of this chapter, four display cases boil their refrigerant at one suction pressure that a rack of compressors holds, and a condenser on the roof rejects their heat together with the compressors' work. A thermostat on each case's air opens and closes that case's refrigerant supply, and an electronic expansion valve on each coil holds the superheat at its outlet. Unless a scheduler is switched on, nothing in the program connects one case's thermostat to another's.

Each of these loops is a textbook loop, and the simulator is designed to correct a belief that loops which behave well alone keep behaving well together. Four identical thermostats started from seeded random phases fall into step after 65.2 min, because the suction pressure they share moves with each of them, and the rack then starts and stops against all four at once, 14 starts in the first hour and 30 in the next two. A superheat loop that holds (5.98 ± 0.16) K at a 6 K setpoint swings between 0.7 and 7.8 K at 2 K with the same gains, because below a minimum stable superheat the evaporator's gain and lag rise under the loop. Neither effect appears in the transfer function of one loop. The first belongs to the utility, since a suction side twenty times larger still locks the cases, sooner, in 52.6 min, and only a pressure that no case can move leaves them apart, and the second belongs to the operating point. Practical and simulation studies of supermarket systems have reported the first, valve actions that coincide under distributed hysteresis control and a compressor that periodically works hard at low efficiency and high wear [2].

The class of problem is a population of on-off or limit-cycling loads coupled through a supply they share. Thermostatically controlled loads on an electric grid are one example, and in heterogeneous populations of them only a spread in the loads' characteristic frequencies damps the aggregate power, whose free response also beats [3]. In a population of coupled limit-cycle oscillators some lock their phases once the coupling passes a threshold, and a review of 25 years of work on the Kuramoto

model traces that transition [4]. Inside the vapor-compression cycle a relative gain analysis has shown couplings between inputs and outputs strong enough that independent single-loop control is inadequate [5]. Compressors on one air header, chillers on one chilled-water loop, boilers on one steam header and heat pumps on one district loop share the shape, a common pressure or temperature that each unit moves and every other unit feels. In a line of cars in Volume 7, Chapter 14 each follower settles well alone and the line can still pass a disturbance back larger, and on the solar feeder of Volume 17, Chapter 8 a tap changer and the inverters act on one voltage while two of its timers chase each other.

However, the supermarket literature poses these questions on benchmark and plant models and answers them with optimization [6–9], and its evidence for synchrony comes from field data and simulations that a student cannot rerun after changing one setting [2]. This chapter maps where a superheat loop loses its stability and how a shared suction pressure pulls independent thermostats into step, on a browser simulation of a four-case pack described completely enough for a reader to build it again. The thermostats, the superheat loop, the sequencer and the floating pressures follow established practice. Measured on the model, the time four cases take to lock moves less with the suction volume and the rack's tuning than with where the cases start (52.6 to 65.2 min from 5000 to 250 L and 54.7 to 78.9 min across the suction loop's gain and integral time, against 31.9 to 94.0 min over five other seeds), and the opening at which a fixed valve meets the stable line moves 1.76 points as the case air cools by 0.28 K and frost builds. The model has 62 states, its charge and case energies close to 4.9×10^{-15} of their throughput, and its controller reads only its instruments. Ten cases and a graded day each isolate one decision.

Section 12.2 follows the history of expansion valves, racks and synchrony. Section 12.3 gives the model of the cases, the suction side, the rack and the condenser, and Section 12.4 works out what the stable line and the shared pressure imply. Section 12.5 sets out the thermostats, the superheat, suction and condenser loops and the supervisor as programmed. Section 12.6 reports the cases, from a valve opened by hand to the summer day, and Section 12.6.10 adds a hybrid predictive scheduler. Section 12.7 describes the real sensors, valves, compressors and controllers of a store, and Sections 12.8 and 12.9 give the limitations and the conclusions, and the exercises in Section 12.10 run the cases.

12.2 Expansion Valves, Compressor Racks and Cases in Step

The stability of the expansion valve was already a modeling subject in 1985, when a lumped model of an evaporator, built from its balance equations for enthalpy, mass and momentum, explained the random fluctuations of the liquid dry-out point, showed that a conventional thermostatic expansion valve often operates unstable, and showed with a process computer connected on line that a controller adapted to the evaporation process raised the evaporator's performance, and therefore its efficiency, by at least 5 % [10]. The thermostatic valve meters liquid on the signal of a remote bulb charged with a refrigerant liquid and vapor mixture, and mathematical models of that sensor were used in 1987 to find the simplest representation of the valve that remained adequate [11].

Electronic valves then took over the metering. A stepper-motor valve on a machine of less than 6 kW, tested at evaporating temperatures from −20 to 10 °C and compressor speeds from 1000 to 3000 rpm, was run under proportional-derivative and qualitative optimal regulation and held its superheat at the set value in steady operation, in a study that names the valve's primary task as metering the evaporator's flow without hunting [12]. In 2002 a control system built on the minimum stable superheat was investigated experimentally on an evaporator [13]. A later controller with a single outlet temperature sensor used the variance of that temperature, which grows at low superheat, to drive the superheat low enough for close to optimal filling of the evaporator [14]. The stable line of Section 12.3 is a teaching version of that boundary, and a real one is found on its evaporator by experiment or from the outlet temperature's variance.

Models for control came alongside. A lumped moving-interface model of the evaporator and condenser predicted where the two-phase region ends, which the superheat depends on, and a relative gain analysis of its linearization found cross-couplings strong enough to indicate that independent single loops are inadequate [5]. A first-principles model of a transcritical air-conditioning system of 11 states was reduced to 5 by singular perturbation once its time scales had been separated [15].

The supermarket pack puts many evaporators on one rack. How its cases were found to fall into step, and how it became a benchmark for hybrid control, is told in the history box below.

> ⧖ HISTORY **2005 to 2008: Cases in Step Become a Benchmark**
>
> A doctoral thesis on the control of refrigeration systems, completed in 2005, set out the traditional supermarket pack: parallel compressors on one suction manifold, a suction pressure controller that switches compressors only outside a dead band, and on each display case a hysteresis controller that opens and closes the case's inlet valve [16]. It reported that in practice the cases tend to synchronize, which loads the central plant periodically and raises its energy use and wear. Hysteresis stayed in use because a thermostatic expansion valve rules out a continuous cascade, and because the off periods let frost melt, so defrosts are needed less often [16]. A study published in 2008 records that the system had become a benchmark problem of the European Network of Excellence on Hybrid Control [17].

Optimization followed. A steady-state model of a simple refrigeration system was used to find the suboptimal condenser pressure that minimizes energy consumption [18]. Nonlinear predictive control parameterized the on/off actions of the valves and compressors by the times at which they switch, so that the hybrid problem became a nonlinear program [7], predictive hybrid control was applied to the benchmark process itself [8], and an economic predictive controller used the thermal capacity of the goods to move the load against the daily outdoor temperature and electricity price and to offer ancillary services to a balancing market [9]. The floating pressures, the scheduler and the staggered defrost of Section 12.5 are rule-based counterparts of these optimizers, and each can be switched off to measure what it earns.

12.3 The Pack as Modeled

Figure 12.1 shows the pack. Four multideck cases, named dairy, meat, deli and produce, take the store's heat in through their air curtains, walls and lights and give it to a coil in the base. A thermostat on each case's air enables that case's electronic expansion valve, which a superheat loop opens and closes. The vapor from the four coils passes an accumulator into one suction side, a rack of two fixed-speed compressors (C1 and C2) and one variable-speed compressor (C3) pumps it to a roof condenser with three fans, and the liquid returns through a high side to the valves. Of the 62 integrated states, 24 belong to the cases (six each), 4 to the pack and 34 are ledgers that integrate every heat and refrigerant flow so that the balances can be audited. The signals are classified in Table 12.1. Time in the equations below is in seconds, and the engine, which counts minutes, scales each rate by 60 in a single place.

12.3.1 A Display Case

Each case is three lumped heat capacities and a refrigerant charge. The air, together with the case's own metal, exchanges heat with the store, the goods and the coil, as shown in Equation (12.1).

$$\begin{aligned} C_a \frac{\mathrm{d}T_a}{\mathrm{d}t} &= Q_{load} - UA_g(T_a - T_g) - Q_{aw} - Q_{sh} \\ C_g \frac{\mathrm{d}T_g}{\mathrm{d}t} &= UA_g(T_a - T_g) \\ C_w \frac{\mathrm{d}T_w}{\mathrm{d}t} &= Q_{aw} - Q_{tp} + Q_{def} - Q_{melt} \end{aligned} \tag{12.1}$$

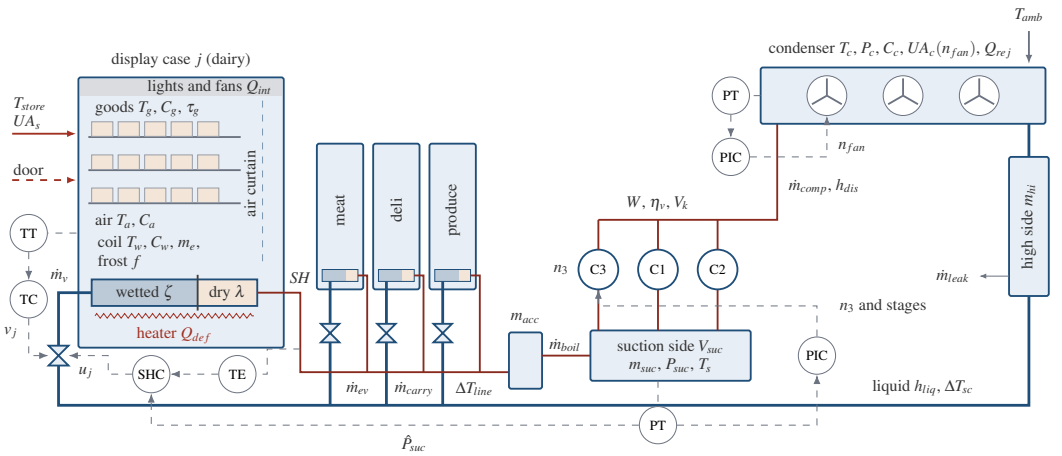


Figure 12.1 The pack, with the dairy case drawn in full as display case j and the meat, deli and produce cases as its copies on the same liquid and suction lines. The store heat enters through UA_s from T_{store}, the lights and fans add Q_{int} and an open door adds a conductance. The air T_a exchanges heat with the goods T_g and with the coil T_w, whose charge m_e wets a share ζ and leaves a dry length λ that sets the outlet superheat SH. A thermostat (TT, TC) gives each valve its call v_j and a superheat loop (TE, SHC, with the suction transmitter) its opening u_j. The coils' evaporation $\dot{m}_{ev}$ and any liquid carried over, $\dot{m}_{carry}$, reach the accumulator m_{acc} and the suction side m_{suc} at P_{suc}, where PIC sets C3's speed n_3 and stages C1 and C2. The compressors deliver $\dot{m}_{comp}$ at h_{dis} to the condenser T_c, whose fans n_{fan} PIC sets against T_{amb}, and the high side m_{hi} returns liquid at h_{liq} less any leak $\dot{m}_{leak}$. The liquid line is drawn heavier than the vapor lines, and the signal lines are dashed.

Table 12.1 Signals of the Pack by Role. Defaults Are the Reference Settings of Every Case

Symbol	Role	Default	Range
$T_{a,j}$	CV of TC-j, state	SP 2 °C, band 2 K	−25–10 °C, 0.5–6 K
v_j	MV of TC-j, valve call	thermostat	0, 1
SH_j	CV of SHC-j, state	SP 6 K	1–15 K
u_j	MV of SHC-j, valve opening	—	5–100 %, shut 0
T_s	CV of PIC-1, saturated suction	SP −10 °C, dead band 2 K	−30–0 °C
n_3, s_k	MVs of PIC-1, C3 speed and stages	—	30–100 %; on, off
T_c	CV of PIC-2, state	SP 40 °C	20–55 °C
n_{fan}	MV of PIC-2, fan speed	—	0–100 %
T_{amb}	DV, measured	22 °C ± 6 K, peak 15:00	±10 K, heat wave
UA_{door}	DV, a door held open	0	0.25 kW/K on one case
—	DV, restocking	—	40 % of goods at 10 °C
N_{out}	DV, condenser fans failed	0	0–3
$\dot{m}_{leak}$	DV, refrigerant leak	0	0–20 kg/h
—	DV, stuck valve, defrost that runs on	off	events
$T_{g,j}$	state, not measured	2 °C	—
$T_{w,j}$, $m_{e,j}$	states, coil and charge	equilibrium	—
f_j	state, frost	0	—
m_{suc}, m_{acc}	states, suction and accumulator	—	—
m_{hi}	state, high-side refrigerant	—	—
$\hat{T}_{a,j}$	measured air	lag 20 s	noise 0.05 K
$\widehat{SH}_j$	measured superheat	lag 6 s	noise 0.08 K
$\hat{P}_{suc}$, $\hat{P}_c$	measured pressures	no lag	noise 0.005 bar
R	monitor, thermostats' order parameter	—	0–1

where the case index j is dropped, $Q_{load} = (UA_s + UA_{door})(T_{store} - T_a) + Q_{int}$ is the store load, $UA_g = C_g/\tau_g = 0.194\,\mathrm{kW/K}$ couples the goods to the air, $Q_{aw} = UA_{aw}(T_a - T_w)/(1 + f/f_{50})$ is the heat the coil takes from the air through a frost layer f, Q_{sh} is the heat that superheats the vapor, Q_{tp} is the boiling heat, Q_{def} is the defrost heater while it runs and $Q_{melt} = k_{melt} T_w \min(1, f/0.05\,\mathrm{kg})$ melts

frost while $f > 0$ and $T_w > 0\,^{\circ}\mathrm{C}$. At 2 °C the store load is 4.91 kW a case, 19.64 kW for the four.

The coil's charge sets how much of it boils, as shown in Equation (12.2).

$$\begin{aligned} \frac{\mathrm{d}m_e}{\mathrm{d}t} &= \dot{m}_v - \dot{m}_{ev} - \dot{m}_{carry}, \qquad \zeta = \min(1, m_e/m_{max}) \\ Q_{tp} &= UA_{wr}\,\zeta'\,(T_w - T_s), \qquad \dot{m}_{ev} = Q_{tp}/\Delta h \\ \dot{m}_v &= u\,\dot{m}_{v,100}\sqrt{\frac{\rho_l(P_c - P_{suc})}{\rho_{ref}\Delta P_{ref}}}\,\min\left(1, \frac{m_{hi}}{m_{hi,min}}\right) \end{aligned} \tag{12.2}$$

where ζ is the wetted share of the coil, $\zeta' = \zeta$ when $T_w \geq T_s$ and $\max(\zeta, 0.3)$ when the coil is colder than saturation, so that a cold coil condenses vapor back and caps the suction pressure near the coldest coil while the rack is stopped, and $\Delta h = h_g(T_s) - h_{liq}$ while the coil boils and $h_g - h_f$ at T_s while it condenses. Liquid fed past a full coil leaves in about 2 s, $\dot{m}_{carry} = (m_e - m_{max})/2\,\mathrm{s}$ for $m_e > m_{max}$. In the valve law ρ_l is the liquid density at the valves, $\rho_{ref} = 1150\,\mathrm{kg/m^3}$ and $\Delta P_{ref} = 8\,\mathrm{bar}$, and the last factor starves the valves when a leak empties the high side. Frost grows at $r_f \min(1, \zeta/0.3)$ while $T_w < 0\,^{\circ}\mathrm{C}$ and $\zeta > 0.02$, and melts at Q_{melt}/L_f with $L_f = 334\,\mathrm{kJ/kg}$.

12.3.2 The Outlet Superheat and the Stable Line

The vapor leaving the wetted length is warmed by the case air over the dry length λ. Above the declared minimum stable superheat SH_{mss} that warming is the physics, and below it the superheat collapses G times faster with the dry length, which Equation (12.3) expresses.

$$SH^* = \begin{cases} \Delta T\left(1 - \mathrm{e}^{-a\lambda}\right), & \lambda \geq \lambda_{mss} \\ SH_{mss}\max\left(0,\ 1 - G\,\dfrac{\lambda_{mss} - \lambda}{\lambda_{mss}}\right), & \lambda < \lambda_{mss} \end{cases} \tag{12.3}$$

where $\Delta T = T_a - T_s$, $a = UA_{sh}/(\dot{m}_{ev}c_{pv})$, and $\lambda_{mss} = -\ln(1 - SH_{mss}/\Delta T)/a$ is the dry length at which the warming reaches the line, while a line at or above ΔT leaves the first branch in force. The superheat chases its target through a lag that also lengthens below the line, and the dry-out point wanders, as shown in Equation (12.4).

$$\begin{aligned} \tau_{sh}\left[1 + (D-1)\beta\right]\frac{\mathrm{d}SH}{\mathrm{d}t} &= SH^* - SH \\ \lambda &= 1 - \zeta + \sigma_\lambda\,(1 + A\beta)\,w\,\xi \\ w &= \min\left[1,\ (1-\zeta)/0.04\right] \end{aligned} \tag{12.4}$$

where $\beta = \min\{1, \max[0, (SH_{mss} - SH)/1\,\mathrm{K} + 0.5]\}$ blends from 0 above the line to 1 a kelvin under it, w removes the wander from a full coil, and ξ is a seeded noise of unit variance updated every 0.5 s as $\xi \leftarrow \alpha\xi + \sqrt{1-\alpha^2}\,\nu$, $\alpha = \mathrm{e}^{-0.5\,\mathrm{s}/\tau_\lambda}$, with ν a standard normal deviate. With $G = 8$, $D = 3$ and $A = 3$ the line is three declared factors on the gain, the lag and the wander. These are teaching values, a reduction of the empirical stability boundary of Section 12.2 rather than a measured evaporator, and a real line is found per evaporator by experiment [13] or on line [14].

12.3.3 Suction Side, Accumulator and Rack

The suction side is one volume whose vapor mass is a state, as shown in Equation (12.5).

$$\begin{aligned} \frac{\mathrm{d}m_{suc}}{\mathrm{d}t} &= \textstyle\sum_j \dot{m}_{ev,j} + \dot{m}_{boil} - \dot{m}_{comp} \\ \frac{\mathrm{d}m_{acc}}{\mathrm{d}t} &= \textstyle\sum_j \dot{m}_{carry,j} - \dot{m}_{boil} \\ \dot{m}_{boil} &= Q_{acc}\min\left(1, m_{acc}/0.05\,\mathrm{kg}\right)/(h_g - h_f) \end{aligned} \tag{12.5}$$

where the accumulator boils its liquid on $Q_{acc} = 1\,\mathrm{kW}$ of line heat, and the pressure is $P_{suc} = P(m_{suc}/V_{suc})$ from the vapor-density fit of Table 12.2, with $T_s = T_{sat}(P_{suc})$. The suction side is 250 L, a 50 L header with 200 L of suction lines. With the header alone the rack stops on its low-pressure switch 23 times in five hours of case 4.

The rack pumps and compresses the vapor as shown in Equation (12.6).

$$\begin{aligned} \dot{m}_{comp} &= \textstyle\sum_k \dot{m}_k = \sum_k s_k \eta_v V_k n_k \rho_1 \\ \eta_v &= 1 - c\left[(P_c/P_{suc})^{1/1.1} - 1\right] \\ W &= \textstyle\sum_k \dot{m}_k \Delta h_s / \eta_k \\ \Delta h_s &= c_s (T_1 + 273.15)\left[(P_c/P_{suc})^{n} - 1\right] \end{aligned} \tag{12.6}$$

where s_k is 1 while stage k runs, $n_k = 1$ for C1 and C2 and $n_3 \in [0.3, 1]$ for C3, $\rho_1 = \rho_{suc}(T_s + 273.15)/(T_1 + 273.15)$ is the inlet density at $T_1 = T_s + SH_1$, SH_1 is the coils' mixed outlet superheat plus $\Delta T_{line} = 4\,\mathrm{K}$ of line heat gain, η_v is clamped to $[0.05, 1]$, $\eta_k = \eta_{is}$ for C1 and C2 and $\eta_{is}[1 - 0.12(1 - n_3)^2]$ for C3, and the discharge leaves at $h_{dis} = h_1 + \Delta h_s/\eta_{is}$. Each displacement V_k is sized so that the compressor gives its rated cooling (10, 10 and 15 kW) at −10 °C suction, 10 K of total superheat, 40 °C condensing and 3 K of subcooling, which makes them 30.13, 30.13 and 45.20 $\mathrm{m^3/h}$ at full speed. The compressors start and stop at once.

12.3.4 Condenser, High Side and the Day

The condensing temperature and the high side's refrigerant follow Equation (12.7).

$$\begin{aligned} C_c \frac{\mathrm{d}T_c}{\mathrm{d}t} &= \dot{m}_{comp}(h_{dis} - h_{liq}) - UA_c(n_{fan})(T_c - T_{amb}) \\ \frac{\mathrm{d}m_{hi}}{\mathrm{d}t} &= \dot{m}_{comp} - \textstyle\sum_j \dot{m}_{v,j} - \dot{m}_{leak} \end{aligned} \tag{12.7}$$

where $UA_c(n) = UA_{c0} + (UA_c - UA_{c0})[n(N_f - N_{out})/N_f]^{0.8}$ for $N_f = 3$ fans with N_{out} failed, $P_c = P_{sat}(T_c)$, the liquid reaches the valves at $h_{liq} = h_f(T_c - \Delta T_{sc})$ with $\Delta T_{sc} = \min[\varphi_{sc}\max(0, T_c - T_{amb}), 5\,\mathrm{K}]$, and the fans draw $(N_f - N_{out})P_{fan}n^3$. The ambient is $T_{amb} = \bar{T} + A_T \sin[2\pi(t_{clock} - 9\,\mathrm{h})/24\,\mathrm{h}] + \delta$ with a heat wave adding 10 K over an hour, holding it and removing it over two. A high-pressure cut-out trips every compressor at $P_c \geq P_{sat}(56\,^\circ\mathrm{C}) = 15.30\,\mathrm{bar}$ and locks the rack out for 10 min and until P_c has fallen 2 bar, and a low-pressure switch stops it at $P_{suc} \leq P_{sat}(-25\,^\circ\mathrm{C}) = 1.06\,\mathrm{bar}$ for a minute and until $P_{suc} > P_{sat}(-22\,^\circ\mathrm{C})$. Both are teaching values. The application notes that real cut-outs are set below the system's maximum allowable pressure under EN 378-2 and the compressor maker's limits. Energy is the integral of W, the fan power and the defrost heaters, and the specific energy divides it by the integrated case load $\sum_j Q_{load,j}$.

> ✦ DID YOU KNOW? **An Industry Leak Rate of One Quarter**
>
> When a federal partnership with food retailers was launched in 2007, its agency developed an industry average leak rate of 25 % for supermarket refrigerant. The partners' average emissions rate was 12.9 % in 2021, about half of that estimate [19].

12.3.5 Refrigerant Properties and Instruments

Every property comes from a declared fit, listed in Table 12.2, and not from an equation of state. Against the seven-point R134a table that the application's checks carry, from −40 to 60 °C, the fits hold the saturation pressure within 0.32 %, the enthalpies within 0.18 kJ/kg and the vapor density within 0.28 %. Real properties come from the international standard formulation for R134a, a Helmholtz equation of state that represents the vapor pressure to about 0.02 % and the density to about 0.05 % [20]. R404A and R290 are offered as approximate alternatives, and no case uses them.

Table 12.2 Declared R134a Fits, With T in °C, P in bar, h in kJ/kg on the IIR Reference and ρ in kg/m^3. Each Column Is a Cubic $c_0 + c_1 x + c_2 x^2 + c_3 x^3$

x	h_f T	h_g T	ρ_f T	$\ln P$ $\ln \rho_g$
c_0	199.9849	398.617	1294.945	−1.700271
c_1	1.338574	0.5925403	−3.230513	0.9737556
c_2	1.453863×10^{-3}	-1.158092×10^{-3}	-8.782384×10^{-3}	0.04937264
c_3	1.089485×10^{-5}	-1.643745×10^{-5}	-8.253043×10^{-5}	−0.009170624

Saturation $\ln P_{sat} = A - B/(T + C)$ with $A = 9.939613$, $B = 2163.556$ and $C = 244.0$. Vapor heat capacity $c_{pv} = 0.897 + 0.00436\,T_s$ kJ/(kg K). Isentrope $c_s = 0.604$ kJ/(kg K) and $n = 0.1185$. The vapor density $\rho_g(P)$ inverts the $\ln P$ cubic by Newton's method.

Table 12.3 Constants of the Model. Sliders of the Show Model Dialog Are V, Hidden Parameters H and Code Constants C

Symbol	Meaning	Value	Class
Display cases (each)			
C_a, C_g, τ_g	air and metal, goods, goods lag	320, 1400 kJ/K, 120 min	V
UA_s, Q_{int}, T_{store}	store load, lights and fans, store	0.24 kW/K, 0.35 kW, 21 °C	V
UA_{aw}, UA_{wr}, UA_{sh}	air to coil, coil to boiling, air to vapor	1.4, 1.2, 0.21 kW/K	V
C_w, m_{max}	coil metal, charge of a full coil	25 kJ/K, 1.5 kg	V
$\dot{m}_{v,100}$	valve capacity at 8 bar	0.10 kg/s	V
τ_{sh}, SH_{mss}	superheat lag, stable line	18 s, 3 K	V
G, D, A	steepening, lag and wander factors	8, 3, 3	V
σ_λ, τ_λ	wander, its correlation time	0.012, 15 s	V, H
r_f, f_{50}	frost rate, frost halving UA_{aw}	0.5 kg/h, 5 kg	V
Q_{def}, k_{melt}	defrost heater, melting	2 kW, 0.8 kW/K	V, H
Suction side and rack			
V_{suc}	suction side volume	250 L	V
Q_{c1}, Q_{c2}, Q_{c3}	rated cooling of C1, C2, C3	10, 10, 15 kW	V
n_{min}, η_{is}, c	C3 least speed, efficiency, clearance	0.30, 0.65, 0.05	V, V, H
ΔT_{line}, Q_{acc}	line heat gain, accumulator heat	4 K, 1 kW	V, H
T_{LP}	low-pressure cut-out, saturated	−25 °C	H
Condenser and high side			
UA_c, UA_{c0}, C_c	full fans, fans stopped, capacity	2.7, 0.35 kW/K, 60 kJ/K	V
N_f, P_{fan}	fans, power of one at full speed	3, 0.55 kW	H, V
φ_{sc}, $\Delta T_{sc,max}$	subcooling share, largest	0.35, 5 K	V, H
m_{ch}, $m_{hi,min}$	liquid charge, flashing threshold	30, 6 kg	V, H
T_{HP}, t_{lock}	high-pressure cut-out, lockout	56 °C, 10 min	V
Weather			
$\bar{T}$, A_T	daily mean ambient, half range	22 °C, 6 K	V
Instruments			
—	air, outlet sensor lags and noise	20, 6 s; 0.05, 0.08 K	H
—	pressure transmitter noise	0.005 bar	H
Code constants			
UA_{door}, L_f	open door, latent heat of frost	0.25 kW/K, 334 kJ/kg	C
—	carry-over time, condensing floor	2 s, 0.3	C

The controller reads only its instruments. The case air sensors lag 20 s with 0.05 K of noise, the coil outlet sensors lag 6 s with 0.08 K, the two pressure transmitters carry 0.005 bar of noise and the ambient sensor 0.1 K, while the defrost termination reads the coil temperature directly. The goods are never measured.

12.3.6 Starting State and Numerics

Cases 1 and 2 start every coil at equilibrium, the air at the temperature where the store load equals what the coil takes at the setpoint superheat, which in case 2 is 2.14 °C with the valves at 32.31 % and the coils 82.1 % wet at −6 °C suction. The other cases place each thermostat at a phase of its cycle, seeded or pinned, on a triangle that cools from cut-in to cut-out over the first 60 % of the cycle. The condenser starts at the larger of its setpoint and the ambient plus 8 K, the rack with the stages and C3 speed that carry the cases cooling at $t = 0$, and the fans with the speed that holds that condensing temperature. The suction side starts with the saturated vapor density of its setpoint and the high side with the 30 kg charge less what the coils hold. The fourth-order Runge-Kutta method integrates the 62 states on a 0.5 s step, the controller scans every 2 s, the historian logs every 5 s, and

the wander has its own 0.5 s clock so that a finer step replays the same realization (Section 1.6.1). The seed is 20261010. Over the challenge day of Section 12.6 with the suggested settings, the charge and every case's energy close to 4.9×10^{-15} of their throughput. Halving the step changes the day's 183.51 kWh by less than one part in 10^7 and leaves its 108 compressor starts unchanged, because the controller scans on its own 2 s clock whatever the step, so every thermostat and stage switches on the same scan. The application passes its 36 checks.

12.3.7 Assumptions

Lumped cases. A case is three heat capacities and a charge, with no air curtain dynamics, no temperature profile along a shelf and no product core, so the goods follow the air through one two-hour lag. Frost is laid at a declared rate with no latent heat and no humidity balance, so only its melting appears in the coil's energy.

A declared stable line. The superheat is one exponential over one dry length, and its behavior below the line is three declared factors rather than a two-phase flow simulation. A real line is measured on its evaporator.

One suction volume. The suction side is one pressure, its vapor's density is the saturated value at that pressure, and there is no pressure drop between the cases and the rack.

Ideal machines. The compressors are displacements with a clearance efficiency and a fixed isentropic efficiency, with no oil return, motor heating or starting transient, and the condenser and receiver are one saturation temperature with a lumped capacity.

Fitted properties and teaching thresholds. The refrigerant comes from fits good to a fraction of a percent, and every threshold, from the 3 K line and the 56 °C cut-out to the goods band of 1 to 3 °C and the challenge's limits, is a teaching value set on this model.

12.4 What the Line and the Shared Pressure Imply

12.4.1 A Steep Stretch That Moves

Figure 12.2(a) draws the dairy coil's steady superheat in case 1 against the valve opening, holding the case air, the suction and the frost at their values in the first and fifth steps of the case's manual sequence. Above the line the superheat falls 0.65 K for each percent of opening at the first step's air of 2.36 °C, and past the line it loses the remaining 3 K within 0.17 points of opening, 18 K per percent. That steep stretch is the stable line seen from the valve. Its position is not a property of the valve, because the opening at which a fixed valve meets the line moves from 35.63 % to 33.87 % between the two steps as the air cools to 2.08 °C and 0.20 kg of frost builds. A setpoint chosen as an opening would cross the line by itself.

An open-loop step of the valves from 31.5 to 32.5 % on case 2's conditions, in a model run with every noise off and the case's scripted setpoint steps removed, gives a process gain of −1.03 K per percent on the true outlet superheat, and a two-point fit at 28.3 and 63.2 % of the change gives a lag of about 143 s and 7 s of apparent dead time. The model holds no transport delay, so that dead time is the coil's charge and the 18 s superheat lag acting in series. The superheat loop's gain of 1.2 %/K therefore makes a loop gain of 1.24 at 6 K, and with the noise still off the loop steps from 6 to 8 K to reach 7.8 K after 186 s. Below the line the same opening moves the superheat more than ten times as far, three times as slowly and with three times the wander, so the identical proportional-integral law sees a different plant. The setpoint chooses the plant in the same way in Volume 7, Chapter 3. At a 2 K setpoint it settles into the limit cycle of Figure 12.2(b), whose troughs under 1.5 K come 2.1 to 5.1 min apart.

12.4.2 One Pressure for Four Cases

Every case boils at the saturation temperature of one pressure, so its boiling heat $UA_{wr}\zeta(T_w - T_s)$ falls whenever another case raises T_s. Over the first three hours of case 4, in the 30 seconds after a cut-in with no other valve or stage switching from 10 seconds before to 30 seconds after, the suction

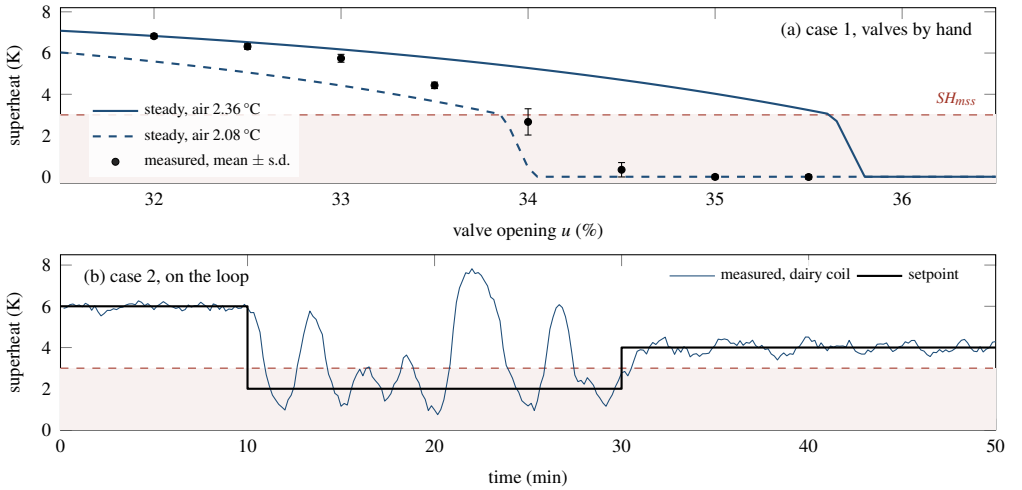


Figure 12.2 The superheat and its stable line, from the engine. (a) The dairy coil's steady superheat in case 1 against the valve opening, at the air and frost of the case's first and fifth steps, with the measured superheat of each six-minute step as a mean and standard deviation over its last four minutes. (b) The dairy coil's measured superheat in case 2, on its loop, as the setpoint steps from 6 to 2 K at 10 min and to 4 K at 30 min. The shaded band lies below the declared 3 K line.

rose a median 1.02 K (11 events) and the cases already cooling lost a median 11.0 % of their boiling heat. A cut-out lowered the suction 0.72 K. The rack's staging answers in the opposite sense and by more, since a compressor start lowered the suction a median 2.38 K within 20 seconds (32 events) and lifted the cooling cases' boiling heat 14.3 %, and a stop raised it 1.86 K and cut that heat by 14.4 %. A case ahead in its cycle is slowed by the cases that open after it, and the timing of every stage change is set by the cases together. Over many cycles the phases draw together.

🕮 **Pause.** Try Exercise 12.1 before reading on.

The pull is strong, and what carries it matters less than where the cases start. From the same seeded phases the four cases are in step after 65.2 min on 250 L, 53.1 on 1000 L and 52.6 on 5000 L, each volume set before the run. Halving or doubling the suction loop's gain gives 56.2 and 62.6 min and its largest setting of 60 %/K gives 78.9 min, and integral times of 15, 60 and 120 s give 67.2, 77.0 and 54.7 min. A 4 K dead band locks them sooner, at 47.9 min, with the measured suction spanning −14.0 to −3.3 °C between its 1st and 99th percentiles from the first half hour to the third hour against −12.1 to −4.2 °C with the shipped 2 K. On five other seeds the cases are in step after 31.9 to 94.0 min, 2.8 to 8.1 of their cycles, a wider spread than any of these settings gives, and on an infinite suction side of 1×10^{12} L, a pressure no case can move that only the application's checks reach because the parameter stops at 5000 L, none of the five ever synchronizes. Two cases alone, the dairy and produce cases with the other two thermostats in Manual and their valves shut, close a gap of 0.39 of a cycle to 0.10 by their fifth cut-in on the shared side and stay within 0.10 after the first hour, while on the infinite side their gap stays between 0.40 and 0.50 for four hours.

12.4.3 What the Condenser Can Carry

The four cases take 19.64 kW from a 21 °C store. Rejecting that load and the work needed to pump it into air at 38 °C needs a condensing temperature of 48.7 °C with all three fans, 52.5 °C with one out and 62.6 °C with two out, against the 56 °C cut-out, so with two fans out no controller holds both the goods and the rack. The cost of head pressure is steep as well. C1 alone at −10 °C suction, with 6 K of superheat at the coil outlet and 3 K of subcooling, has a cycle coefficient of performance of 4.31 at 25 °C condensing, 2.70 at 40 °C and 1.80 at 55 °C, which is what floating the head pressure on the

Table 12.4 Loops and Supervisory Settings With Their Defaults

Loop	CV → MV	Setting
TC-j	$\hat{T}_{a,j} \to v_j$	on above 3 °C, off below 1 °C (SP ± band/2)
SHC-j	$\widetilde{SH}_j \to u_j$	PI, $K_c = 1.2\,\%/\mathrm{K}$, $\tau_I = 60\,\mathrm{s}$, 5–100 %, $a = -1$
	protection	below 1 K, closes at 30 %/min
PIC-1	$\hat{T}_s \to n_3$	PI, $K_c = 25\,\%/\mathrm{K}$, $\tau_I = 30\,\mathrm{s}$, 30–100 %, $a = -1$
	staging	dead band 2 K, requests held 15 s, 0.5 min between changes
	timers	C1, C2 run 2 min, rest 5 min; C3 run and rest 1 min
PIC-2	$\hat{T}_c \to n_{fan}$	PI, $K_c = 3\,\%/\mathrm{K}$, $\tau_I = 60\,\mathrm{s}$, 0–100 %, $a = -1$
FY-1	floating suction	off; +0.05, −0.25 K/min, duty limit 85 %, −14 to −4 °C
	floating head	off; $\max(25\,^\circ\mathrm{C}, \hat{T}_{amb} + 8\,\mathrm{K})$
KY-1	desynchronization	none; offsets 1.0, 0.8, 1.25, 0.65; scheduler gap 2.5 min
	defrost	staggered, every 6 h from 02:00, 20 min or 10 °C, drip 3 min

ambient buys back. In Volume 9, Chapter 8 the cheapest feasible operating point of a cooling plant likewise moves with the weather.

12.5 Thermostats, Valves, Rack and Fans as Programmed

The controller scans every 2 s and reads only the instruments of Section 12.3. Its loops and their settings are listed in Table 12.4, and every loop has its own Manual and Auto switch, while the header's switch throws all of them at once.

12.5.1 The Shared Controller

The superheat, suction and condenser loops run the shared velocity-form law of Section 1.6.2 with no derivative term and no output bias, so Equations (1.14) to (1.17) reduce to Equation (12.8).

$$
\begin{aligned}
&e_k = a(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c e_k \\
&S_k = S_{k-1} + P_k - P_{k-1} \\
&v_k = S_k + I_{k-1}, \qquad u_k = \min[u_{max}, \max(u_{min}, v_k)] \\
&I_k = I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_k - v_k}{T_t}\right]\Delta t_c
\end{aligned}
\tag{12.8}
$$

where $a = \pm 1$ is the action, $\Delta t_c = 2\,\mathrm{s}$, the integral I carries the output's starting value and the tracking time constant T_t equals τ_I. In Manual the integral is solved each scan as $I_k = u_{op} - S_k$, so the return to Auto is bumpless. No loop carries a rate limit. All three loops have $a = -1$, so a measurement above its setpoint raises the output, which the application's direction checks confirm. A superheat above its setpoint opens the valve, a suction above its setpoint speeds C3 up and a condensing pressure above its setpoint speeds the fans, which is direct action in the ISA sense although comments in the app's source call the loops reverse acting.

12.5.2 Cases and Valves

Each thermostat is the switch with hysteresis of Section 1.4.7. It opens its valve call when its measured air passes the cut-in, the setpoint plus half the band, and closes it below the cut-out, half a band under the setpoint, and a defrost holds it closed. While a case calls, its superheat loop sets the opening, and with the superheat loops in Manual every calling valve takes the operator's opening, which the arrow keys move in quarter percents. Between calls the valve is shut and the loop tracks the opening it closed at, so each restart begins where the last call ended. Below a measured superheat of 1 K a low-superheat protection overrides the loop and closes the valve at 30 %/min, with the loop tracking it. A stuck valve stays where it stuck whatever it is told.

12.5.3 Suction and Staging

C3 trims the suction on its speed, and C1 and C2, two integer actuators in the sense of Section 1.4.8, are staged by the framework's stage sequencer, shared/src/framework/core/sequencer.js. A request for

one more stage is made when the measured suction has been above the top of the 2 K dead band for 15 s with C3 off or at full speed, and a request for one fewer when it has been below the bottom for 15 s with C3 off, fixed or at its least speed. The sequencer starts the free stage of lowest priority, C3 first and then whichever of C1 and C2 has been off longest, stops the running stage of highest priority that has run its minimum, holds 0.5 min between changes unless the suction is a full dead band low, and counts every start inside a trailing hour. The minimum run and rest are 2 and 5 min for C1 and C2 and 1 min each for C3, and C3's loop restarts at its least speed each time C3 starts or stops. In Manual the operator sets C3's speed, which stops it at zero, and the number of fixed compressors. With head-pressure unloading on, the rack adds no stage and caps C3's speed in proportion once the condensing pressure is within 1.5 bar of the cut-out, and sheds a stage within 0.5 bar of it. The cap is applied after the output of Equation (12.8), outside the loop's own limits, so the loop's integral is not told of it and can wind up toward 100 % speed while the cap holds.

> ✦ DID YOU KNOW? **Timers That Move the Cycling**
>
> From the simulator: minimum run and rest times on C1 and C2 cut the rack's starts in three hours of case 3 from 97 to 59, but the variable-speed C3 then makes 20 of them, the suction swings from −15.0 to −1.7 °C instead of −11.9 to −7.7 °C, and R averages 0.37 (Table 12.6).

12.5.4 Condenser and Supervisor

The fans hold the condensing temperature at 40 °C, or with floating head pressure at the measured ambient plus 8 K, never under 25 °C, and in Manual the operator sets their speed. Floating suction averages each case's valve-open share over 30 min as a first-order lag, raises the suction setpoint 0.05 K/min while every share is under 85 %, and lowers it 0.25 K/min as soon as one passes 90 % or a case's air is a kelvin past its cut-in, between −14 and −4 °C, so the case with the least room sets the pressure for all four. Desynchronization by offsets scales the four bands by 1.0, 0.8, 1.25 and 0.65, so the thermostats cycle at four periods. The scheduler holds a case that reaches its cut-in until the last opening is 2.5 min old or the case is 0.6 K past its cut-in, and releases the warmest idle case early once its air is 0.2 K above its setpoint, nobody is waiting and the last opening is 3.75 min old. Defrosts start on a clock, staggered in four slots of an interval or all at once, heat the coil at 2 kW until it reaches 10 °C or 20 min have passed, and drip for 3 min, while a defrost whose termination fails runs to a 60 min failsafe.

12.5.5 Monitors and the Models Kept Apart

Two monitors read the controller's own signals. The hunting detector records the measured superheat's turning points with 1 K of hysteresis and flags a coil that holds, in the last 12 min, at least five swings of 2 K or more and three troughs at least 0.5 K under the line, ignoring the first 3 min after each opening. The synchrony monitor gives each case the phase $\varphi_j = 2\pi(t - t_{in,j})/P_j$ since its last cut-in over its last period and computes the order parameter of Equation (12.9).

$$R = \frac{1}{N}\left|\sum_{j=1}^{N} \mathrm{e}^{i\varphi_j}\right| \tag{12.9}$$

where R is 1 when the cases cut in together and near 0 when they are spread, and the status reads SYNCHRONIZED once R has stayed above 0.9 for 20 min. Equations (12.1) to (12.7) are the simulation model. The identified model of the superheat at 6 K is the first-order lag with dead time of Section 12.4, and the controller holds no model of the plant, since the floating suction reads duty as a proxy for room and the scheduler needs only the openings.

12.6 Measured Cases, From a Hand on the Valve to a Hot Day

Table 12.5 lists the cases and Table 12.6 their measured results. Each starts from the reference settings of Tables 12.1 and 12.4 and changes only what its row lists, and among the original cases only case 6

Table 12.5 The Cases, Each Changing the Reference Settings as Listed

#	Change from the reference, duration	Teaching point
1	cases held calling, valves by hand from 32 % up 0.5 % every 6 min, suction −6 °C, 48 min	the line by hand
2	as 1 with the superheat loops on, SP 6, 2 and 4 K, 50 min	hunting below the line
3	staging alone, C3 fixed, no timers, scheduler, 3 h	starts against timers
4	free-running thermostats from seeded phases, 3 h	synchrony
5	started in step, offset differentials, 3 h	desynchronization
6	2 kg of frost, all four defrosted at 01:00, scheduler, 4 h	a scheduled peak
7	floating suction from −12 °C, meat case at 1 °C, scheduler, 6 h	the case that binds
8	13:00 start, two fans out, heat wave at 14:00, floating head, unloading, scheduler, 6 h	an infeasible afternoon
9	meat case restocked at 15 min, scheduler, 5 h	goods the air hides
10	predictive case admission, 3 h	restricted hybrid scheduling
11	a summer day, see Table 12.7	graded challenge

Table 12.6 Measured Results of the Cases and Their Variants

Case and variant	Result
1, 32 to 33.5 %	(6.82 ± 0.11) K to (4.43 ± 0.16) K, coils 77 to 89 % wet
1, 34, 34.5, 35 %	(2.66 ± 0.64) K, (0.34 ± 0.35) K, coil full and 0.18 kg of liquid by 48 min
2, SP 6, 2, 4 K	(5.98 ± 0.16) K; 0.67 to 7.84 K, hunting from 18.9 min; (3.98 ± 0.25) K
3, no timers	97 starts, suction −11.9 to −7.7 °C
3, C1, C2 timed 2 and 5 min	59 starts (C3 20), −15.0 to −1.7 °C, *R* 0.37 mean
3, timed, C3 variable	no start, −10.7 to −9.4 °C
4, free running	in step at 65.2 min, 14 starts in the first hour, 30 in the next two
4, infinite suction side	never in step, *R* averaging 0.45 from hour 1 to 4, no restaging
5, offsets, scheduler, none	24, 8 and 45 starts; *R* 0.33, 0.07 and 1.00 at 3 h
6, all at once, staggered	56.2 kW and 46.4 kW rejected at peak, all four cooling 17.5 and 2.8 min after 01:00
7, floating, fixed	43.21 and 45.42 kWh, 27 and 58 starts, setpoint −9.39 °C mean
8, unloading on, off	no trip, goods out 205 to 208 min; 26 trips, goods to 14.87 °C
9, restocked	goods 5.22 °C, out of band 147 min, air at most 3.58 °C

and the summer day defrost. No interface takes part in the results below, which run the engine from its default seed over each case's own duration unless another span is named. A setting that a variant changes is made at the case's start, before the run, as a learner makes it, and the application gives the same run whether it is made before the first Run or carried through a Reset (Section 1.6.1). A start-up parameter such as the suction side volume, changed during a run, waits for the next Reset. Where the noise matters, the spread over five other seeds is given beside the default seed's value.

12.6.1 The Line by Hand

Case 1 holds all four cases calling with their thermostats in Manual and opens the four valves by half a percent every six minutes from 32 %, with the suction held at −6 °C. From 32 to 33.5 % the superheat falls from 6.82 to 4.43 K and stays within ±0.11 to ±0.19 K, while the coils go from 77 to 89 % wet. The next half percent takes it under the line, to a mean of 2.66 K that wanders between 1.52 and 4.00 K, at 34.5 % it sits at 0.34 K with the coils 98 % wet, and at 35 % they are full and liquid starts to reach the accumulator, 0.18 kg of it by 48 min. The case has no loop, so it floods rather than hunts. The measured points of Figure 12.2(a) follow the steady curve for the air of each step until the line. At 34 % the mean stays above the curve's 0.44 K because the coil is still filling when the step ends, and with the wander and the sensor noise switched off its true superheat is still falling, from 4.32 K at 24 min to 0.83 K at 30 min.

12.6.2 The Loop That Hunts

Case 2 puts the same coils on their superheat loops at 6 K, where the superheat holds (5.98 ± 0.16) K with the valves between 31.7 and 32.6 %. At 10 min the setpoint drops to 2 K, a kelvin under the line. The superheat never settles. It swings between 0.67 and 7.84 K, averaging (3.27 ± 1.98) K, the valves travel from 17.5 to 38.0 %, the low-superheat protection acts for 6 % of the time from 12 to 30 min, and the loop is flagged hunting from 18.9 min. Raised to 4 K at 30 min, a kelvin above the line, the

superheat holds (3.98 ± 0.25) K, as shown in Figure 12.2(b). The cure is margin above the line, not a smaller gain.

Pause. Try Exercise 12.2 before reading on.

12.6.3 Starts Against Timers

Case 3 holds the suction at −10 °C by staging alone, with C3 at fixed speed, a 2 K dead band, no timers and the scheduler keeping the cases apart. The rack makes 97 starts in three hours, 48 of C1 and 49 of C2, and after the first half hour the suction stays between −11.9 and −7.7 °C at its 1st and 99th percentiles. Minimum run and rest times of 2 and 5 min on C1 and C2 cut the starts to 59, but C3 takes up the cycling with 20 of them, the suction swings from −15.0 to −1.7 °C and lies outside its dead band 49 % of the time, and the swings begin to pull the thermostats toward step, R averaging 0.37 after the first half hour against 0.09 untimed. Five other seeds give 97 to 104 starts untimed and 54 to 67 timed. Giving C3 its speed back removes every start on the default seed, leaves 2 to 7 on the other five, and holds the suction between −10.7 and −9.4 °C.

12.6.4 Synchrony and Its Cures

Case 4 runs the four thermostats free from the seeded phases 0.59, 0.67, 0.38 and 0.08 of a cycle. R climbs from 0.59 at 20 min to 1.00 at an hour, and the cases are in step from 65.2 min, 5.7 cycles of 11.4 min, as the screenshot of Figure 12.3 shows at two hours. In step, the rack makes 14 starts in the first hour and 30 in the next two, and the suction swings from −12.4 to −3.0 °C in the last hour against −12.2 to −7.5 °C in the first half hour. On the infinite suction side R averages 0.45 from the first hour to the fourth and the rack never restages.

> **TRY IT IN THE APP** **4 · Four Thermostats Fall Into Step**
>
> Load the scenario and press Run. Watch the four hands of the thermostat dial on the Store view draw together. R climbs from 0.59 at 20 min to 1.00 at an hour, the cases are in step from 65.2 min, and the status line then reads SYNCHRONIZED. On the trends the four case air temperatures merge into one trace.

Case 5 starts the cases a hundredth of a cycle apart. With offset differentials R falls from 1 to 0.30 in 30 min and averages 0.40 from the first hour to the third, and the rack makes 24 starts against 45 free running, but the four periods let the phases slide back through one another and R is back to 0.33 at three hours. The scheduler spreads them and keeps them spread, R 0.42 at 30 min, 0.16 at an hour, 0.12 at two and 0.07 at three, with 8 starts in three hours and 3 and 2 early releases of the meat and produce cases. Five other seeds give 22 to 28, 6 to 9 and 45 starts with offsets, the scheduler and neither. The scheduler costs the warmest air 0.47 K, a peak of 4.09 °C against 3.61 °C free running.

> **TRY IT IN THE APP** **5 · Desynchronization: Offsets and a Scheduler**
>
> Load the scenario and press Run. With Desynchronization on Offsets, R falls from 1 to 0.30 in 30 min, the rack makes 24 starts in three hours, and R is back to 0.33 at the end. Press Reset, set Desynchronization to Scheduler on the KY-1 card before pressing Run, and run again: 8 starts, and R 0.07 at three hours.

12.6.5 A Defrost Is a Scheduled Load

Case 6 gives each display case 2 kg of frost, keeps the cases apart with the scheduler and defrosts all four at 01:00. Each heater melts its frost, 670 kJ of latent heat, and warms the coil to 10 °C in 7.2 to 9.0 min, ending by temperature, and a 3 min drip follows. From 01:00 to the end of the four hours all four cool at once for 17.5 min, against 2.8 min with the defrosts staggered 90 minutes apart, and five other seeds give 16.0 to 20.0 against 1.1 to 4.0 min. Both runs bring in all three compressors, but all at once the condenser rejects 56.2 kW at its peak against 46.4 kW, the electric power peaks at 15.4 kW against 13.5 kW, and the condensing temperature reaches 49.8 °C against 44.3 °C. The

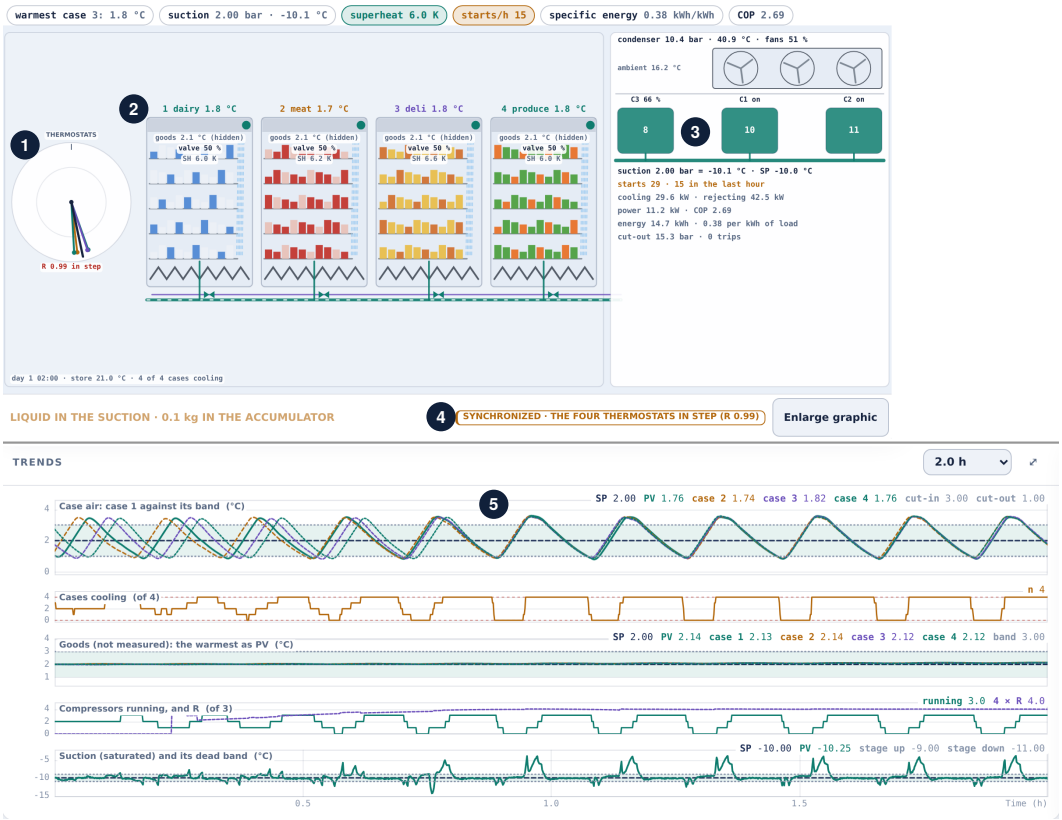


Figure 12.3 Case 4, Four thermostats fall into step, at $t = 2.00$ h in the Store view with the 2 h trend window and no control changed, a composite of the app's plant view with its status line and its trend strips; the header, the control panel and the Learn panel are cut. (4) The status line reads SYNCHRONIZED with the four thermostats in step at R 0.99, and (1) the thermostat dial draws the four hands together. (2) All four cases cool, their air at 1.8, 1.7, 1.8 and 1.8 °C, every valve at 50 % and measured superheats of 6.0 to 6.6 K, with the hidden goods at 2.1 °C. (3) The rack runs C3 at 66 % with C1 and C2, after 8, 10 and 11 starts, 29 in all and 15 in the last hour, holding the suction at 2.00 bar, −10.1 °C, and the condenser at 10.4 bar, 40.9 °C, with the fans at 51 % against a 16.2 °C ambient. The coils take 29.6 kW, the condenser rejects 42.5 kW and the pack draws 11.2 kW of electricity, the COP over the compressors' power is 2.69, and the run so far has used 0.38 kWh per kWh of case load. (5) In the trends the four case air temperatures cycle apart in the first half hour and as one trace by the second hour, the cases cooling step between none and four, the compressors running follow them as four times R climbs to 4.0, and the suction swings each time the rack restages.

defrost brings the cases back in step, and the scheduler spreads them again, R averaging 0.22 from 01:40 to 02:40 against 0.15 staggered.

12.6.6 The Case That Binds a Floating Suction

Case 7 sets the meat case at 1 °C and the other three at 2 °C on a cautious fixed −12 °C, with floating suction on. The meat case binds from the first hour to the last, reaching a duty of 78 % at six hours while the others sit at 66 to 72 %, and the setpoint floats between −10.81 and −8.12 °C, averaging −9.39 °C from the first hour to the sixth. The six hours cost 43.21 kWh against 45.42 kWh fixed, 4.9 % less, the compressors' share 4.9 % less for the 2.6 K gained, and the compressors start 27 times

instead of 58. Five other seeds save 5.4 to 6.3 % with 19 to 25 starts against 66 to 72. The coldest case taxes every case on the line.

12.6.7 An Afternoon No Setting Saves

Case 8 starts at 13:00 with two of the three fans out of service, and at 14:00 a heat wave lifts the ambient 10 K over an hour, to 38 °C at 15:00, holding until 18:00. Section 12.4 showed that the load needs 62.6 °C of condensing temperature with two fans out against the 56 °C cut-out. With head-pressure unloading the rack never trips. It holds the condensing temperature to 54.7 °C by refusing capacity for 292 min, and from 15:32 every case's goods leave their band, for 205 to 208 min, reaching 5.13 °C. Without unloading the rack trips 26 times, the first at 14:24 when 4 K of the heat wave has arrived, and the goods reach 14.87 °C. Five other seeds trip 31 to 33 times, the first between 13:01 and 13:16 before the heat wave begins, so the time of the first trip depends on the noise while the outcome does not. With one fan repaired at the start unloading holds every case in its band, and with both repaired it is never needed. Unloading chooses warm goods over a stopped rack, and only a repaired fan avoids both.

12.6.8 Goods the Thermostat Cannot See

Case 9 replaces two fifths of the meat case's goods with goods at 10 °C at 15 min. The goods reach 5.22 °C and take 147 min to return under 3 °C, 140 to 145 min on five other seeds, while the air never leaves its ordinary cycling, reaching at most 3.58 °C after the restock, and the thermostat's only sign is a half-hour duty that rises from 59 % just before the restock to 74 % 50 minutes after it. A thermostat on the air holds the air, and no tuning of it holds a goods band.

🕮 **Pause.** Try Exercise 12.3 before reading on.

12.6.9 The Summer Day

The challenge runs a day on fixed suction (−10 °C) and condensing (40 °C) setpoints with free-running thermostats, a 2 K superheat setpoint and all four cases defrosted at once at 04:00, 10:00, 16:00 and 22:00. A heat wave arrives at 14:00 and the meat case is restocked at 16:00. The target asks for no high-pressure trip, at most 150 compressor starts, every case's goods in their band for all but 240 min, at most 20 min of hunting and 90 min with liquid in the accumulator, and at most 0.41 kWh per kWh of case load. The day is judged as given, so a changed model parameter or disturbance setting, or an extra disturbance button, voids the attempt. Table 12.7 compares the configurations, and Figure 12.4 draws three of its columns against the limits. As shipped the rack trips 9 times, the first at 16:11, just after the 16:00 defrost and two hours into the heat wave, makes 383 starts with the cases synchronized for 848 min, leaves the worst goods out of their band for 473 min at up to 9.82 °C, and floods the accumulator for 877 min. The suggested configuration, floating suction and head pressure, the scheduler, a staggered defrost and a 6 K superheat setpoint, passes with no trip, 108 starts and 188 min out of band, all of them in the restocked case. On five other seeds the shipped day also trips 9 times, with 385 to 391 starts, and the suggested one passes with 100 to 113 starts. Undoing the scheduler, the staggered defrost or the 6 K setpoint fails the target, and so does fixing both pressures, while either float alone is enough. With the defrost all at once the rack trips 11 times, with the thermostats free running it makes 178 starts, at 2 K the loops hunt for 340 min, and with both pressures fixed the day costs 0.417 kWh per kWh, against 0.405 with only the suction fixed and 0.407 with only the head pressure fixed, and offset differentials in place of the scheduler also pass, with 139 starts. Removing the trip alone does not rescue the shipped day, which still makes 354 starts staggered at 6 K, while head-pressure unloading can stand in for the staggering.

12.6.10 An Eight-Minute Hybrid Scheduler

The Desynchronization control also offers an MPC choice, which addresses the shared load by choosing case calls while the rack still enforces its compressor timers. Every minute a mixed-integer linear program selects one eight-minute profile per case from 00000000, 11111111, 11001100 and

Table 12.7 The Summer Day in Twelve Configurations. Limits: No Trip, 150 Starts, 240 Goods Minutes, 20 Hunting, 90 Liquid, 0.41 kWh per kWh

Configuration	Trips	Starts	Goods	Hunt	Liquid	kWh/kWh
as shipped	9	383	473	17	877	0.443
suggested	0	108	188	0	19	0.392
suggested, unloading	0	115	191	3	26	0.392
suggested, defrost at once	11	151	471	0	218	0.389
and unloading	0	117	164	0	29	0.392
suggested, 2 K superheat	0	128	177	340	192	0.394
suggested, free running	0	178	188	0	40	0.393
suggested, offsets	0	139	200	2	29	0.392
suggested, fixed suction	0	95	197	7	2	0.405
suggested, fixed head	0	91	189	6	7	0.407
suggested, fixed pressures	0	66	205	1	5	0.417
shipped, staggered, 6 K	0	354	219	8	179	0.439

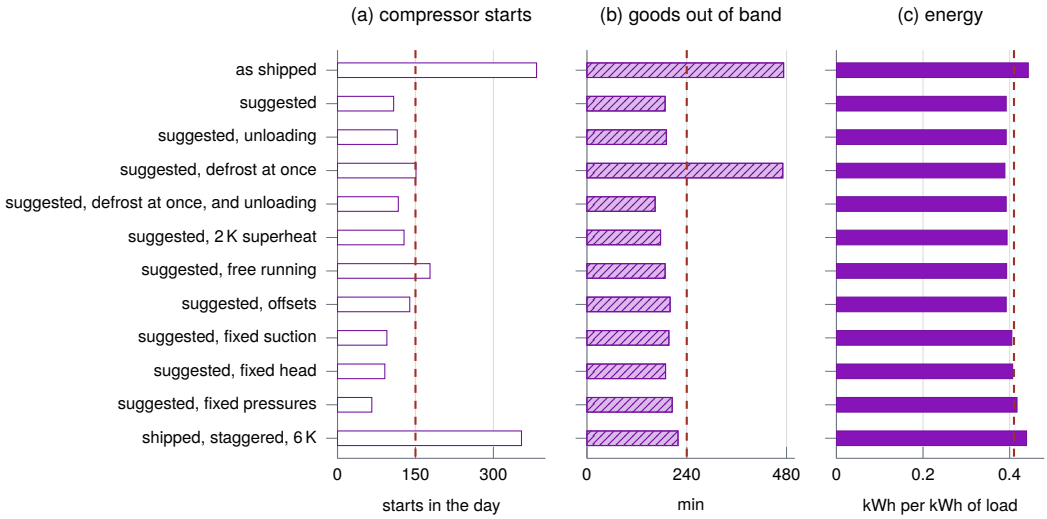


Figure 12.4 The summer day of Table 12.7 in its twelve configurations: (a) compressor starts, (b) the worst case's goods minutes out of band and (c) the specific energy, with the challenge limits dashed. The table prints the values. Notice that fixing both pressures gives the fewest starts and fails on energy alone, while the free-running thermostats fail on starts alone.

00110011. Only the first call is applied before replanning. This is a restricted profile library, not unrestricted compressor commitment or an electrical energy optimum. The mixed-integer formulation makes a small receding decision problem explicit alongside the usual MPC architecture [21], whose form in the apps is set out in Section 1.6.3.

Each case has nominal air, goods and coil estimates. With minutes as time, the predictor uses

$$320\dot{\hat{T}}_a = 60[.24(21-\hat{T}_a)+.35-U_g(\hat{T}_a-\hat{T}_g) \\ -1.4(\hat{T}_a-\hat{T}_w)]+320b, \tag{12.10}$$

$$1400\dot{\hat{T}}_g = 60U_g(\hat{T}_a-\hat{T}_g), \tag{12.11}$$

$$25\dot{\hat{T}}_w = 60[1.4(\hat{T}_a-\hat{T}_w)-.78v(\hat{T}_w-T_s)], \tag{12.12}$$

where $U_g = 1400/7200$ kW/K, v is the binary call and T_s is the current measured suction temperature, held throughout the forecast. The factor $.78 = 1.2(.65)$ fixes the wetted fraction at .65. Exact zero-order holds propagate each on/off interval. These coefficients stay at declared defaults when the

Table 12.8 Matched Three-Hour Profile Comparison

Method	Energy (kWh)	Rack starts	Air IAE (K min)
None	22.0537	45	557.51
Offsets	21.9225	24	481.92
Scheduler	21.5574	8	540.19
MPC	21.5184	24	384.14

learner changes the plant. Goods and coil estimates start at measured air temperature, then follow the nominal balances and the controller's own calls. Air is assimilated each two-second control scan. A rate-residual filter with gain .15 and bounds ± 1 K/min supplies b. It is a bounded disturbance correction, motivated by offset compensation [22], rather than a proof of zero offset for this profile library.

For each candidate, the cost sums absolute air error, twenty times air-band exceedance, .15 per on-minute and .4 per transition. Two continuous variables represent the peak number of simultaneous calls and of simultaneous openings. Their default prices are four and two. Four one-hot rows and sixteen coupling inequalities constrain sixteen binary and two continuous variables. Current manual calls are fixed, and an active defrost fixes its case off. Bands are soft. The defrost, superheat loops, rack staging, run/rest timers, pressure switches and manual stations keep authority. Measured air more than .6 K beyond either band edge overrides the call. The two-minute episodes inside the library therefore do not establish a dwell time on the executed sequence.

Only an optimal, finite, integral solution satisfying every bound and row, with a consistent objective and gap, is accepted. A rejected solve holds the current call vector on its first scan, then thermostat hysteresis takes over until a valid plan recovers. Historian marks and notes identify both transitions. The PIs continue their existing operation and tracking because the scheduler changes call permissions rather than their continuous commands. The first optimization occurs at initialization. Eight appended CSV fields include solver iterations, nodes, band exceedance, status, planning time and the largest predicted air temperature at the horizon. The scene badge displays that endpoint, which is not a prediction of the warmest goods.

An independent enumeration of all 256 combinations agrees with the MILP objective on twelve varied measurement packets. A separate RK4 integration agrees with the nominal exact hold within 1.5×10^{-14} K. These algebra checks do not validate the reduction against refrigerant transients. Removing boiling immediately when a call ends ignores liquid remaining in the coil. At $\hat{T}_w = -2$ and $T_s = -10$ degrees C with a .65 wetted fraction, that omission changes the coil rate by 14.976 K/min. Frost, superheat heat removal, future defrost, doors, weather and restocking are also absent. Goods remain unmeasured.

Table 12.8 and Figure 12.5 use case 10's nearly synchronous start, the default seed and three hours under each method. The MPC reduces measured air IAE summed across the cases from 557.51 to 384.14 K min and lowers energy and starts against free running. The gap scheduler makes fewer starts. A simple change of its gap to 1.5 min gives five starts at 21.8414 kWh, so low start counts do not require prediction. MPC makes 106 case cut-ins against 66 with the default gap scheduler. Pricing peak calls does not minimize motor starts.

A wider comparison of 44 runs covers defrost, floating suction, restocking, heat-wave operation, a half-size physics step, four scheduler gaps and two alternative concurrency prices. At 180 min in the heat wave, MPC gives 31.95 min of goods outside band against 27.88 with the gap scheduler and 16.14 with None. Halving air-to-coil conductance raises MPC air IAE to 232.78 K min against 163.19 for None. On the suggested day, MPC uses 175.3384 kWh and 103 starts against 183.5111 kWh and 108 for the suggested scheduler, but goods outside band rise from 188.22 to 215.02 min. Both pass. On the as-loaded day both fail with nine trips, despite MPC reducing starts from 383 to 184. Half-step runs preserve the rounded three-hour results. Six Chrome captures time sixty decisions

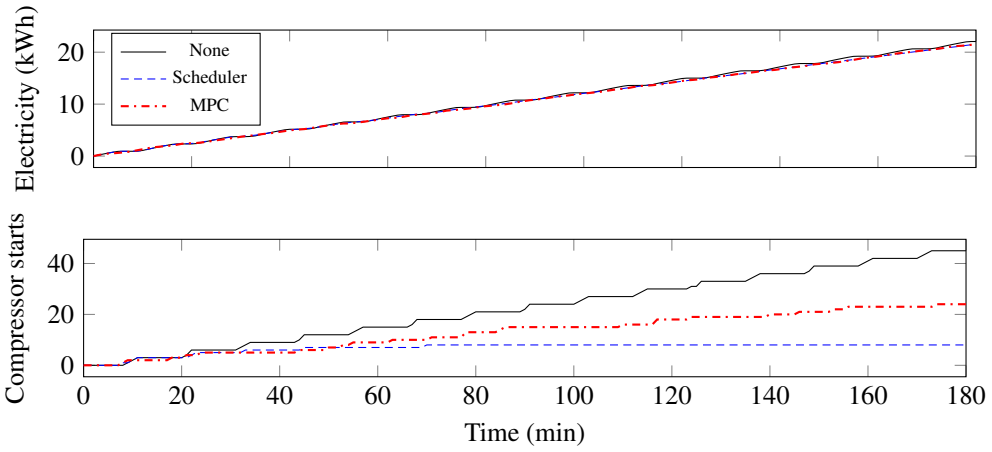


Figure 12.5 Case 10 uses the same instruments and initial phases under each choice: None is thin solid, the scheduler densely dashed and MPC thick dash-dotted. A smaller cumulative energy demand accompanies more compressor starts for MPC than for the gap scheduler.

each at .332–.360 ms per complete plan, .0332–.0360 percent of one core at one simulated minute per wall second. At the 900-fold top speed the same work costs 29.9–32.4 percent, excluding observation, physics and drawing. Prediction buys one trade, rather than replacing equipment capacity or the simpler scheduler.

12.7 Technology in Practice

What Measures It

In the basic control structure of a refrigeration system the suction pressure is measured at the compressor inlet, and with no pressure drop in the suction line it equals the evaporating pressure [16]. This chapter's model makes that assumption, one suction pressure for four cases. Each display case carries an air temperature sensor next to the goods, used in its control loop as an indirect measure of the goods temperature [16]. Case 9 shows what that indirection hides. The superheat of each coil needs a second pair of readings, the outlet temperature and the evaporation pressure.

> **TECHNOLOGY** **Sensor: Suction Pressure and a Clamped Thermometer**
>
> Superheat is not measured directly. It is the evaporator outlet temperature minus the evaporation temperature, and the evaporation temperature is normally found from a measurement of the evaporation pressure [14]. On one laboratory refrigeration system the pressure transmitters gave a voltage in proportion to the pressure, and platinum resistance thermometers were mounted on the copper tubes under insulation, so they read the tube rather than the refrigerant and added a small time constant [16]. Close to a flooded coil the outlet temperature grows restless: on two test systems its variance rose considerably at low superheat and fell again once the evaporator flooded [14]. The 6 s lag and 0.08 K noise of $\widehat{SH}_j$ in Table 12.1 stand for that clamped sensor.

What Moves It

The rack's compressors are switched on and off to hold the suction pressure, and one of them may run on a frequency converter for continuous action. The condenser pressure is held the same way, by switching fans or by changing their speed on a frequency converter [16]. The pack of this chapter has both kinds: C1 and C2 are staged while C3 trims on its speed, and PIC-2 sets the speed of three fans. Figure 12.6 draws the loops on that hardware.

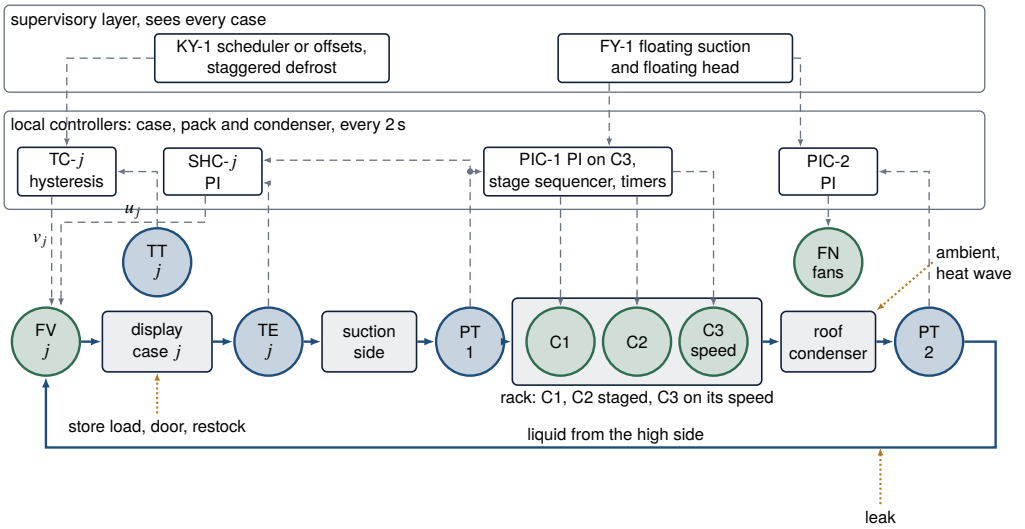


Figure 12.6 The chapter's loops on store hardware. In each case controller the thermostat TC-j reads TT-j and gives the inlet valve FV-j its call v_j, and SHC-j sets its opening u_j from the outlet temperature TE-j and the suction pressure PT-1. The pack controller holds the suction with C3's speed and stages C1 and C2 under their timers, and the condenser controller sets the fan speed from PT-2. The supervisory layer spreads the cases' calls and defrosts and floats both pressure setpoints. Dotted arrows mark disturbances.

> **TECHNOLOGY** **Final Element: An Inlet Valve That Opens and Pulses**
>
> In a traditional display case an electronic on/off inlet valve, opened and closed by the case's temperature controller, sits in series with a thermostatic expansion valve that holds the superheat while the inlet valve is open. In some versions the expansion valve is left out, and the inlet valve regulates the superheat itself by pulse width modulation [16]. A thermostatic valve meters the flow on the temperature difference across the evaporator, keeps liquid from returning to the compressor, and is stable enough that it typically needs no adjustment [23]. The response from opening to superheat is in general very nonlinear, which makes an electronic valve's loop hard to tune [14]. This chapter's valve carries both roles, the thermostat's call v_j and the superheat loop's opening u_j.

Where the Controller Runs

The control of a supermarket pack is distributed. The rack has its suction pressure controller, the condenser its pressure controller, and each display case a superheat and a temperature controller. A monitoring center receives data from the local controllers through a gateway and can change their setpoints and parameters, but the infrastructure allows only limited sharing of information directly between the local loops [16]. A building automation system of the same kind records historical data and alerts staff to a device failure or an alarm, and in such operational technology the supervisory and lower loops run with cycle times from milliseconds to minutes [24]. The chapter's controller scans every 2 s. Read as store hardware, the thermostats and superheat loops belong in the case controllers, PIC-1 and the stage sequencer in the pack controller, and the floating pressures, the scheduler and the staggered defrost in a supervisory layer that must see every case.

IN PRACTICE **Head Pressure, Superheat and Charge**

A federal retrofit guide for grocery stores sets the routine. Head pressure is set from the maker's chart for the refrigerant and the desired temperature, and lowered only in small increments while product temperatures are watched, because not every system meets its load at a lower head pressure. A check of the superheat confirms that the expansion valve works, and the suction pressure then falls in line. The charge is checked and the system inspected for leaks at least yearly. An under-charged system runs its compressor continuously at low suction and head pressures and misses its setpoints, and an over-charged one raises head pressure [23]. The leak of Table 12.1 starves this chapter's valves once the high side runs low.

What Is Modernizing It

Three developments touch this pack. Defrost cycles typically run on a time schedule whether a defrost is needed or not, and adaptive defrost controllers learn a baseline after installation and skip the unneeded cycles [23]. Head pressure reset combined with electronic expansion valves can lower the minimum condensing temperature to 50 to 60 °F, with a variable-speed drive on the condenser fan [23], which is the floating head pressure of Section 12.5. New systems have moved from conventional centralized direct expansion toward distributed, secondary loop, cascade and transcritical carbon dioxide systems, which typically leak less or use refrigerants with better environmental profiles [19]. Model-based setpoint optimization drove a test refrigeration system close to its optimum and cut its power consumption by up to 20 % [16], and the optimizers of Section 12.2 carry the same idea to the switching of whole packs.

12.8 Limitations

The cases are lumped, with no air curtain, shelf profile or product core, and frost carries no latent heat. The stable line is three declared factors on one exponential, a teaching reduction of a boundary that real evaporators show by experiment [13], and its numbers describe no particular coil. The suction side is one pressure with saturated vapor density and no line pressure drop, the compressors are ideal displacements that start at once, and the condenser is one temperature. The refrigerant comes from fits rather than from the standard formulation [20], and the R404A and R290 alternatives are approximate.

OPEN CHALLENGE **Controlling a Temperature Field, Not a Point**

Hysteresis control of display cases is kept mainly because it gives a uniform temperature through the case and a less harmful frost on the evaporator. If the frost and the temperature distribution in a case could be modeled, continuous control that does not synchronize could be developed. Such a model would probably be a set of partial differential equations, and the task would become control of a temperature distribution rather than of one point [16]. Case 9 shows the gap: the goods leave their band for 147 min while the air thermostat sees only its ordinary cycling (Table 12.6).

Every threshold is a teaching value. The 3 K line, the 56 °C cut-out, the −25 °C low-pressure stop, the goods band of 1 to 3 °C and the challenge's limits were set on this model, and real cut-outs follow the system's pressure rating, the compressor maker and the standard the application names. The weak dependence of the lock on the suction volume and the rack's tuning is measured on this model and its rack logic only. The thermostats' supervisor is rule-based. The profile MILP of Section 12.6.10 omits suction inventory and cannot reproduce the full hybrid optimizers in the benchmark literature [7, 9]. Its band penalties and warm-air overrides provide no goods-temperature guarantee. The numbers describe this model and no particular store.

OPEN CHALLENGE Keeping Many Cases Apart in Real Time

A desynchronizing predictive controller has held two display cases apart in simulation, but its complexity, and with it the computation time, grows with the size of the system, so a larger pack needs simplifications to make it practical [16]. A nonlinear scheme that tuned the cases' own controllers from above worked on a system of moderate size but not on larger ones, because its search space grew combinatorially, and it handled sudden large changes of the disturbances poorly [17]. Progress would be a coordinator that splits the problem into parallel subproblems of lower complexity, as the thesis proposed [16], and still keeps every case's goods in their band.

12.9 Conclusion

The simulated pack shows two failures of loops that behave well alone. A superheat loop that holds (5.98 ± 0.16) K at 6 K swings between 0.67 and 7.84 K at 2 K with the same gains, because below the 3 K line the coil's response steepens more than tenfold, slows threefold and wanders, and the opening at which a valve meets that line drifts 1.76 points in 24 minutes. Four independent thermostats lock in 65.2 min through the one suction pressure, and a larger suction volume does not prevent it (52.6 minutes at 5000 L), nor does the rack's tuning, but the lock never forms on a pressure no case can move. The cures are measured beside the faults. The scheduler cuts three hours' starts from 45 to 8, staggering cuts a defrost's peak rejection from 56.2 to 46.4 kW, floating suction saves 4.9 % of six hours' energy, and together with a 6 K superheat and floating head pressure they turn a day of 9 trips and 383 starts into one of none and 108. The model is lumped, its stable line and thresholds are teaching values, and its supervisor is rule-based. The hybrid scheduler lowers energy against free running in the matched case, while the gap scheduler makes fewer starts and model mismatch can warm goods. A fuller model would add humidity, refrigerant inventory and the latent load of frost to the predictor, line pressure drops, and frozen cases on a second suction group. The pack lets a reader measure the coupling that a shared utility puts between loops before meeting it in a plant.

KEY POINTS

- Four thermostats that never exchange a signal still pull each other into step, because every case boils at the one suction pressure the rack holds.
- A larger suction volume does not prevent the lock (52.6 minutes at 5000 L), but a pressure no case can move does.
- A superheat loop that holds (5.98 ± 0.16) K at 6 K hunts at 2 K with the same gains, and the cure is margin above the line, not a smaller gain.
- The scheduler cuts three hours' starts from 45 to 8, and staggering cuts a defrost's peak rejection from 56.2 to 46.4 kW.

Where else this applies. The rack's minimum run and rest times (Table 12.4) are the minimum dwell that the cooldown of Chapter 14 applies to a fleet of servers.

References

[1] S. A. Tassou, Y. Ge, A. Hadawey, and D. Marriott, "Energy consumption and conservation in food retailing," *Applied Thermal Engineering*, vol. 31, no. 2-3, pp. 147–156, 2011.

[2] L. F. S. Larsen, C. Thybo, R. Wisniewski, and R. Izadi-Zamanabadi, "Synchronization and desynchronizing control schemes for supermarket refrigeration systems," in *2007 IEEE International Conference on Control Applications*, 2007, pp. 1414–1419.

[3] D. Docimo and H. K. Fathy, "Demand response using heterogeneous thermostatically controlled loads: Characterization of aggregate power dynamics," *Journal of Dynamic Systems, Measurement, and Control*, vol. 139, no. 10, p. 101009, 2017.

[4] S. H. Strogatz, "From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators," *Physica D: Nonlinear Phenomena*, vol. 143, no. 1-4, pp. 1–20, 2000.

[5] X.-D. He, S. Liu, and H. H. Asada, "Modeling of vapor compression cycles for multivariable feedback control of HVAC systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 119, no. 2, pp. 183–191, 1997.

[6] L. F. S. Larsen, R. Izadi-Zamanabadi, and R. Wisniewski, "Supermarket refrigeration system - benchmark for hybrid system control," in *2007 European Control Conference (ECC)*, 2007, pp. 113–120.

[7] D. Sarabia and C. de Prada, "Hybrid predictive control of supermarket refrigeration systems: an optimal time switching strategy," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 10 916–10 922, 2008.

[8] N. L. Ricker, "Predictive hybrid control of the supermarket refrigeration benchmark process," *Control Engineering Practice*, vol. 18, no. 6, pp. 608–617, 2010.

[9] T. G. Hovgaard, L. F. S. Larsen, K. Edlund, and J. B. Jørgensen, "Model predictive control technologies for efficient and flexible power consumption in refrigeration systems," *Energy*, vol. 44, no. 1, pp. 105–116, 2012.

[10] W.-D. Gruhle and R. Isermann, "Modeling and control of a refrigerant evaporator," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 4, pp. 235–240, 1985.

[11] K. A. James and R. W. James, "Transient analysis of thermostatic expansion valves for refrigeration system evaporators using mathematical models," *Transactions of the Institute of Measurement and Control*, vol. 9, no. 4, pp. 198–205, 1987.

[12] A. Outtagarts, P. Haberschill, and M. Lallemand, "The transient response of an evaporator fed through an electronic expansion valve," *International Journal of Energy Research*, vol. 21, no. 9, pp. 793–807, 1997.

[13] W. Chen, Z. Chen, R. Zhu, and Y. Wu, "Experimental investigation of a minimum stable superheat control system of an evaporator," *International Journal of Refrigeration*, vol. 25, no. 8, pp. 1137–1142, 2002.

[14] K. Vinther, C. H. Lyhne, E. B. Sorensen, and H. Rasmussen, "Evaporator superheat control with one temperature sensor using qualitative system knowledge," in *2012 American Control Conference (ACC)*, 2012, pp. 374–379.

[15] B. P. Rasmussen and A. G. Alleyne, "Control-oriented modeling of transcritical vapor compression systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 126, no. 1, pp. 54–64, 2004.

[16] L. F. S. Larsen, "Model based control of refrigeration systems," Ph.D. dissertation, Department of Control Engineering, Aalborg University, Aalborg, Denmark, 2006, iSBN 87-90664-29-9. [Online]. Available: https://vbn.aau.dk/ws/files/549479206/thesis.pdf

[17] C. Sonntag, A. Devanathan, and S. Engell, "Hybrid NMPC of a supermarket refrigeration system using sequential optimization," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 13 901–13 906, 2008.

[18] L. S. Larsen and C. Thybo, "Potential energy savings in refrigeration systems using optimal set-points," in *Proceedings of the 2004 IEEE International Conference on Control Applications*, 2004, pp. 701–704.

[19] U.S. Environmental Protection Agency, "GreenChill: Keeping cool for fifteen years (2007–2022)," U.S. Environmental Protection Agency, Report, 2022. [Online]. Available: https://www.epa.gov/system/files/documents/2022-09/GreenChill-Keeping-Cool-for-15-Years-2022_1.pdf

[20] R. Tillner-Roth and H. D. Baehr, "An international standard formulation for the thermodynamic properties of 1,1,1,2-tetrafluoroethane (HFC-134a) for temperatures from 170 K to 455 K and pressures up to 70 MPa," *Journal of Physical and Chemical Reference Data*, vol. 23, no. 5, pp. 657–729, 1994.

[21] 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.

[22] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[23] National Renewable Energy Laboratory, "Advanced energy retrofit guide: Practical ways to improve energy performance; grocery stores," U.S. Department of Energy, Building Technologies Program, Report DOE/GO-102012-3655, revised, 2013. [Online]. Available: https://www.osti.gov/servlets/purl/1045045

[24] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (OT) security," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-82r3, 2023.

12.10 Exercises

12.1 Concept check · 2 min

In Case 4 the four thermostats run free from seeded phases and share nothing but the suction line. Why are they in step after about an hour?

(a) The suction loop is tuned too aggressively, and its own oscillation entrains the four cases

(b) The four cases carry identical loads, so their periods match and the phase gaps they start with persist

(c) The scheduler releases each case early once the last opening is old enough, which lines them up

(d) A case that opens raises the suction pressure and slows the cases already cooling, so the phases draw together

12.2 Concept check · 2 min

In Case 2 the superheat loop holds 6 K and then hunts after its setpoint is lowered to 2 K, a kelvin under the 3 K stable line. Which change does the chapter show restores a steady superheat?

(a) Raising the setpoint to 4 K, a kelvin above the line, with the same gains

(b) Lowering the loop gain at the 2 K setpoint until the swings die out

(c) Holding the valves in Manual at the opening where the steady curve meets the line

(d) Shortening the lag of the outlet sensor so that the loop sees each trough sooner

12.3 Concept check · 1 min

In Case 9 two fifths of the meat case's goods are replaced by goods at 10 °C, and the goods take about two and a half hours to return under 3 °C. Why does the air thermostat not prevent this?

(a) Its 2 K band is too wide, so the air is allowed to warm along with the goods

(b) It measures the air, which recovers within its usual cycling while the unmeasured goods lag by hours

(c) The scheduler holds back the meat case's cut-ins, so the case cannot cool while the goods are warm

(d) The restock raises the suction pressure, which takes cooling away from the meat case

12.4 Theory and calculation · 5 min

The order parameter of Equation (12.9) measures how close the thermostats are to cutting in together.

(a) Case 4 starts the four thermostats at phases 0.59, 0.67, 0.38 and 0.08 of a cycle, so φ_j is 2π times each. Compute R. *Checkpoint:* $R = 0.30$

(b) Show that for two cases a gap of g cycles gives $R = |\cos(\pi g)|$, and evaluate it for the dairy and produce cases of Section 12.4, which start 0.39 of a cycle apart and close to 0.10. *Checkpoint:* 0.34 and 0.95

(c) The status reads SYNCHRONIZED once R has stayed above 0.9 for 20 min. What is the largest gap between two cases that $R > 0.9$ allows, in cycles and in minutes of an 11.4 min cycle? *Checkpoint:* 0.144 of a cycle, 1.6 min

12.5 Theory and calculation · 7 min

Equation (12.3) joins two branches at the dry length λ_{mss}.

(a) Differentiate both branches with respect to λ at λ_{mss}, and show that the slope below the line is the slope above it times $\rho = G\,SH_{mss}/\left[(\Delta T - SH_{mss})(-\ln(1 - SH_{mss}/\Delta T))\right]$, whatever a is. *Checkpoint:* ρ as given

(b) Evaluate ρ at the start of Case 2: air at 2.14 °C, suction at −6 °C, $G = 8$ and $SH_{mss} = 3$ K. *Checkpoint:* $\rho = 10.2$

(c) The superheat loop has a loop gain of 1.24 at 6 K. Taking the slope above the line as a rough stand-in for the slope at 6 K, estimate the loop gain just below the line, and give the superheat lag there from Equation (12.4) with $\beta = 1$. *Checkpoint:* about 13, 54 s

12.6 Theory and calculation · 6 min

The condenser's conductance at fan speed $n = 1$ with N_{out} of its three fans failed follows the $UA_c(n)$ law of Section 12.3 with $UA_c = 2.7$ kW/K and $UA_{c0} = 0.35$ kW/K.

(a) Compute the conductance with no fan, one fan and two fans out. *Checkpoint:* 2.70, 2.05 and 1.33 kW/K

(b) Section 12.4 finds that the four cases at 2 °C need a condensing temperature of 48.7 °C with all fans into air at 38 °C. How much heat does the condenser reject there? *Checkpoint:* about 28.9 kW

(c) Holding that heat, estimate the condensing temperature with one and with two fans out. Compare with the chapter's 52.5 and 62.6 °C and the 56 °C cut-out, and say what your estimate leaves out. *Checkpoint:* 52.1 °C and 59.8 °C

12.7 App challenge · 8 min

Load Scenario 2, *The superheat loop and the stable line*, at the default seed and change nothing. Run its 50 min at speed 5 (about 10 s) and export the CSV file.

(a) From column `SH1_K`, the dairy coil's measured superheat, give its mean over the first 10 min and its lowest and highest values from 10 to 30 min. *Checkpoint:* 5.98 K; 0.74 and 7.82 K

(b) At what time does column `hunting` first read 1? *Checkpoint:* 18.9 min

(c) Give the mean of `SH1_K` from 36 to 50 min, and explain the difference between the three periods. *Checkpoint:* 3.99 K

12.8 App challenge · 8 min

Load Scenario 4, *Four thermostats fall into step*, at the default seed and change nothing. Run the 3 h at speed 20 (about 9 s) and export the CSV file.

(a) Read `sync_R` at 20 min and the first time at which column `status` reads SYNCHRONIZED. *Checkpoint:* 0.59; 65.2 min

(b) Read `metric_starts` at 60 and at 180 min, and give the starts in the first hour and in the next two. *Checkpoint:* 14 and 30

(c) In the `sync_R` column, find when R last rose above 0.9 before the status changed, and explain the difference from (a). *Checkpoint:* 45.2 min

12.9 App challenge · 7 min

Load Scenario 7, *Floating suction: the energy and the case that sets it*, at the default seed and change nothing. Run the 6 h at speed 60 (about 6 s), and read the Energy and Compressor starts tiles and the case the FY-1 card names as binding. Then reload the scenario, set Floating suction pressure to Off before Run, and run the 6 h again.

(a) Give the energy and the starts of both runs. *Checkpoint:* 43.2 and 45.4 kWh; 27 and 58

(b) What share of the energy does the float save? *Checkpoint:* 4.9 %

(c) Which case binds, and why that one? *Checkpoint:* case 2, the meat case

12.10 Application area · 4 min

Case 6 gives each coil 2 kg of frost and defrosts all four at once.

(a) What share of the clean coil's air-to-coil conductance is left with 2 kg of frost, and after how many hours of frosting at the declared rate does a wet, cold coil carry that much? *Checkpoint:* 71 %; 4 h

(b) How much heat does melting 2 kg take, and how long does a 2 kW heater need for that alone? Compare with the measured defrost length. *Checkpoint:* 668 kJ; 5.6 min

(c) A store engineer asks why the defrosts should be staggered. Answer from the case's results. *Checkpoint:* 56.2 kW against 46.4 kW

12.11 Application area · 4 min

Floating head pressure lets the condensing temperature follow the ambient down to a floor of 25 °C.

(a) With the cycle coefficients of performance of Section 12.4, 4.31, 2.70 and 1.80 at 25, 40 and 55 °C condensing, what compressor power carries the 19.64 kW of case load at each, and what does floating from 40 to 25 °C save? *Checkpoint:* 4.56, 7.27 and 10.91 kW; 37 %

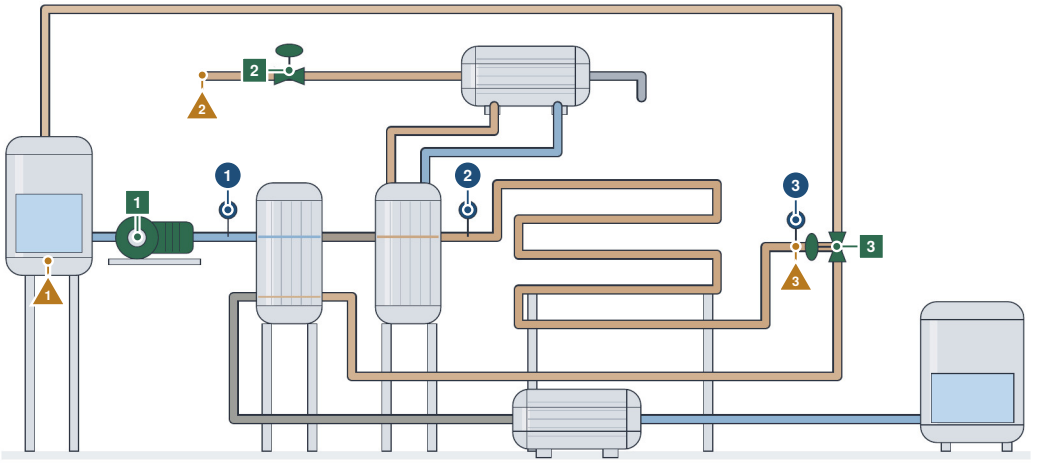
(b) The valve law makes the flow at a given opening proportional to $\sqrt{P_c - P_{suc}}$. With the saturation fit of Table 12.2 and the suction at −10 °C, what share of its capacity at 40 °C condensing does a valve keep at 25 °C, ignoring the change in liquid density? Why is there a floor? *Checkpoint:* about 0.75

12.12 Review the agent · 5 min

An AI assistant reviewing a store's rack writes: "The cases fall into step because each valve opening kicks a small 250 L suction side. Add a 5000 L buffer volume. The suction pressure will then move twenty times less per opening, the coupling will be twenty times weaker, and the cases will stay apart without any scheduler." Find the flaw, and say how you would check the claim in the app.

CHAPTER 13

Milk Pasteurizer Simulator for Product History, Regeneration and Flow Diversion



The controller moves 1 the pumps that move and time the milk (one is drawn)
2 the steam valve that heats the hot water 3 the flow diversion valve, forward or back to the tank

The controller reads 1 the flow meter after the pump 2 the milk temperature leaving the heater
3 the recorder and indicating thermometers at the end of the tube

What disturbs it 1 colder raw milk in the balance tank 2 a sag in the steam supply
3 both end-of-tube thermometers drifting high

Raw milk is warmed by treated milk, heated, held in a long tube and then sent on or back by one valve, so notice that the only judge of each parcel is the thermometer pair just before that valve. Illustration, not to scale.

IN PLAIN TERMS Hot Enough, Long Enough, and Proven

The main idea. A dairy heats raw milk to kill germs and keeps it hot for a set time in a long pipe. A valve at the end of the pipe sends the milk on only when the thermometers say it was treated, and sends it back to be heated again otherwise.

The challenge. The controller sees the thermometers but never the milk itself. Faster flow cuts the time in the pipe while the milk stays hot, and two thermometers that drift together can pass untreated milk while every check looks fine.

Why it matters. Milk that has gone on cannot be called back, so the plant must be sure of every part of it before it leaves.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact in the Introduction, the history in Section 13.2, the two try-its and the fact in Section 13.6, and the key points
Study	45 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 13.6), Technology in Practice (Section 13.7), Limitations (Section 13.8) and the open challenge

The control question A pasteurizer is judged on the time and temperature each parcel of milk received, which no controller measures. What can the temperature cascade, the diversion supervisor and its instruments guarantee about that history, and what can none of them see?

Learning objectives

- Compute the holding time and the legal flow from the tube volume and the recipe.
- Derive the regenerator's effectiveness and its recycle gain, and use them to explain why the plant is slow and sensitive to steam.
- Explain how conditional integration on the outer loop shortens the return from a diversion.
- Distinguish a fault that a comparison of two thermometers catches from one it cannot see.
- Choose a setpoint margin, a feedforward and a tuning that ride out a campaign without a diversion or excess heat.

Prerequisites Sections 1.4.2 (the dose as a guard), 1.4.7 (switching, latch and persistence), 1.6.2 (the PI algorithm and its anti-windup) and 1.6.4 (scenarios, scorecard and data).

The app https://apmonitor.com/apps/food-pasteurization/

See also Volume 3, Chapter 2 for matching a loop structure to where a disturbance enters, Volume 3, Chapter 3 for the speed a cascade's inner loop must reach, and Volume 5, Chapter 7 for diverting the product of a continuous line on a residence time.

13.1 Introduction

A process controller is usually judged by how closely a measured variable follows its setpoint. A pasteurizer is judged by something no controller measures, the time and temperature that each unit of product actually received, and the plant is allowed to send product back whenever that is in doubt. Thermal processing of foods is full of this shape. Retorts for canned foods are controlled on line during sterilization [1], high-temperature short-time (HTST) milk systems have been studied with control strategies built on the lethality delivered [2], and an automated HTST pasteurizer has been run with its flow diversion valve driven by a lethality controller [3].

The HTST milk pasteurizer is the familiar industrial case. Raw milk from a constant-level balance tank is warmed in a regenerator by milk that has already been treated, brought to the recipe temperature with hot water, and held in a tube sized so that the product spends the recipe time in it. A flow diversion valve (FDV) at the end of the tube then sends each part of the stream forward to be cooled, or back to the balance tank. The stringency of the time and temperature standards was set on the heat resistance of *Coxiella burnetii* [4, 5]. The simulator assumes a recipe of 72 °C held for 15 s, which its briefing takes from the U.S. Pasteurized Milk Ordinance minimum for milk (161 °F, rounded up). The same ordinance sets the standards against which the holding time of an HTST plant is checked [6].

DID YOU KNOW? From Fifteen Seconds to a Hundredth

The ordinance accepts more than one recipe for continuous flow. Milk held at 72 °C for 15 s is pasteurized, and so is milk held at 89 °C for 1.0 s or at 100 °C for 0.01 s. The temperature rises by 3 °C for a product with ten percent fat or more, eighteen percent total solids or more, or added sweeteners [7].

Three properties separate this plant from a temperature loop with a dead time. The holding time is the tube volume divided by the flow, so a flow increase shortens the treatment of every parcel while the temperature trend looks unchanged, and a real holding tube passes a spread of residence times whose fastest particle decides compliance [6, 8]. The regenerator returns heat from the product to the product, which makes the plant slow and sensitive to its heat input, and the return stops the moment product is diverted. The diversion itself rests on instruments that lag, drift and can fail together, while product that has passed them cannot be recalled. Product can also be contaminated after the tube,

since over 16 000 culture-confirmed cases of salmonellosis in Illinois were traced to pasteurized milk from a single dairy plant, with evidence suggesting that the strain had persisted in the plant and contaminated milk after pasteurization [9].

The simulator addresses the belief that product at its temperature setpoint is treated product. The class of problem is general, since a retort must also deliver a sterilizing treatment that is a history rather than a reading [1], and any continuous process adds a residence that follows the throughput. Volume 5, Chapter 7 routes the product of a continuous drug-substance line to a divert tank on the same question of which material an upset touched. A transfer function whose dead time is a constant represents neither. However, predictive control studies of pasteurizers address the temperature loop [10, 11] rather than the routing of the product and the record of what it received. The record is the dose idea of Section 1.4.2 applied to every parcel of a continuous stream.

This chapter presents a regenerative HTST milk pasteurizer simulator in which the product moves as parcels that each keep their own record, so that the holding time follows the flow, the regenerator exchanges heat parcel against parcel, and the supervisor that routes the product can be compared with what the product actually received. HTST pasteurization, cascade control with feedforward, conditional integration and diversion logic are standard practice. The simulator rests on a parcel model with an exactly conservative counterflow regenerator and a truth log kept apart from the instruments, and the chapter gives enough detail to rebuild it. Eight cases and a challenge separate what the controller, the supervisor and the instruments can and cannot guarantee.

Section 13.2 reviews the standards, the holding-time checks and earlier computer control of HTST plants. Section 13.3 sets out the parcel model: its sections, heat transfer, hot-water circuit, pressure differential, parcel records and parameters. Section 13.4 treats the regenerator as a filter and a recycle, and Section 13.5 gives the instruments, the cascade with feedforward and the supervisor as implemented. Section 13.6 measures the eight cases, the supervisor's margins and the challenge. Section 13.7 describes the real thermometers, diversion device and controllers of a plant, Section 13.8 states the limitations, Section 13.9 concludes, and the exercises of Section 13.10 run the cases.

13.2 Pasteurization, Holding Time and Diversion

The history of milk pasteurization and of the standards that govern it has been reviewed [12]. The standards name a temperature and a time. Their stringency was set on the requirements for *Coxiella burnetii*, regarded as the most heat-resistant organism of public health significance, and the review that traced this choice also questioned it [5]. The pasteurization of milk containing the organism of Q fever had been studied in 1957 [4]. The history cited above judges the present standards adequate for public health, with the exception of the endospores of *Bacillus cereus*, provided good manufacturing practices are followed [12], which places as much of the burden on the plant as on the recipe.

> ⧖ HISTORY **1924 to 2025: A Model Ordinance for Milk**
>
> Studies by the United States Public Health Service at the turn of the century on the role of milk in spreading disease led to research on sanitary measures, including studies that improved the pasteurization process. In 1924 the Service developed a model regulation, the Standard Milk Ordinance, for voluntary adoption by state and local agencies, and a Code published in 1927 gave the details of compliance. Its successor, the Grade "A" Pasteurized Milk Ordinance, reached its 34th revision in 2025. In 1938 milkborne outbreaks were 25 % of all outbreaks traced to infected foods and contaminated water, and milk and fluid milk products are now associated with less than 1 % of them [7].

Continuous flow turned the time half of the recipe into a flow problem. A method for checking the holding time of short time high temperature pasteurizers appeared in 1939 [13], and the residence time of milk products in HTST holding tubes was measured in 1968 [8]. Tracer tests on a pilot-scale HTST pasteurizer, run to establish whether it met the Pasteurized Milk Ordinance standards, found the fastest moving particle leaving a nominal 21.5 s holding section after 18.6 s, with the section

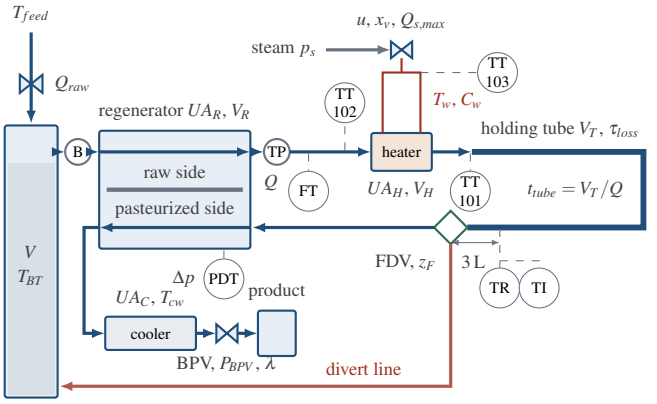


Figure 13.1 Regenerative HTST pasteurizer as modeled. Raw milk at T_{feed} enters the balance tank through a float valve (Q_{raw}), passes the booster pump B, the raw side of the regenerator, the timing pump TP that sets the flow Q, the flow meter FT and TT-102, and is heated by hot water at T_w fed by the steam valve at position x_v. TT-101 reads the heater outlet at the tube inlet. The recorder TR and the indicating thermometer TI read the product 3 L before the FDV, which sends it back through the pasteurized side of the regenerator, the cooler and the back-pressure valve BPV to the product tank, or along the divert line to the balance tank. PDT measures the pressure differential Δp across the regenerator plates.

45 % back-mixed [6]. Compliance is therefore judged on the fastest particle rather than on the mean residence.

Computer control of HTST pasteurizers was evaluated against the regulated instruments. A computer-based system controlled the hot-water temperature, the flow through a centrifugal timing pump and the diversion of underprocessed product, compared single-loop, cascade and multivariable strategies, drove the flow diversion valve with a lethality controller, and logged temperatures within ±0.1 °C of the safety thermal limit recorder [3]. Lethality-based control strategies for HTST systems followed [2], and model predictive control of the pasteurization temperature has been demonstrated [10, 11]. On the equipment side, dynamic models of plate heat exchangers follow milk fouling through the channels [14]. Milk fouling involves protein denaturation and aggregation, is reduced by lower surface temperatures and higher velocities, and costs billions of dollars every year [15].

13.3 Parcel Model of the Pasteurizer

The simulated plant processes 9000 L/h of milk at 6 °C through a 45 L holding tube, an 18.0 s hold, with a 900 kW steam valve on the hot-water circuit. The balance tank, the regenerator with its raw and pasteurized sides, the timing pump, the heater and hot-water circuit, the two passes of the holding tube, the FDV, the cooler and every instrument are shown in Figure 13.1.

Table 13.1 lists the variables by role. Eight lumped states are integrated together with the temperatures of the parcels, which occupy 128 state slots of which 96 are in use at the design flow. The positions, routes and records of the parcels are discrete plant state, updated after each integration step.

13.3.1 Parcels and Sections

The milk is a chain of parcels of volume $\Delta V = 2\,\text{L}$. Each parcel i has a center s_i, measured in liters along its path, a temperature T_i and a record. Three paths exist, the line from the balance tank to the FDV, the forward path from the FDV to the product tank and the divert line back to the balance tank. After every step of length Δt the parcels on the first path advance by $Q\Delta t$, those on the forward path advance by the same amount only while the FDV points forward, and those on the divert line only while it points to divert. A parcel reaching the FDV joins the path the valve points to, a parcel

Table 13.1 Roles and Design Values of the Pasteurizer Variables

Variable	Role	Design value	Range
T_{101}	CV, TT-101	73.5 °C SP	72–78
T_w	inner CV, TT-103	77.0 °C	60–95
u	MV, steam valve	12.5 %	0–100
$T_{w,SP}$	MV of TIC-101	77.0 °C	60–95
FDV	MV of the supervisor	forward	or divert
Q_{SP}	timing pump setpoint	9000 L/h	6000–12 000
T_{feed}	DV, raw feed	6 °C	1–20
p_s	DV, steam supply	100 %	0–110
r_f	DV, fouling rate	0 %/min	0–5
b_R, b_I	DV, thermometer biases	0 °C	±3
λ_{set}	DV, BPV spring loss	0 %	0–60
Q, V	states, flow, tank volume	9000 L/h, 140 L	—
VT_{BT}	state, tank heat content	$T_{BT} = 6$ °C	—
T_w, x_v	states, hot water, valve	77.0 °C, 0.125	—
R_f, Δp, λ	states	0, 25.0 kPa, 0	—
T_i	states, parcels	128 slots	—

reaching the product tank is delivered, a diverted parcel reaching the balance tank is mixed into it, and new parcels leave the tank at its temperature T_{BT}. Residence time is therefore whatever the flow makes it, parcel by parcel.

Each section X (heater, tube, regenerator side or cooler) occupies an interval $[X_0, X_1]$ of its path, and a parcel feels it in proportion to its overlap, as shown in Equation (13.1).

$$w_{X,i} = \frac{\left|\left[s_i - \frac{\Delta V}{2},\, s_i + \frac{\Delta V}{2}\right] \cap [X_0, X_1]\right|}{\Delta V} \tag{13.1}$$

The weights and the regenerator pairs below are frozen through one integration step, an eighth of the 0.8 s a parcel takes to move its own length at 9000 L/h.

13.3.2 Heat Transfer on Each Parcel

On the line from the balance tank, a parcel is heated in the heater and loses heat in the tube, and on the forward path it is chilled in the cooler, as shown in Equation (13.2).

$$\begin{aligned} \frac{\mathrm{d}T_i}{\mathrm{d}t} &= w_{H,i}\, k_H\, (T_w - T_i) + \frac{w_{T,i}}{\tau_{loss}}\, (T_{amb} - T_i) + \dot{T}_{R,i} \\ \frac{\mathrm{d}T_j}{\mathrm{d}t} &= w_{C,j}\, k_C\, (T_{cw} - T_j) - \dot{T}_{R,j} \end{aligned} \tag{13.2}$$

where i is a raw-side parcel, j a forward parcel, $k_H = UA_H/[(1+R_f)\,\rho c_p V_H]$ and $k_C = UA_C/(\rho c_p V_C)$ are rate constants in s^{-1}, R_f is the fouling resistance as a fraction of the clean heater resistance, T_{amb} the room temperature and T_{cw} the ice water. Parcels on the divert line neither gain nor lose heat. Evaluated at steady flow, the heater term reproduces the familiar outlet relation $T_{out} = T_w - (T_w - T_{in})\,\mathrm{e}^{-NTU_H}$ with $NTU_H = UA_H/[(1+R_f)\rho c_p Q]$, which is 1.38 at the design flow.

The regenerator couples every raw parcel i to every pasteurized parcel j that overlaps it along the exchanger, as shown in Equation (13.3).

$$\dot{T}_{R,i} = \sum_j k_R\, \ell_{ij}\, (T_j - T_i), \qquad \dot{T}_{R,j} = \sum_i k_R\, \ell_{ij}\, (T_j - T_i) \tag{13.3}$$

Here $k_R = UA_R/(\rho c_p \Delta V)$, and ℓ_{ij} measures the overlap of the two parcels as a fraction of the regenerator length, with both measured from the raw inlet so that the pasteurized parcels run in

counterflow. Because the two parcels of a pair have the same volume, the heat one side gains is exactly what the other loses, and energy is conserved pair by pair. At balanced flow the arrangement approaches the counterflow effectiveness of Equation (13.4) [16].

$$\varepsilon = \frac{NTU_R}{1+NTU_R}, \qquad NTU_R = \frac{UA_R}{\rho c_p Q} \tag{13.4}$$

At 9000 L/h, $NTU_R = 5.67$ and $\varepsilon = 0.850$, and the parcel regenerator delivers 0.847 at the settled steady state of the first case, within the 1 % that the application's own check requires.

13.3.3 Hot Water, Steam, Pump and Balance Tank

The hot-water circuit is one well-mixed capacitance heated by steam and cooled by the parcels in the heater, as shown in Equation (13.5).

$$\begin{aligned} C_w \frac{\mathrm{d}T_w}{\mathrm{d}t} &= Q_{s,max}\, x_v \frac{p_s}{100} - Q_h - UA_{wl}(T_w - T_{amb}) \\ Q_h &= \rho c_p \Delta V \sum_i w_{H,i}\, k_H\, (T_w - T_i) \end{aligned} \tag{13.5}$$

Here $Q_{s,max}$ is the heat the fully open valve passes at design supply pressure, $x_v \in [0,1]$ the valve position and p_s the supply pressure in percent of design, so that a pressure loss removes heat at an unchanged valve position. The valve, the timing pump and the balance tank follow Equation (13.6).

$$\begin{aligned} \tau_v \frac{\mathrm{d}x_v}{\mathrm{d}t} &= \frac{u}{100} - x_v, \qquad \tau_p \frac{\mathrm{d}Q}{\mathrm{d}t} = Q_{SP} - Q \\ \frac{\mathrm{d}V}{\mathrm{d}t} &= Q_{raw} - Q, \qquad \frac{\mathrm{d}(VT_{BT})}{\mathrm{d}t} = Q_{raw} T_{feed} - Q\, T_{BT} \\ Q_{raw} &= \mathrm{clamp}\big(K_{fv}(V_{set} - V),\, 0,\, Q_{raw,max}\big) \end{aligned} \tag{13.6}$$

where Q_{raw} is the flow through the float valve. A diverted parcel arriving at the tank adds ΔV to V and $\Delta V\, T_i$ to VT_{BT} at once, so a long diversion fills the tank with hot product and closes the float valve. Fouling grows at the rate set by the disturbance slider, $\mathrm{d}R_f/\mathrm{d}t = r_f/6000$ with r_f in percent of the clean resistance per minute, so that the heater keeps $1/(1+R_f)$ of its clean UA_H.

13.3.4 Pressure Differential and the Diversion Valve

The pasteurized side of the regenerator is held above the raw side by the back-pressure valve and the friction downstream of the regenerator, while the booster pump pressurizes the raw side. The differential follows its algebraic target through a one-second lag, given by Equation (13.7).

$$\begin{aligned} \tau_{\Delta p} \frac{\mathrm{d}\Delta p}{\mathrm{d}t} &= \Delta p^* - \Delta p, \qquad q = Q/Q_0 \\ \Delta p^* &= P_{BPV}(1-\lambda) + \phi\, P_f\, q^2 \\ &\quad - \beta\left(P_B - P_d\, q^2\right) - (1-\beta)\, P_{suc} \end{aligned} \tag{13.7}$$

where $Q_0 = 9000\,\mathrm{L/h}$, $\phi = 1$ in forward flow and 0 otherwise, $\beta = 1$ while the booster runs and 0 when it is stopped, and the spring loss creeps toward its disturbance value by $\tau_{cr}\,\mathrm{d}\lambda/\mathrm{d}t = \lambda_{set} - \lambda$. At the design flow $\Delta p = 25.0\,\mathrm{kPa}$, and with the booster stopped during a diversion it is 90.0 kPa. The FDV position z_F moves toward 1 (forward) or 0 (divert) at a rate of $1/\tau_{FDV}$ and routes a crossing parcel forward while $z_F \geq 0.5$.

13.3.5 The Parcel Record and the Truth Log

Each parcel carries two clocks for the holding tube. In every step the fraction of the step its center spends inside the tube is $f_i = |[s_i, s_i + Q\Delta t] \cap [T_0, T_1)|/(Q\Delta t)$, exact for steady flow, and the clocks

Table 13.2 Parameters of the Pasteurizer. The Parameter Panel Shows and Adjusts the Visible Ones and Omits the Hidden Ones

Symbol	Meaning	Value	Class
T_{rec}, t_{hold}	recipe	72.0 °C, 15 s	visible
Δp_{min}	interlock trip	6.9 kPa	visible
V_T	holding tube	45 L	hidden
τ_{loss}	tube heat loss	3000 s	visible
UA_H, V_H	heater, clean	14.0 kW/K, 12 L	vis., hid.
UA_R, V_R	regenerator, per side	57.4 kW/K, 30 L	vis., hid.
UA_C, V_C	cooler	25.3 kW/K, 20 L	hidden
T_{cw}, T_{amb}	ice water, room	1.5, 20 °C	hidden
C_w	hot-water capacity	400 kJ/K	visible
$Q_{s,max}$, τ_v	steam valve	900 kW, 3 s	visible
UA_{wl}	hot-water loss	100 W/K	hidden
V_{set}, K_{fv}	float valve	150 L, 0.25 s^{-1}	hidden
$Q_{raw,max}$, τ_p	float capacity, pump	21 600 L/h, 2 s	hidden
P_{BPV}, P_B	BPV, booster	80, 80 kPa	visible
P_f, P_d, P_{suc}	friction, droop, suction	20, 5, −10 kPa	hidden
$\tau_{\Delta p}$, τ_{cr}	pressure, creep	1, 60 s	hidden
τ_{FDV}	FDV travel	0.5 s	visible
τ_{101}, τ_R, τ_{TI}	thermometer lags	2, 1.5, 2 s	visible
τ_{102}, τ_{103}	thermometer lags	2, 3 s	hidden
τ_{FT}, τ_{PDT}	flow, Δp lags	1, 0.5 s	hidden
—	plausibility tolerance	0.5 °C	visible
ρ, c_p	milk	1030 kg/m^3, 3.93 kJ/(kg K)	fixed
ΔV	parcel, slots	2 L, 128	fixed

advance as shown in Equation (13.8).

$$t_{T,i} \leftarrow t_{T,i} + f_i \Delta t, \qquad t_{h,i} \leftarrow t_{h,i} + f_i \Delta t \ \text{ if } T_i \geq T_{rec} \tag{13.8}$$

where $t_{T,i}$ is the time in the tube and $t_{h,i}$ the continuous hold. One step below T_{rec} sets $t_{h,i}$ to zero for good. At the FDV the parcel is treated if $t_{h,i} \geq t_{hold}$, and otherwise underprocessed, by temperature if its hold was broken and by time if it was not. The time in the tube is therefore $t_{tube} = V_T/Q$, 18.0 s at 9000 L/h, and the largest flow that still holds every parcel for the recipe time is $Q_{legal} = V_T/t_{hold} = 10\,800\,\text{L/h}$. Two more records are kept. The heat above the recipe, $E_i = \int \max(0, T_i - T_{rec})\,dt$ over the whole path in °C s, measures overprocessing without quality kinetics, and a forward parcel inside the pasteurized side of the regenerator while $\Delta p < 0$ is flagged as a raw-contact risk, since a pinhole would then leak raw milk into it. No leak is simulated. The scorecard counts these records, not the instruments, and the controller never reads them.

13.3.6 Numbers, Geometry and Initialization

The parameters are listed in Table 13.2. The fixed pipe runs are 4 L from the balance tank to the regenerator, 6 L from the regenerator to the heater with TT-102 at their middle, 3 L from the heater to the tube with TT-101 at the tube inlet, 6 L from the FDV to the pasteurized side, 8 L from there to the cooler, 4 L on to the product tank and 24 L in the divert line. The recorder and the indicating thermometer sit 3 L before the FDV, inside the end of the tube, so that the 1.2 s of travel from them to the valve covers most of the recorder's lag and the valve's travel. Every scenario starts from the closed-form balanced steady state of the line, with linear profiles in the regenerator, the balance tank at $V = V_{set} - Q/K_{fv} = 140\,\text{L}$ and the steam valve at the position that balances the heater duty, 12.29 %. Integration is classical fourth-order Runge-Kutta with a 0.1 s step. Halving the step moves the steady heater inlet by 0.004 °C at 400 s of the first case, inside the 0.02 °C that one of the application's 36 passing checks allows.

13.3.7 Assumptions

Plug flow holds in every section, so each parcel moves at the mean velocity and its time in the tube is exactly V_T/Q. A real holding tube has a velocity profile, and its fastest particle can leave well before the mean residence [6, 8], so the model overstates the hold of a real tube at the same flow.

Heat transfer is lumped per section with constant milk properties, one well-mixed hot-water temperature and no heat capacity in the plates. Plate metal would slow the loss of preheat after a diversion without preventing it, and a channel-by-channel model is available where that detail matters [14].

Fouling is a prescribed growth of resistance, not a deposition model. Real milk fouling depends on the surface temperature [15], which the cascade raises as it compensates, so a real run fouls faster toward its end.

The recipe is one temperature and one time. No microbial kinetics (decimal reduction times and z values) are modeled, so a parcel held at 71.9 °C counts as fully underprocessed.

Pressures are algebraic with a one-second settling, and a raw-contact risk is a flag rather than a contamination. The instruments are first-order lags with colored noise and an additive bias, with no failed-open or failed-shut modes, and the timing pump is a first-order drive with no pump curve.

13.4 Regeneration as a Recycle

The regenerator does three things at once, and each shapes the control problem. Averaged from 60 to 300 s of case 1, it transfers 576 kW against 107 kW in the heater, 84 % of the heating. The steam supplies 112.8 kW, of which 5.7 kW is lost from the hot-water circuit, and the plant uses 12.5 kWh of steam per 1000 L while the milk takes up about 76 kWh per 1000 L on its way from 6 to 73.5 °C.

It is a filter for feed disturbances. With the product temperature fixed, a change at the raw inlet reaches the heater inlet multiplied by $1-\varepsilon = 0.15$.

It is also a recycle. A change in the heater outlet T_{101} travels through the tube, the FDV and the pasteurized side and returns to the heater inlet about 24 s later at 9000 L/h, multiplied by ε and the tube loss, and the heater passes $\mathrm{e}^{-NTU_H} = 0.251$ of its inlet to its outlet. With the measured effectiveness the loop gain is therefore $g = \varepsilon\,\mathrm{e}^{-NTU_H}\,a_T = 0.211$, with $a_T = \mathrm{e}^{-V_T/(Q\tau_{loss})} = 0.994$, and the steady gain from the hot water to the product rises from the heater's own $1-\mathrm{e}^{-NTU_H} = 0.749$ to Equation (13.9).

$$\begin{aligned} \frac{\partial T_{101}}{\partial T_w} &= \frac{1-\mathrm{e}^{-NTU_H}}{1-g} = 0.950 \\ \frac{\partial T_{101}}{\partial T_{feed}} &= \frac{\mathrm{e}^{-NTU_H}(1-\varepsilon)}{1-g} = 0.0486 \end{aligned} \tag{13.9}$$

The recycle matters most to the hot water. With the valve fixed, the heater draws $\rho c_p Q(1-\mathrm{e}^{-NTU_H}) = 7.6\,\mathrm{kW/K}$ from the circuit, which with the circuit's own loss would let it settle in about 52 s, with 1.17 °C of hot water for each percent of valve. Most of that heat comes back through the regenerator, and only the heat that leaves with the product, the tube loss and the circuit's own loss is really removed, 1.61 kW/K in all. In manual, a 1 % step of the steam valve from 12.5 to 13.5 %, taken as the difference from an identical run without the step so that the slow drift of a valve held just below its settled balance of 12.53 % cancels, raises the hot water 0.82 °C in the first minute, 2.82 °C in five minutes and 5.57 °C in the end, with 63 % of the rise taking 429 s, and the energy balance of the step closes to within 0.01 kW of the added 9 kW. The regenerator makes the plant almost five times as sensitive to steam and about eight times slower. The same property makes a small manual error persist, since a valve held at 12.4 % instead of 12.53 % lets the product settle 0.67 °C low over about 40 min.

The recycle is broken by the routing. In a diversion nothing flows through the pasteurized side, the raw milk reaches the heater at the balance-tank temperature, and the heater must supply $\rho c_p Q(T_{SP} - T_{BT})$ by itself, 683 kW from a 6 °C tank. The 900 kW valve was sized for that duty.

13.5 Implemented Controller

The loops and the supervisor run on every third 0.1 s physics step, a control period $\Delta t_c = 0.3\,\mathrm{s}$, and the integrals advance with that value. A priming call seeds them at $t = 0$, and they then execute at 0.3, 0.6, 0.9 s and so on. Within each step any scripted event that is due is applied first, then the controller when its period is due, then the integration, then the parcel motion and routing, the order of the engine step in Section 1.6.1. The historian logs at 2 Hz, and the speed runs from 1 to 60 s of simulated time per second of real time, 5 by default.

13.5.1 Instruments

No loop reads a parcel. Every measurement is the true value at a fixed point of the line passed through an exact first-order lag, updated at each controller execution, plus colored noise with a 0.5 s correlation time and, for the recorder and the indicating thermometer, an additive bias. The lags are 2 s for TT-101 and TT-102, 3 s for TT-103, 1.5 s for the recorder, 2 s for the indicating thermometer, 1 s for the flow meter and 0.5 s for the PDT. The noise has a standard deviation of 0.03 °C on TT-101 and TT-102, 0.05 °C on TT-103, 0.02 °C on the two tube thermometers, 0.3 % of reading on the flow meter and 0.3 kPa on the PDT. The recorder and the indicating thermometer read the same point, so only their lags, their noise and their biases differ.

13.5.2 Cascade, Feedforward and the PI Algorithm

TIC-101 holds the heater outlet by setting the hot-water setpoint, and TIC-102 holds the hot water by moving the steam valve. With feedforward on, TIC-101 adds the hot-water temperature that the clean heater needs at the measured flow Q_m and heater inlet $T_{102,m}$, as shown in Equation (13.10).

$$T_{w,ff} = \frac{T_{SP} - T_{102,m}\,\mathrm{e}^{-N_m}}{1 - \mathrm{e}^{-N_m}}, \qquad N_m = \frac{UA_H}{\rho c_p Q_m} \tag{13.10}$$

clamped to 20–130 °C. The feedforward knows nothing of fouling, the recycle or the losses, and the integral trims what it misses. Both loops are PI controllers in the velocity form of Section 1.6.2, executed every Δt_c with no derivative term and no rate limit. In Equation (1.16) the output bias u_b is $T_{w,ff}$ for TIC-101 and zero for TIC-102, and the back-calculation of Equation (1.17) uses the tracking time $T_t = \tau_I$. The limits are 60–95 °C on the hot-water setpoint and 0–100 % on the valve. Each loop is reverse acting under the ISA convention, since a product below its setpoint must raise the hot-water setpoint and a cold hot-water circuit must open the valve. The defaults are $K_c = 1.0\,^\circ\mathrm{C}$ per °C and $\tau_I = 20\,\mathrm{s}$ for TIC-101, and $K_c = 6\,\%$ per °C and $\tau_I = 30\,\mathrm{s}$ for TIC-102.

The outer loop has a second anti-windup mechanism. While the valve command sits at 100 % with the product below its setpoint, or at zero with the product above it, TIC-101 runs with $\tau_I = \infty$ for that scan, so its integral term stops while the proportional term still acts. This is conditional integration against the saturation of the inner loop, the remedy for cascade windup [17]. In manual the operator sets the valve, TIC-102 tracks the operator's valve position as in Section 1.6.2, and TIC-101 tracks the measured hot water, so the return to automatic is bumpless. In local mode TIC-102 holds the operator's hot-water setpoint and TIC-101 tracks it. Switching the feedforward shifts the integral by the old minus the new feedforward, so the hot-water setpoint does not jump. The timing pump receives the flow setpoint directly, and there is no flow loop.

13.5.3 The Supervisor

The FDV is commanded by the state machine shown in Figure 13.2, a hybrid automaton in the sense of Section 1.4.7. Forward flow requires that the recorder read at least $T_{rec} + \delta$, 72.2 °C with the default $\delta = 0.2\,^\circ\mathrm{C}$, that the measured flow not exceed Q_{legal}, that the Δp interlock be untripped, and that the two tube thermometers not disagree by more than 0.5 °C for 3 s, a persistence that keeps the difference between their lags during a fast swing from counting as a fault. Any failure diverts at once. When all four conditions hold again the supervisor keeps diverting in recovery until $n_{flush} V_T$, 90 L at

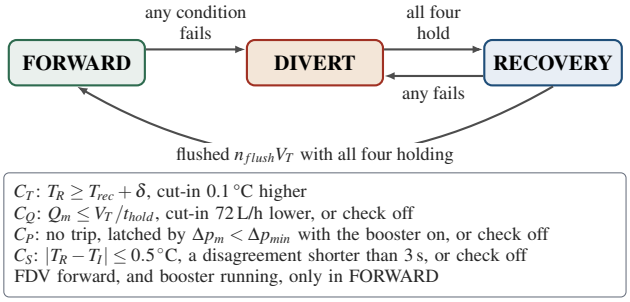


Figure 13.2 Supervisor that routes the product. The four conditions are evaluated on the instruments every 0.3 s, and only the transition from forward to divert counts a diversion. A Δp trip also stops the booster and stays latched until an operator reset, and the flush is counted with the measured flow.

the default of two tube volumes, has passed, so parcels that were in the tube during the fault never go forward on a later good reading. The Δp trip, set below 6.9 kPa while the booster runs, stops the booster, which drops the raw side to suction, and clears only on an operator reset, a latch in the sense of Section 1.4.7.

13.5.4 Simulation, Control and Internal Models

The parcel model of Section 13.3 is the simulation model. The controller's internal model is Equation (13.10), the clean heater alone, with no regenerator, no fouling and no losses. The supervisor's only model is the legal flow V_T/t_{hold}. No model is identified inside the application, but step tests on it give the control model the tuning faces. Through the inner loop in local mode, a 1 °C step of the hot-water setpoint, taken as the difference from an identical run without the step, leaves the heater inlet within 0.001 °C of its old value for 24.4 s, and the product settles 0.950 °C higher, 0.749 °C of it from the heater and the rest returned by the regenerator. The hot water overshoots to 1.32 °C above its old value at 20 s on the way.

13.6 Scenarios and Measured Results

The eight cases and the challenge are listed in Table 13.3. Every case starts from the reference configuration, feedforward and outer-loop anti-windup on, a 73.5 °C product setpoint, two flush volumes and every check on, apart from the changes named. Each change a case names is made after the case is loaded and before the run starts. Reset keeps the reader's settings (Section 1.6.1), and the app gives one run whether a setting is chosen before the run or carried through Reset. All values below were read from headless runs of the engine at the default seed, with times in seconds of simulated time. Values that move with the measurement noise are averages over the stated window, and events are timed to the 0.1 s step.

13.6.1 Stable Production and a Colder Feed

In the first case the regenerator warms the milk from 6 to 62.9 °C, and the heater, with hot water at 77.0 °C and the valve at 12.5 %, adds the last 10.6 °C. Each parcel spends 18.0 s in the tube, 3 s more than the recipe, and reaches the recorder 0.30 °C cooler, at 73.20 °C. A parcel that makes the whole trip carries 30.0 °C s of heat above the recipe, and the scorecard's mean over the parcels delivered by 300 s is 29.9 °C s, since the parcels already in the line at $t = 0$ start with the heat records of the steady profile.

When the feed switches from 10 to 2 °C at 60 s, the balance tank passes 6 °C after 39 s, and the heater inlet falls only from 63.55 °C to 62.31 °C by 480 s, a drop of 1.24 °C. The heater attenuates that again, and the heater outlet departs from its setpoint by at most 0.08 °C with the cascade and feedforward and 0.10 °C without the feedforward. The heater duty rises from 101 to 113 kW, and the

Table 13.3 Scenario Sequence

#	Scenario and event	Teaching point
1	Stable production	where the heat comes from
2	Feed 10 to 2 °C at 60 s	the regenerator filters
3	10 500 L/h, then 11 500 L/h	a flow limit is a time limit
4	Steam to 60, then 15 %	inner-loop headroom
5	Fouling at 3 %/min	a controller hides decline
6	Recorder, then both, +1.5 °C	what redundancy cannot see
7	BPV spring creeps	the interlock and its margin
8	40 s steam trip, no anti-windup	the return is the upset
9	1800 s campaign	graded challenge

Table 13.4 Outcomes of the Cases at the Default Seed

Case	Measured outcome
1, mean 60–300 s	regenerator 576, heater 107, steam 112.8 kW
1, per parcel	hold 18.0 s, tube loss 0.30 °C, 30.0 °C s
2, heater inlet	63.55 °C to 62.31 °C by 480 s
2, product departure	cascade 0.08, no FF 0.10, local 0.25 °C
3, 10 500 L/h	hold 15.4 s, dip 0.66 °C, no diversion
3, 11 500 L/h	divert 241.6 s, low 64.9 °C, forward 447.4 s
3, same	576 L diverted, heater peak 682 kW
3, flow check off	134 parcels (268 L) held 14.1 s
4, 60 % steam	valve 12.5 to 20.9 %, dip 0.66 °C
4, 15 % steam	divert 268.6 s, low 52.5 °C, 284 L
4, no outer anti-windup	456 L diverted
5, 1200 s	UA_H 62.5 %, T_w 77.0 to 81.1 °C
5, product	within 0.13 °C, integral 4.0 °C
6, recorder drift	divert 62.8 s, 382 L, forward 216.1 s
6, common drift	128 parcels (256 L) by 540 s, undetected
6, check off, one drift	27 parcels by 180 s
7, BPV creep	trip 103.9 s at 6.61 kPa, 506 L
7, interlock off	reversal 132.2 s, 151 parcels (302 L)
8, no anti-windup	3 diversions, 506 L, 45.6 °C s
8, anti-windup on	1 diversion, 236 L, 33.5 °C s

product leaving the cooler falls from 3.0 to 2.4 °C. With the cascade opened (Local) the hot water holds its 77.0 °C default rather than the 76.8 °C the cascade was asking for, so the product has risen to 73.60 °C when the feed switches, on its way to the 73.66 °C that the same hot water holds without the switch. It then settles at 73.27 °C, the 0.39 °C below that level that the recycle gain of Equation (13.9) predicts for the 8 °C change, and bottoms 0.25 °C below its setpoint at 282 s. The colder feed is a small upset by physics, because it enters before the regenerator. The case serves to locate where a disturbance enters rather than to test the loop, the question Volume 3, Chapter 2 asks of every loop structure.

13.6.2 Flow and the Holding Time

A flow step to 10 500 L/h at 60 s shortens the hold from 18.0 to 15.4 s and dips the product 0.66 °C without a diversion. At 240 s production asks for 11 500 L/h, where the hold would be 14.1 s however hot the product is, and the flow condition diverts at 241.6 s. The diversion is expensive. With nothing flowing through the pasteurized side, the heater inlet falls to 12.0 °C, the heater duty rises to 682 kW and the heater outlet sinks to 64.9 °C. After production returns to 9000 L/h at 330 s, the supervisor enters recovery at 342.7 s, falls back to divert twice as the loops restore the temperature, and returns to forward flow at 447.4 s, with 576 L diverted. With the flow check off, no diversion occurs and 134 parcels, 268 L, go forward by 330 s at the right temperature and held only 14.1 s.

🕮 **Pause.** Try Exercise 13.1 before reading on.

13.6.3 Steam Loss and Fouling

A sag of the steam supply to 60 % at 60 s is ridden out by the inner loop. The valve opens from 12.5 to 20.9 %, passing 20.5 % at 74.1 s, the product dips 0.66 °C, and the recorder never reads below 72.56 °C. At 240 s the supply falls to 15 %, where the valve needs about 80 % to give the heater its 107 kW. The hot water cools while the valve travels, and the FDV diverts at 268.6 s, before the valve reaches 80 % at 272.1 s. The diversion removes the regenerator's hot side, and by 300 s the heater is taking 385 kW against the 135 kW a wide-open valve can give, so the heater outlet has sunk to 53.9 °C when the supply returns at 300 s and reaches 52.5 °C at 302.4 s. Forward flow resumes at 382.0 s after 284 L. With the outer-loop anti-windup off the same event costs 456 L and forward flow resumes at 451.3 s.

Fouling at an accelerated 3 % per minute leaves the heater with less than 70 % of its clean UA_H from 857 s and 62.5 % after 1200 s. The cascade raises the hot water from 77.0 to 81.1 °C to push the same heat through the poorer surface, the product stays within 0.13 °C of its setpoint, and the valve stays between 11.9 and 13.5 % because the duty has not changed. The feedforward still assumes a clean heater and stays at 77.0 °C, and the integral of TIC-101 carries 4.0 °C. The approach, hot water minus product, grows from 3.5 °C in the first 30 s to 7.6 °C in the last 30 s, and it is the only trend that shows the deposit. A fouling rate of this size would not occur in a dairy, but it serves to compress a production run into twenty minutes. Volume 2, Chapter 4 follows fouling on a steam-heated exchanger.

13.6.4 What the Instruments Cannot See

When the recorder alone reads 1.5 °C high at 60 s, the plausibility check diverts at 62.8 s with the product at 73.17 °C. A comparison of two instruments can say that one of them is wrong but never which one, so good product is lost, 382 L by the time the recorder is recalibrated at 180 s and forward flow resumes at 216.1 s. At 360 s both thermometers read 1.5 °C high together, they agree, and the check has nothing to compare. At 390 s the operator trims the setpoint to 72.0 °C to bring the recorder back to its usual reading. The product at the recorder falls below 72 °C at 419.5 s, and the first underprocessed parcel goes forward at 421.3 s. At 500 s, as shown for the recorder and the product in Figure 13.3, the recorder reads 73.17 °C and the indicating thermometer 73.10 °C while the product is at 71.76 °C, and by 540 s 128 parcels, 256 L, have gone forward underprocessed, every one while all four conditions held. With the plausibility check off and the same trim made at 120 s, the single drifting recorder passes 27 parcels by 180 s. Redundancy catches independent faults and not common ones.

> **TRY IT IN THE APP 6 · Sensor Bias**
>
> Load the scenario and press Run. Watch the FDV and the tube strip of the trends. When the recorder alone reads high at 60 s the plant diverts. After 360 s both thermometers read high together, the FDV stays forward, and the product trace falls below the recipe line while both thermometer traces stay above it. By 540 s 128 parcels have gone forward underprocessed.

Pause. Try Exercise 13.2 before reading on.

13.6.5 The Pressure Differential

The differential starts at 25.0 kPa. At 60 s the back-pressure valve spring begins to creep toward 45 % below its setting, and the interlock trips at 103.9 s, when the transmitter reads 6.56 kPa and the true differential is still 6.61 kPa. The booster stops, the raw side drops to suction and the FDV diverts. The trip latches through the repair at 240 s until the operator reset at 270 s, and forward flow resumes at 306.1 s after 506 L. With the interlock off the differential reverses at 132.2 s, and 151 parcels, 302 L, sit in the regenerator with the raw side at the higher pressure and go on to the product tank, 150 of

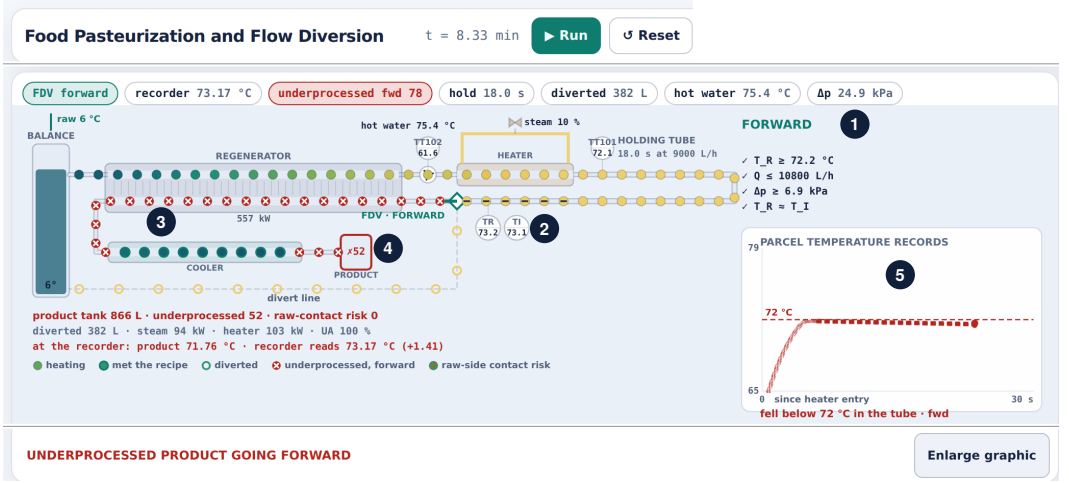


Figure 13.3 The pasteurizer application in case 6, sensor bias, at $t = 8.33$ min (500 s), 140 s after both tube thermometers began reading 1.5 °C high and 110 s after the setpoint was trimmed to 72.0 °C. The picture is a composite of the app's header, plant view and status strip, and the control panel, the trends and the Learn panel are cut. (1) The FDV is forward and the supervisor's four conditions are all ticked, while the tally under the scene gives the product at the recorder as 71.76 °C against a recorder reading of 73.17 °C at (2) the recorder and the indicating thermometer. The six parcels in the last 11 L of the holding tube carry the bar of a record already broken, (3) crossed red parcels fill the pasteurized side of the regenerator and surround nine treated parcels in the cooler, (4) the product tank counts 52 of the 78 underprocessed parcels forwarded so far, and (5) the parcel records run just below the 72 °C line. The status strip reads underprocessed product going forward, and the tally shows 382 L diverted, 866 L delivered and no raw-contact risk.

them by the repair and the last 0.1 s after it, while the differential recovers. The margin between the 6.9 kPa trip and zero is what turns a slow mechanical failure into a safe diversion.

13.6.6 Coming Back From a Diversion

Case 8 trips the steam for 40 s at 30 s with the outer-loop anti-windup off, and the FDV diverts at 56.5 s. When the steam returns at 70 s the valve is wide open, the hot water overshoots to 98.3 °C at 105.4 s, and the heater outlet peaks at 77.7 °C at 110.4 s, while the plant is still diverting. Recovery begins at 115.0 s and forward flow resumes at 151.3 s. Then the regenerator restarts with a balance tank at 56.1 °C, the heater outlet rises to 75.96 °C at 168 s, and the valve stays below 0.5 % from 168.6 to 214.5 s, because the hot water can cool only as fast as the milk carries heat out of it. Without anti-windup the integral of TIC-101 keeps pushing the hot-water setpoint down, to 70.3 °C, and when the tank refills with cold milk the product falls below the cut-out, with second and third diversions at 230.5 and 398.8 s and 506 L diverted by 420 s. With the anti-windup on, the setpoint bottoms at 71.5 °C and one diversion of 236 L is the whole cost, while the heat above the recipe per delivered parcel falls from 45.6 to 33.5 °C s. The largest overshoot of the case therefore belongs to the diversion, and the upset that causes the later diversions is the smaller one after the return.

Table 13.5 Supervisor and Plant Margins. Runs Last 600 s for Case 3, 420 s for Case 8, 1800 s for the Challenge, and 900 s With the Smaller Valve

Configuration	Diversions	Diverted	Under
case 3, flush 0 volumes	8	478 L	0
case 3, flush 1 volume	8	786 L	1
case 3, flush 2 volumes (default)	1	576 L	0
case 3, flush 3 volumes	1	622 L	0
case 8, cut-out margin 0 °C	3	494 L	1
challenge as shipped, margin 0 °C	4	1524 L	6
case 3, 500 kW valve	4	1058 L	0
case 4, 500 kW valve	3	932 L	0

TRY IT IN THE APP 8 · Recovery From Diversion

Load the scenario and press Run. With the outer-loop anti-windup off the steam trip costs 3 diversions and 506 L by 420 s. Press Reset, set Outer-loop anti-windup to On on the TIC-101 card before pressing Run, and run again: one diversion of 236 L is the whole cost.

Pause. Try Exercise 13.3 before reading on.

13.6.7 The Supervisor's Margins

What the defaults of the supervisor and the plant protect against is measured in Table 13.5. The counts of underprocessed parcels depend on the measurement noise, so every configuration was also run at seven other seeds, and the spread is given below. With no flush, case 3 returns to forward flow the moment the recorder recovers and cycles through 8 diversions, which forwards no underprocessed parcel at the default seed but up to 2 at others. One tube volume still cycles 8 times, diverts 786 L and forwards one underprocessed parcel (none at the other seven seeds). Two volumes return in one diversion of 576 L with none forwarded at any seed, and three add 46 L without improving safety. Without the 0.2 °C cut-out margin, the recorder's 1.5 s lag lets 1 underprocessed parcel through in case 8 (1 to 4 over the seeds) and 6 in the shipped challenge (3 to 7), while a 0.5 s recorder lets none through in either (at most one at other seeds). With a 500 kW valve, set in the parameter panel before the run, the heater cannot replace the regenerator during a diversion, and case 3 is still diverting at 900 s after 4 diversions and 1058 L. At four of the eight seeds it regains forward flow near 572 s after 2 diversions, 125 s later than with the full valve. Moving the recorder to 0.5 L before the valve in a copy of the model, with everything else unchanged, forwards 3 underprocessed parcels in case 3 with a one-volume flush (2 to 4 over the seeds), where the 3 L position forwards 1 (0 or 1). Neither position forwards one in the shipped challenge at the default seed, and the 0.5 L position forwards one at a single other seed.

13.6.8 The Campaign Challenge

The challenge runs 1800 s with two silo switches, four production-rate changes and two sags of the steam supply, starting from a detuned cascade (K_c 0.5 and 60 s outside, 3 and 60 s inside) with no feedforward, no outer anti-windup and a 73.0 °C setpoint. It is met with no underprocessed or raw-contact-risk parcel forwarded, at most 100 L diverted, and at most 40 °C s of heat above the recipe per delivered parcel. The measured configurations are listed in Table 13.6. As shipped the plant stays safe but diverts 4 times, 1554 L in all, and delivers 3220 L for 83.8 kWh, 26.0 kWh per 1000 L. The suggested configuration never diverts, keeps the heater outlet between 72.84 and 74.90 °C and the tube outlet above 72.59 °C, and delivers 4774 L for 63.9 kWh, 13.4 kWh per 1000 L, with 34.9 °C s per parcel, 5.1 °C s inside the limit. Each element pays for itself. A higher setpoint alone avoids every diversion but overheats the product, the suggested tuning at the old setpoint diverts 7 times (6

Table 13.6 Challenge Configurations Over 1800 s. Heat Is the Mean Heat Above 72 °C per Delivered Parcel

Configuration	Div.	Diverted	Heat	Delivered
as shipped	4	1554 L	74.2	3220 L
suggested[a]	0	0	34.9	4774 L
setpoint 74.5 °C only	0	0	49.9	4774 L
suggested at 73.0 °C	7	2084 L	25.6	2688 L
suggested, no feedforward	1	304 L	39.7	4470 L
suggested, no anti-windup	0	0	34.8	4774 L

[a] feedforward and anti-windup on, 1.0 and 20 s, 6 and 30 s, 73.8 °C

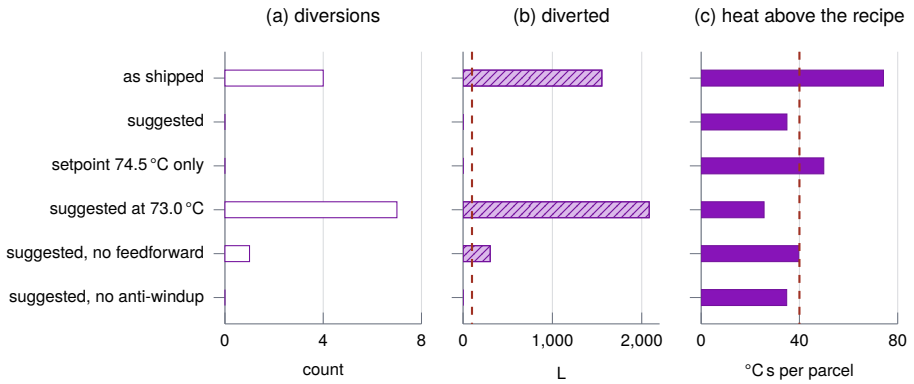


Figure 13.4 The campaign challenge of Table 13.6 in its six configurations: (a) diversions, (b) volume diverted and (c) heat above the recipe per delivered parcel, with the challenge limits dashed. The table prints the values. Notice that no tuning buys the margin that the setpoint gives, and that the margin is paid for in heat: the higher setpoint alone never diverts but overheats, and the suggested tuning at the old setpoint diverts most of all.

or 7 over eight seeds), and dropping only the feedforward costs one diversion of 304 L. The outer anti-windup changes little in the suggested run, because the valve command sits at zero only from 906.4 to 914.8 s, after the cut to 8000 L/h, and the frozen integral moves the mean heat by 0.1 °C s. Figure 13.4 draws the diversions, the volume diverted and the heat of each configuration against the limits.

> ✦ **DID YOU KNOW? Better Tuning, More Diversions**
>
> From the simulator: the suggested tuning, run at the shipped 73.0 °C setpoint, diverts 7 times and sends 2084 L back, where the detuned plant as shipped diverts 4 times and 1554 L. The setpoint, not the tuning, buys the margin (Table 13.6).

13.7 Technology in Practice

What Measures It

In a real HTST plant the sensor of the thermal-limit controller sits at the outlet of the holding tube. It is set and sealed so that forward flow cannot start unless the temperature there is above the pasteurization temperature, nor continue once it falls below, and no milk may bypass it [7]. That is the cut-out of Figure 13.2, read by the chapter's TR. The time half of the recipe is held by a positive-displacement timing pump or by a timing system built on a magnetic flow meter [7], the TP and FT of Figure 13.1.

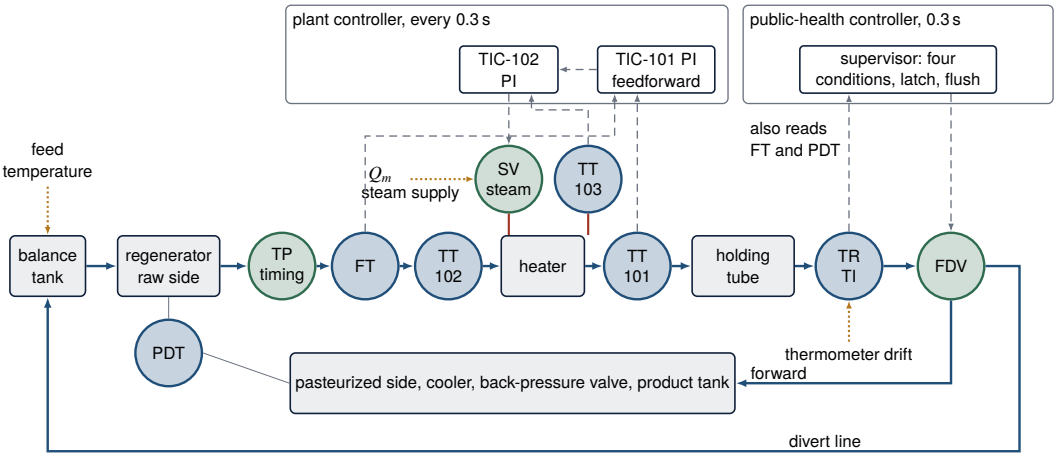


Figure 13.5 The loops of the pasteurizer drawn on plant hardware, with the chapter's tags. Sensors are drawn blue, final elements green and disturbances dotted. The temperature cascade (TIC-101 on TT-101 sets the setpoint of TIC-102 on TT-103, which moves the steam valve) runs on the plant controller, and the supervisor that commands the FDV runs on its own controller, as the ordinance requires of a public-health computer. Notice that the cascade never reads TR and TI, and the supervisor never moves the steam.

> **TECHNOLOGY** **Sensor: A Recorder-Controller and Its Witness**
>
> The ordinance asks for two thermometers at the end of the holding tube. The recorder-controller must read within 0.5 °C near the temperature at which it diverts, and a digital sensor may drift no more than 0.5 °C in three months against a certified source. Its self-diagnostic circuits blank the reading on a failed component. The indicating thermometer must read within 0.2 °C, and in a bath test its reading must climb from 11 °C to 4 °C below the bath in four seconds or less [7]. TR and TI of Section 13.5 model the pair as lags of 1.5 and 2 s with a bias each, and a common bias is the fault the pair cannot see.

What Moves It

Three final elements act on the milk: the steam valve of the hot-water circuit, the timing pump and the flow diversion device. The booster pump on the raw side is a fourth, switched rather than modulated. The ordinance lets it run only while the timing pump runs, the diversion device is forward and the pasteurized milk is at least 6.9 kPa above the most the booster can develop [7]. The chapter's interlock trips at the same 6.9 kPa (Table 13.2) and stops the booster. Figure 13.5 draws the loops on this hardware.

> **TECHNOLOGY** **Final Element: The Flow Diversion Device**
>
> The flow diversion device sits downstream of the holding tube and diverts the milk automatically in response to a sub-legal condition. A failure of its motive power must divert the flow, so the valve fails safe. A leak-escape on the forward side of the seat carries any leakage outside or back to the balance tank, and the flow-control sensor sits at most 46 cm upstream of the valve inlet. Motive pumps must stop when the milk is below the pasteurization temperature and the valve is not fully diverted [7]. The chapter's FDV is a two-position valve with a 0.5 s travel.

Where the Controller Runs

Computers commonly manage the public-health devices of a pasteurizer, and the ordinance treats such a computer as a replacement for hard wiring. A computer or programmable logic controller used for public-health control is dedicated to one pasteurizer, cannot be commanded by any other

computer, puts every public-health output in its fail-safe position on loss of power, and keeps its program sealed. It must monitor all inputs and update all outputs at least once every second [7]. The chapter's supervisor scans every 0.3 s, inside that limit. Read as plant hardware, the cascade of TIC-101 and TIC-102 belongs in the plant controller, which may only request a device, and the four conditions, the latch and the flush belong in the public-health controller.

> **IN PRACTICE** **Cut-In, Cut-Out and the Salt Test**
>
> At the start of each day the operator finds the cut-in and cut-out temperatures on the indicating thermometer and enters them on the recorder chart, and checks the recorder against the indicating thermometer, adjusting it to read no higher. The holding time is the flow time of the fastest particle. It is tested at maximum flow by timing a salt solution injected into water through the tube, on installation, at least every six months, after any change that may affect it, and whenever the seal on the timing pump's speed is broken [7].

What Is Modernizing It

Three developments touch this plant. Many pasteurization systems time the hold with a magnetic flow meter. Its flow alarm diverts whenever the flow would make the hold shorter than the legal time, and once the legal flow returns a delay keeps the device diverted for at least 15 or 25 seconds, depending on the product [7]. The chapter's supervisor has both, as its flow condition and its flush. Electronic collection and storage of data can replace the circular chart recorder, provided the records keep the purpose of their paper counterparts [7]. In-line fouling sensors and predictive models aim to time cleaning-in-place from the state of the surface [18], the trend that the approach of case 5 stands in for.

13.8 Limitations

Plug flow gives every parcel the mean residence. A real holding tube is judged on its fastest particle, which the model does not represent [6].

The recipe is a threshold with no microbial kinetics, so the model cannot credit extra lethality at higher temperature or grade a short excursion by its effect. The overprocessing measure, heat above the recipe in °C s, stands in for quality kinetics.

The fouling rate is prescribed rather than tied to the surface temperature [15], and no leak is simulated behind the raw-contact flag. The instruments drift only through the bias sliders and the scenario events.

> **OPEN CHALLENGE** **Seeing the Deposit Before It Matters**
>
> Fouling still cuts the efficiency of dairy heat treatment, raises its energy use and complicates its cleaning, and in severe cases a plant can fail to meet its required thermal profile. Cleaning is usually decided from a fall in heat transfer or a rise in pressure drop, which sees fouling late, and cleaning systems cannot tell when the deposit is gone, so they nearly always overclean. Acoustic, ultrasonic, electrochemical and spectroscopic sensors and learned predictors exist, but most remain at the experimental or prototype stage [18]. Progress would be in-line sensing on real exchangers that sets the cleaning schedule. Case 5 shows the gap: only the approach reveals the deposit.

13.9 Conclusion

A regenerative HTST pasteurizer is a plant in which the controlled variable is not the thing that must be guaranteed, and the simulator described here keeps the two apart by carrying the milk as parcels with their own records. The regenerator supplies 84 % of the heat, and the recycle it creates raises the steady gain of the hot water to 5.57 °C per percent of valve with a 429 s response, against 1.17 °C per percent and 52 s without it, while a diversion removes it and asks the heater for up to 682 kW. A flow of 11 500 L/h holds each parcel 14.1 s, and without the flow check 134 parcels go forward at the right temperature. Two thermometers drifting together pass 128 underprocessed parcels while every

condition holds. Conditional integration on the outer loop cuts the return from a steam trip from 3 diversions and 506 L to 1 and 236 L, and the challenge moves from 4 diversions and 1554 L to none. The model assumes plug flow, a threshold recipe and prescribed fouling. A fuller model would add a residence time distribution with a fastest particle, microbial kinetics for the recipe, a fouling law driven by the wall temperature and a simulated leak. The application is an accessible browser plant for learning where a temperature loop ends and a certified treatment begins.

KEY POINTS

- A pasteurizer must guarantee the time and temperature each parcel of milk received, which no controller measures.
- The regenerator supplies 84 % of the heat, and its recycle makes the plant almost five times as sensitive to steam and about eight times slower.
- The holding time is the tube volume divided by the flow, so at 11 500 L/h every parcel is held 14.1 s however hot it is.
- Two thermometers drifting together pass 128 underprocessed parcels while every condition of the supervisor holds.
- Conditional integration on the outer loop cuts the return from a steam trip from 3 diversions and 506 L to 1 and 236 L.

Where else this applies. The cut-out margin above the recipe, sized to the recorder's lag (Table 13.5), is a guard band of the same kind as the one that narrows the fill limits before a vial is accepted in Chapter 9.

References

[1] A. A. Teixeira and G. S. Tucker, "On-line retort control in thermal sterilization of canned foods," *Food Control*, vol. 8, no. 1, pp. 13–20, 1997.

[2] A. Negiz, P. Ramanauskas, A. Çinar, J. E. Schlesser, and D. J. Armstrong, "Modeling, monitoring and control strategies for high temperature short time pasteurization systems—2. Lethality-based control," *Food Control*, vol. 9, no. 1, pp. 17–28, 1998.

[3] J. E. Schlesser, D. J. Armstrong, A. Cinar, P. Ramanauskas, and A. Negiz, "Automated control and monitoring of thermal processing using high temperature, short time pasteurization," *Journal of Dairy Science*, vol. 80, no. 10, pp. 2291–2296, 1997.

[4] J. B. Enright, W. W. Sadler, and R. C. Thomas, "Pasteurization of milk containing the organism of Q fever," *American Journal of Public Health and the Nations Health*, vol. 47, no. 6, pp. 695–700, 1957.

[5] O. Cerf and R. Condron, "*Coxiella burnetii* and milk pasteurization: an early application of the precautionary principle?" *Epidemiology and Infection*, vol. 134, no. 5, pp. 946–951, 2006.

[6] P. M. Tomasula and M. F. Kozempel, "Flow characteristics of a pilot-scale high temperature, short time pasteurizer," *Journal of Dairy Science*, vol. 87, no. 9, pp. 2761–2768, 2004.

[7] *Grade "A" Pasteurized Milk Ordinance (PMO), 2025 Revision*, U.S. Food and Drug Administration and National Conference on Interstate Milk Shipments, 2025. [Online]. Available: https://www.fda.gov/media/193438/download

[8] R. W. Dickerson, A. M. Scalzo, R. B. Read, and R. W. Parker, "Residence time of milk products in holding tubes of high-temperature short-time pasteurizers," *Journal of Dairy Science*, vol. 51, no. 11, pp. 1731–1736, 1968.

[9] C. A. Ryan, M. K. Nickels, N. T. Hargrett-Bean, M. E. Potter, T. Endo, L. Mayer, C. W. Langkop, C. Gibson, R. C. McDonald, R. T. Kenney *et al.*, "Massive outbreak of antimicrobial-resistant salmonellosis traced to pasteurized milk," *JAMA: The Journal of the American Medical Association*, vol. 258, no. 22, pp. 3269–3274, 1987.

[10] J. J. Ibarrola, J. M. Sandoval, M. García-Sanz, and M. Pinzolas, "Predictive control of a high temperature–short time pasteurisation process," *Control Engineering Practice*, vol. 10, no. 7, pp. 713–725, 2002.

[11] S. Niamsuwan, P. Kittisupakorn, and I. M. Mujtaba, "Control of milk pasteurization process using model predictive approach," *Computers & Chemical Engineering*, vol. 66, pp. 2–11, 2014.

[12] V. H. Holsinger, K. T. Rajkowski, and J. R. Stabel, "Milk pasteurisation and safety: a brief history and update," *Revue Scientifique et Technique de l'OIE*, vol. 16, no. 2, pp. 441–451, 1997.

[13] D. M. Roger, "A method for checking the holding time in short time high temperature pasteurizers," *Journal of Milk Technology*, vol. 2, no. 4, pp. 191–192, 1939.

[14] M. C. Georgiadis and S. Macchietto, "Dynamic modelling and simulation of plate heat exchangers under milk fouling," *Chemical Engineering Science*, vol. 55, no. 9, pp. 1605–1619, 2000.

[15] B. Bansal and X. D. Chen, "A critical review of milk fouling in heat exchangers," *Comprehensive Reviews in Food Science and Food Safety*, vol. 5, no. 2, pp. 27–33, 2006.

[16] R. K. Shah and D. P. Sekulić, *Fundamentals of Heat Exchanger Design*. Wiley, 2003.

[17] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[18] J. Tarapata, T. R. Murphy, E. W. Finnegan, T. F. O'Callaghan, and J. A. O'Mahony, "Approaches for measuring and predicting fouling during thermal processing of dairy solutions," *Comprehensive Reviews in Food Science and Food Safety*, vol. 24, no. 4, p. e70209, 2025.

13.10 Exercises

13.1 Concept check · 2 min

In Scenario 3 production asks for 11 500 L/h at 240 s. With the flow check off, the cascade keeps the product near its setpoint and the recorder never reads below the recipe temperature. Is the milk that goes forward pasteurized?

(a) Yes, because every parcel stays at or above 72 °C all the way through the holding tube

(b) Yes, provided the feedforward raises the hot water to match the higher flow

(c) No, because the tube now holds each parcel 14.1 s, under the 15 s of the recipe

(d) No, because the regenerator recovers less heat at the higher flow and the product cools

13.2 Concept check · 2 min

In Scenario 6 both tube thermometers start reading 1.5 °C high at 360 s, and the operator trims the setpoint to 72.0 °C at 390 s. Underprocessed parcels go forward from 421.3 s while all four conditions of the supervisor hold. Why does the plausibility check not divert?

(a) Its 3 s persistence filters out a drift as gradual as this one

(b) The trim of the setpoint switches the check off until the next recalibration

(c) The recorder's lag hides the drift until the product has left the tube

(d) The two thermometers agree, so a comparison between them has nothing to detect

13.3 Concept check · 1 min

In Scenario 8 the outer-loop anti-windup is off. After forward flow resumes at 151.3 s the steam valve stays shut while the product runs hot, and two more diversions follow by 420 s. What does the outer-loop anti-windup change?

(a) It freezes the product loop's integral while the valve is pinned, so the hot-water setpoint is not driven down

(b) It limits how fast the steam valve may open when the steam returns at 70 s

(c) It keeps the FDV in divert until the balance tank has cooled to the feed temperature

(d) It narrows the hot-water setpoint range so that the hot water cannot overshoot

13.4 Theory and calculation · 4 min

The holding tube holds $V_T = 45\,\text{L}$, the recipe asks for 15 s at 72.0 °C, and the parcels move in plug flow.

(a) Find the time in the tube at 9000 L/h, 10 500 L/h and 11 500 L/h. *Checkpoint:* 18.0 s, 15.4 s and 14.1 s.

(b) Find the legal flow Q_{legal}, and the flow below which the flow condition holds again after a trip, given its cut-in 72 L/h lower. *Checkpoint:* 10 800 L/h and 10 728 L/h.

(c) With all four conditions holding, how long does the default flush of two tube volumes take at 9000 L/h? Compare with Scenario 6, where the recorder is recalibrated at 180 s and forward flow returns at 216.1 s. *Checkpoint:* 36 s, as measured.

13.5 Theory and calculation · 7 min

The regenerator has $UA_R = 57.4\,\mathrm{kW/K}$ and the clean heater $UA_H = 14.0\,\mathrm{kW/K}$. Milk has $\rho = 1030\,\mathrm{kg/m^3}$ and $c_p = 3.93\,\mathrm{kJ/(kg\,K)}$, the flow is 9000 L/h, and the tube loss has $\tau_{loss} = 3000\,\mathrm{s}$ (Equations (13.4) and (13.9)).

(a) Find $\rho c_p Q$, NTU_R, ε, NTU_H and e^{-NTU_H}. *Checkpoint:* 10.12 kW/K, 5.67, 0.850, 1.38 and 0.251.

(b) With the parcel effectiveness 0.847, find a_T, the recycle loop gain g and the steady gains of T_{101} to the hot water and to the feed. Derive the second from a balance on the heater and the regenerator. *Checkpoint:* 0.994, 0.211, 0.950 and 0.0486.

(c) In Scenario 2 with the cascade opened (Local), the hot water holds 77.0 °C while the feed falls 8 °C. Predict the steady fall of the product. *Checkpoint:* 0.39 °C, as measured (73.66 to 73.27 °C).

13.6 Theory and calculation · 7 min

Each percent of the 900 kW steam valve adds 9 kW at design pressure. The hot-water circuit has $C_w = 400\,\mathrm{kJ/K}$ and loses $UA_{wl} = 100\,\mathrm{W/K}$, and the heater passes $1 - \mathrm{e}^{-NTU_H} = 0.749$ of the difference between the hot water and its inlet to the milk (Section 13.4).

(a) Ignore the recycle, so that the heater inlet stays fixed. Find the heat the circuit loses per kelvin of hot water, the steady rise per percent of valve and the time constant of the hot water. *Checkpoint:* 7.68 kW/K, 1.17 °C per percent, 52 s.

(b) With the recycle, a kelvin of hot water raises T_{101} by 0.950 and the heater inlet by εa_T times that, with $\varepsilon = 0.847$ and $a_T = 0.994$. Find the net heat lost per kelvin and the steady rise per percent, and compare with the measured 5.57 °C. *Checkpoint:* 1.62 kW/K and 5.6 °C per percent.

(c) In a diversion the heater must bring 6 °C milk to 73.5 °C by itself. Find the duty, and compare it with what the wide-open valve passes at full steam pressure and at 15 % of it. *Checkpoint:* 683 kW against 900 kW and 135 kW.

13.7 App challenge · 7 min

Open Scenario 3, "3 · Flow increase", at the default seed, and run at 30× (30 s of plant time per wall second).

(a) Run it as loaded to 600 s (20 s of wall time). Read the Diversions and Diverted tiles. *Checkpoint:* 1 diversion, 576 L, both fixed once forward flow returns at 447.4 s.

(b) Press Reset, set Divert above the legal flow to Off on the Timing pump card, and run to 400 s. Read Underprocessed forwarded and Diverted. *Checkpoint:* 142 parcels, 0 L. The count is fixed from 336 s.

(c) The chapter counts 134 parcels by 330 s. Explain where the other parcels come from. *Checkpoint:* 8 parcels in the tube when production returned to 9000 L/h.

13.8 App challenge · 7 min

Open Scenario 8, "8 · Recovery from diversion", at the default seed, and run at 10×.

(a) Run it as loaded, with Outer-loop anti-windup Off, and pause between 490 and 560 s. Read the Diversions and Diverted tiles, and the lowest hot-water setpoint on the product trend after forward flow resumes at 151.3 s. Tolerance 0.2 °C. *Checkpoint:* 3 diversions, 670 L, lowest setpoint 70.3 °C.

(b) Press Reset, set Outer-loop anti-windup to On on the TIC-101 card, and run to the same time. Read the same values. *Checkpoint:* 1 diversion, 236 L, lowest setpoint 71.5 °C.

(c) Run the first case on past 560 s. What happens, and why does the chapter quote 506 L by 420 s? *Checkpoint:* A fourth diversion at 568 s. The third diversion is still under way at 420 s.

13.9 App challenge · 8 min

Open the challenge, "Challenge · variable-feed campaign", at the default seed, and run at 60× (30 s of wall time). Pause at 1800 s. The counts and volumes do not change between 1740 and 1860 s. Read the heat at 1800 s, to within 1 °C s.

(a) Run it as loaded. Read Underprocessed forwarded, Diversions, Diverted and Excess heat. *Checkpoint:* 0 parcels, 4 diversions, 1554 L, 74.2 °C s. Target not met.

(b) Press Reveal suggested tuning, then Reset, and run again. Read the same tiles. *Checkpoint:* 0 parcels, 0 diversions, 0 L, 34.9 °C s. Target met.

(c) Set Product temperature setpoint to 73.0 °C, press Reset and run again. Read the same tiles and explain the result. *Checkpoint:* 0 parcels, 7 diversions, 2084 L, 25.6 °C s.

13.10 Application area · 4 min

Tracer tests on a pilot HTST pasteurizer found the fastest particle leaving a nominal 21.5 s holding section after 18.6 s (Section 13.2). Assume the tube of this chapter has the same ratio of fastest to nominal time.

(a) Find the hold of the fastest particle at 9000 L/h and at the legal flow of 10 800 L/h. *Checkpoint:* 15.6 s and 13.0 s.

(b) Find the largest flow at which the fastest particle is still held 15 s, or the tube volume that would allow 10 800 L/h. *Checkpoint:* About 9340 L/h, or about 52 L.

13.11 Application area · 5 min

In Scenario 5 fouling grows at 3 % of the clean resistance per minute, so R_f rises by $r_f/6000$ each second and the heater keeps $1/(1+R_f)$ of its clean UA_H. The heater inlet stays near 62.9 °C because the duty does not change.

(a) Find when the heater falls below 70 % of clean, and what is left at 1200 s. *Checkpoint:* 857 s and 62.5 %.

(b) From the outlet relation with $NTU_H = 1.38/(1+R_f)$, find the hot water that holds the product at 73.5 °C when clean and at 1200 s, and compare with the run. *Checkpoint:* 77.0 °C and 81.2 °C, against 77.0 and 81.1 °C.

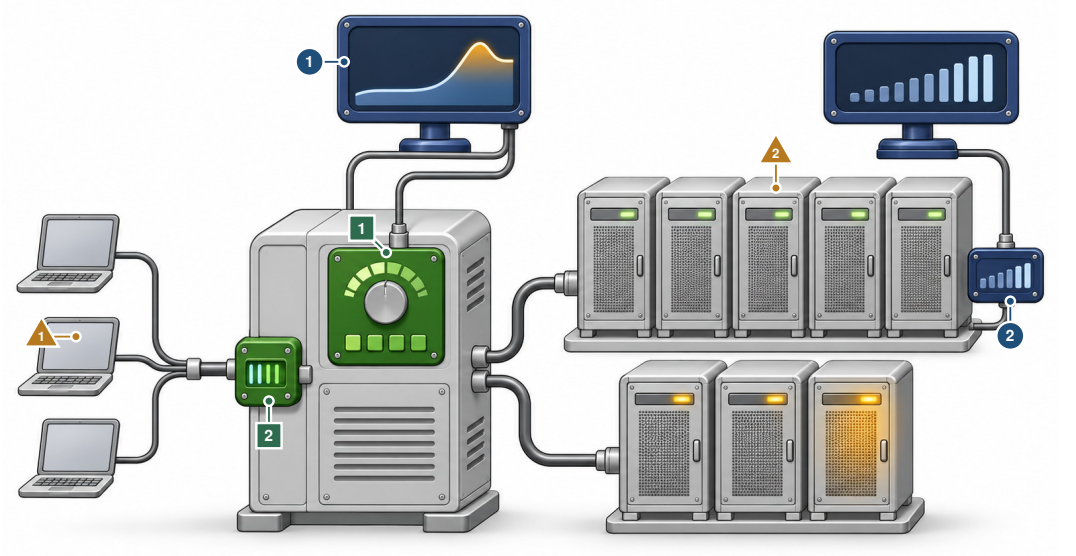
(c) The hot-water setpoint is limited to 95 °C. At what share of the clean UA_H does the cascade run out of room, and when would that happen at this rate? Say what an operator should watch. *Checkpoint:* About 29 %, after about 82 min. The approach.

13.12 Review the agent · 5 min

An AI assistant reviews the challenge with the suggested tuning at a 73.0 °C setpoint and writes: "The 7 diversions come from the 0.2 °C cut-out margin, which trips the FDV on good milk. Set Cut-out margin above the recipe to 0 °C. The recorder then diverts only below the 72.0 °C of the recipe, the run drops to 3 diversions and 686 L, the heat falls to 21.6 °C s, and nothing underprocessed can pass, since the recorder still enforces the recipe." Find the flaw, and say how you would check it in the app.

CHAPTER 14

Integer Servers, Boot Delays and Retry Storms in a Cloud Autoscaling Simulator



The controller moves 1 the autoscaler's order of whole servers, which boot before they serve
2 admission control at the load balancer

The controller reads 1 the minute's 99th-percentile latency against its objective
2 the busy fraction of the ready servers, the metric the autoscaler reads

What disturbs it 1 a flash crowd of users 2 a preemption that removes the spot servers

Requests flow from the users through the balancer to the ready servers, so notice that the booting servers on the lower shelf serve nothing until their boot ends, which is why servers ordered in a crowd arrive minutes after it began. Illustration, not to scale.

IN PLAIN TERMS **Ordering Servers That Arrive Late**

The main idea. A web service runs on a group of computers. A program watches how busy they are and adds or removes whole computers as the traffic rises and falls.

The challenge. A new computer needs about three minutes to start before it can answer anyone. A sudden crowd can come and go before that help arrives. When answers are slow, users press again, and their repeats can keep the service jammed after the crowd has left.

Why it matters. Shops, banks and news sites need a service that stays quick at the busiest moment, which is when extra computers are slowest to help.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 14.2, the fact in Section 14.4, the two try-its and the fact in Section 14.5, and the key points
Study	50 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 14.5), Technology in Practice (Section 14.7), Limitations (Section 14.8) and the two open challenges

The control question An autoscaler orders whole servers that serve only after a boot, on a metric that lags and cannot read above full, while the objective is a latency percentile that stays flat until the servers are nearly full. When does capacity ordered by feedback arrive too late, and what keeps the objective when it does?

Learning objectives

- Compute the 99th-percentile latency from the declared queueing term, and locate the knee and the objective on it.
- Evaluate the platform's desired-replica rule on given readings, and explain why a saturated CPU reading limits each decision.
- Compare the rule, a PI loop and a forecast floor on a ramp by breach minutes and bill.
- Explain how client retries hold a service in overload after its cause has gone, and how a retry budget and admission control break the loop.
- Judge when no autoscaler can meet an objective without shedding.

Prerequisites Sections 1.4.7 (hysteresis and dwell), 1.4.8 (integer actuators), 1.6.2 (the PI law as implemented) and 1.6.4 (scenarios and data). Section 1.6.3 is needed only for the predictive controller.

The app https://apmonitor.com/apps/cloud-autoscaling/

See also Volume 5, Chapter 2 for a disturbance shorter than the dead time of the loop that answers it, Volume 4, Chapter 2 for a protective override, and Volume 2, Chapter 8 for metering a queue short of its breakdown.

14.1 Introduction

At 20:00 in the fourth case of the simulator of this chapter, the traffic to a web service quadruples in two minutes. The autoscaler sees it in its next published metric and orders servers at once, and the first new server begins to serve at 20:03:45, because every server needs 180 s to boot. The crowd holds for six minutes and is gone by 20:12. The fleet keeps growing after that, to the 200-server quota at 20:16, while only 17.2 % of the users' requests between 20:00 and 20:15 were answered within the clients' 2 s timeout. The service spends 14 minutes above its 300 ms latency objective, and 23.7 % of the users' requests from 19:45 to 21:45 are finally dropped. Nothing in the rule was wrong. The capacity it ordered arrived after the need for it had ended. A pellet grill whose lid is opened for three minutes meets the same difficulty, with fuel that burns after the lid has shut (Volume 5, Chapter 2).

Cloud computing made capacity a quantity that software can order by the second [1], and elastic services commonly carry an autoscaler that measures how busy its servers are and adds or removes them [2, 3]. These loops are feedback controllers in the full sense, with a sensor, a setpoint, an actuator and a disturbance [4], and they inherit the classic difficulties of process control in an unfamiliar vocabulary. The actuator is an integer. It arrives after a dead time set by the boot, which motivated capacity managers that keep servers on rather than pay that delay at the worst moment [5]. The measurement is an average that lags, and a busy fraction cannot read above 100 %. The quantity users care about is a tail percentile of latency [6], and queueing theory says that it stays nearly flat until the servers are close to full and then grows as $1/(1-\rho)$ [7, 8]. A request that fails is often sent again, which turns an overload into extra load [9, 10].

The word elastic encourages a belief that this simulator sets out to correct, namely that capacity in the cloud follows the demand. In this model a rule that keeps the average utilization at 70 % follows the morning ramp of case 2 without a single breach minute, yet the same rule leaves 14 breach minutes in a twelve-minute crowd, and with the fleet capped at 60 servers in case 7 the clients' retries keep the service in overload until 22:10, two hours after the crowd ended. The multi-server queue behind that curve is the standard model of a server farm [8], and the same pairing of a discrete, slow capacity with a queue whose wait explodes near saturation appears in call centers staffed by shifts and in hospital wards that open beds. A block diagram of a proportional loop with a delay shows the lag but none of the knee, the integer rounding, the saturated sensor or the retries.

The literature a control student would read treats these pieces in separate places. Surveys classify autoscaling policies by technique [2, 3], control-theoretic work analyzes one loop at a time [4, 11, 12], and the retry storm appears in operations practice and in the study of metastable failures [9, 10, 13]. This chapter takes one browser simulator, in which a declared queueing model, an integer fleet with a boot delay, the platform's metric pipeline, the published scaling rule and the clients' retries run together, and measures on it how each of the standard remedies (a lower target, a cooldown, a forecast, admission control and a retry budget) changes the breach minutes, the dropped requests and the bill. None of the pieces is new. The rule is the documented algorithm of a public open-source platform [14], the latency term is the Kingman and Sakasegawa approximation [7, 15], and the retry storm and the retry budget are described in the sources cited above. The simulator joins them in a day-long fluid model whose request ledgers close to round-off, and the measurements on it include a paired-seed comparison in which the retry budget removes every storm minute on ten of ten seeds. The app scores the nine cases and the graded day by completion rules (Section 1.6.4).

Section 14.2 follows the domain from admission control to the desired-replica rule. Section 14.3 gives the model of the service: traffic, the request fluid, the latency knee, retries, the integer fleet and the metric pipeline. Section 14.4 gives the autoscalers: the platform's rule, a PI loop, the forecast floor, the predictive controller and the two overrides. Section 14.5 measures the cases and the day, Section 14.6 sorts the evidence, Section 14.7 describes the real metrics pipeline, machine launches and controllers, Section 14.8 states the limitations, and Section 14.9 draws the conclusions. The Exercises at the end of the chapter run the cases.

14.2 From Admission Control to the Desired-Replica Rule

Web servers met feedback control before cloud computing existed. In 2002 classical feedback control was applied to a web server end system for overload protection and per-class response-time and throughput guarantees, with a model of the server and named equivalents of its sensors and actuators [11]. Two years later a book collected the modeling, transfer functions, proportional and PID control and state-space design of computing systems in one text [4]. Capacity in that era was bought in advance, so the loops managed admission and resource shares inside a fixed machine rather than the number of machines.

Elastic computing changed the actuator. A view of cloud computing from Berkeley, published in 2010, named the illusion of capacity on demand and pay for use on a short-term basis among the properties that made the cloud new [1]. Control researchers saw that the new actuator was coarse. A study of automated control in clouds treated the virtual machine count as a discrete actuator and proposed proportional thresholding, a target range for an integral controller that depends on the size of the fleet [12]. Forecasting entered next, with a model-predictive algorithm that forecast the workload and allocated resources ahead of it [16], and a capacity manager for multi-tier data centers kept just enough spare servers on to absorb bursts in the request rate [5]. By 2014 a review could classify the techniques into threshold rules, reinforcement learning, queueing theory, control theory and time-series analysis [2], and a 2018 survey built a taxonomy of web-application autoscalers from the challenges that remained open [3].

> ⧖ HISTORY **2011: Elasticity Gets a Federal Definition**
>
> In September 2011 the National Institute of Standards and Technology published a definition of cloud computing built from five essential characteristics, three service models and four deployment models [17]. One characteristic is rapid elasticity. Capabilities are provisioned and released, in some cases automatically, to scale outward and inward with demand, and to the consumer they often appear to be unlimited and available in any quantity at any time. Another is measured service, a metering capability through which the use of resources is monitored, controlled and reported, typically on a pay-per-use basis [17]. Elasticity is the actuator of this chapter's loop and metering is its sensor. The 200-server quota and the 180 s boot of Table 14.2 are what the appearance of unlimited capacity leaves out.

Table 14.1 Variables by Role, With the Values at the Screenshot Instant of Case 4 (20:06)

Symbol	Role	20:06	Unit
p_{99}	CV, minute 99th percentile, objective 300 ms	91.56	s
u	measured, busy fraction or requests per server	1.00	—
u_{SP}	SP of the autoscaler	0.70	—
n_{rep}	MV, replicas, active plus booting	75	servers
n_{act}, n_{boot}	states, active and booting servers	52, 23	servers
λ_{user}	DV, users' first attempts (truth)	9614	req/s
λ_{retry}	state flow, retries arriving (truth)	18 715	req/s
λ_{off}	measured, offered to the balancer	28 329	req/s
—	refused at the queue limit	23 129	req/s
q	state, backlog (truth)	156 000	requests
f_{out}	output, share timing out (truth)	0.904	—
—	answered in time (truth)	498	req/s
—	bill rate	15.00	$/h
—	DVs, crowd, preemption, slow dependency, deploy, forecast miss	—	events

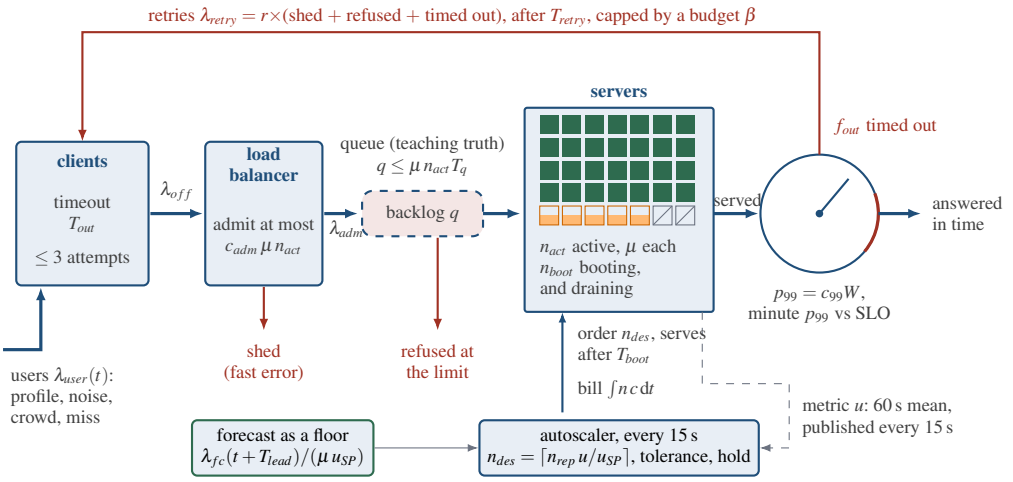
The rule that became the default was the simplest. Container orchestration grew from a decade of internal cluster managers into an open-source platform [18]. The horizontal pod autoscaler of that platform, Kubernetes, multiplies the current replica count by the ratio of the measured metric to its target and rounds up, skips any change within a 10 % tolerance, runs every 15 s, counts pods that are not yet ready as idle when it scales up, holds a scale-down to the highest recommendation of a 300 s stabilization window, and limits a scale-up to doubling or four pods per period [14]. In control terms it is a proportional controller on the load with a deadband, a hold and an integer output. This simulator writes that algorithm out from its public documentation.

What the loop protects changed as well. Large services learned that a tail percentile, not the mean, decides what users experience, because one slow component among many delays a whole request [6], and site-reliability practice turned the percentile into a service-level objective with an error budget, together with rules for handling overload such as refusing work early, capping the attempts per request at three and giving each client a retry budget of 10 % of its requests [9]. The overload that retries sustain after its trigger has gone was later named a metastable failure [10], and recent work analyzes how such a congestive collapse is triggered and why it persists after the trigger is removed [13]. The queueing law under all of it is older. Kingman's 1961 heavy-traffic result gives the mean wait of a single-server queue as $\rho/(1-\rho)$ times the variability and the service time [7], and Sakasegawa's 1977 formula extends the form to several servers [15].

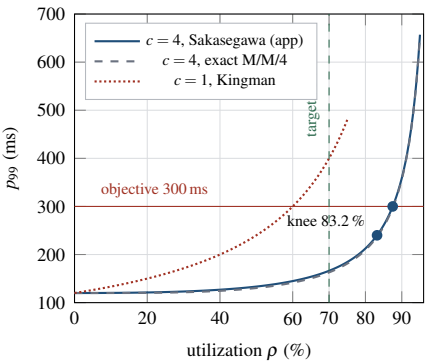
14.3 The Service as Modeled

The simulated system is one stateless web service, drawn left to right as the application's Service view draws it, in Figure 14.1(a). Users send requests at $\lambda_{user}(t)$. Clients wait at most T_{out} for an answer and send a failed request again after T_{retry}, up to three attempts. A load balancer admits at most a share c_{adm} of the active capacity when admission control is on, refuses the rest at once, and holds what it admits in a queue limited to T_q seconds of the fleet's capacity. Servers are integers that pass from ordered to booting to active to draining. The latency of the admitted flow is reported as a per-minute 99th percentile against the objective, and an autoscaler reads the platform's metrics every 15 s and orders servers. The plant integrates in seconds with a step of $\Delta t = 1$ s, and a run starts at the case's clock time.

The quantities that cross the boundary of the loop are classified in Table 14.1. The controlled quantity is the minute's p_{99}, held under an objective as a constraint, while the autoscaler's own setpoint is a utilization target u_{SP} on the metric it reads. The manipulated variable is the integer replica count n_{rep}.



(a)



(b)

Figure 14.1 The autoscaled web service as modeled, and its latency knee. (a) Users at λ_{user} and retries at λ_{retry} arrive as the offered flow λ_{off}. The balancer admits λ_{adm}, sheds the excess over $c_{adm}\mu n_{act}$ when admission control is on, and refuses what would wait beyond $\mu n_{act} T_q$. Active servers (solid squares) serve at μ each, booting servers (half-filled squares) serve after T_{boot}, and drained servers are outlined and crossed. Failures return as retries, and the autoscaler orders n_{des} from the published metric u, with the forecast as a floor. The backlog q is hidden from every controller. (b) p_{99} against utilization from the application's own functions, for four workers per server with the Sakasegawa term the application uses (solid) and with the exact M/M/4 wait (dashed), and for one worker (dotted). The knee at 83.2 % is where queueing doubles the unloaded 120 ms, and the 300 ms objective falls at 87.5 %.

14.3.1 Traffic

The users' traffic is a declared diurnal profile with colored noise, a forecast-miss level and a flash crowd, as shown in Equation (14.1).

$$\begin{aligned}\lambda_{user}(t) = \lambda_{peak}\,D(t_c)\,\big(1+\sigma\,\xi(t)\big) \\ \times L(t)\,\big(1+(X-1)\,F(t-t_X)\big)\end{aligned} \tag{14.1}$$

where t_c is the clock time, D is a periodic Catmull-Rom spline through hourly knots from 0.17 of the peak at 04:00 to 1.0 at 21:00 (1920 and 2400 req/s at 15:00 and 21:00 for $\lambda_{peak} = 2400\,\text{req/s}$, a daily mean of 1574 req/s), ξ is a unit-variance first-order colored noise with a 120 s time constant

drawn on a fixed 10 s grid from the case's seed and interpolated, $\sigma = 0.05$, L ramps from 1 to $1+m$ over 10 min after a forecast miss of size $m = 0.25$, and F rises from 0 to 1 in 2 min, holds 6 min and falls over 4 min after a crowd of multiple $X = 4$ starts at t_X. The step uses the traffic at its midpoint, $t + \Delta t/2$, so that a finer step sees the same traffic.

14.3.2 One Step of the Request Fluid

Requests are tracked by attempt generation $g \in \{0,1,2\}$, the users' first attempts and their first and second retries, with arrival rates a_g and backlogs q_g. Over one step, with the active capacity $C = \mu n_{act}$, the balancer and the servers act as shown in Equation (14.2).

$$\begin{aligned} s &= \begin{cases} \min\big(1,\ c_{adm} C / \textstyle\sum a_g\big) & \text{admission on} \\ 1 & \text{off} \end{cases} \\ A_g &= s\,a_g, \qquad \text{shed}_g = (a_g - A_g)\,\Delta t \\ S &= \min\big(\textstyle\sum_g (q_g + A_g \Delta t),\ C\Delta t\big), \quad S_g \propto q_g + A_g \Delta t \\ w_g &= q_g + A_g \Delta t - S_g, \qquad q_g^{+} = w_g \min\Big(1,\ \frac{C T_q}{\sum w_g}\Big) \end{aligned} \tag{14.2}$$

where A_g is the admitted rate, S the requests served in the step, shared in proportion to what waits, and $w_g - q_g^{+}$ the overflow that the balancer refuses at the queue limit. With $\lambda_{off} = \sum_g a_g$ and $\lambda_{adm} = \sum_g A_g$, the step is exact for flows held constant over it, so the integration has no truncation error beyond the sampling of the traffic.

14.3.3 Latency: Flat, Then the Knee

A server runs c workers, each serving one job of T_{svc} at a time, so its capacity is $\mu = cB/T_{svc}$ with $B = 1$ for a web server. The latency of the step is the service time, the steady-state queueing wait at the present utilization and the fluid backlog drained at the fleet's rate, as shown in Equation (14.3).

$$\begin{aligned} \rho &= \frac{\lambda_{adm}}{\mu\, n_{act}}, \qquad U(\rho, c) = \frac{\rho^{\sqrt{2(c+1)}-1}}{c\,(1-\rho)} \\ W &= T_{svc}\big(1 + V\,U(\min(\rho, 0.98), c)\big) + \frac{\bar{q}}{\mu\, n_{act}} \\ p_{99} &= c_{99}\, W, \qquad f_{out} = 100^{-T_{out}/p_{99}} \end{aligned} \tag{14.3}$$

where $V = (c_a^2 + c_s^2)/2$ is the variability of arrivals and service, $\bar{q}$ is the mean backlog over the step, c_{99} is a declared ratio of the 99th percentile to the mean, and f_{out} is the share of served requests whose latency exceeds the client timeout under a tail declared exponential through p_{99}. At $c = 1$, U is Kingman's $\rho/(1-\rho)$ [7], and the exponent is Sakasegawa's multi-server form [15]. The term is held at $\rho = 0.98$ because above it the backlog carries the overload. With $c = 4$, $T_{svc} = 40$ ms, $V = 1$ and $c_{99} = 3$, the unloaded p_{99} is 120 ms, and the p_{99} is 133, 166, 213, 261, 359 and 657 ms at 50, 70, 80, 85, 90 and 95 % utilization, as shown in Figure 14.1(b). Queueing doubles the unloaded latency at 83.2 %, which the application calls the knee, and the 300 ms objective is crossed at 87.5 %. With one worker per server the knee would sit at 50 % and the objective at 60 %, so no target near 70 % could hold it, which is why the builder chose four.

The approximation is declared, so its error is stated. Against the exact M/M/4 mean wait from the Erlang C formula [8], at $V = 1$ where the two describe the same queue, the term is 28 % high at $\rho = 0.5$ (0.112 against 0.087 service times, a wait too small to matter), 3.5 % high at 0.8 and 1.1 % high at 0.9. The exact queue crosses the 300 ms objective at 87.7 % against the term's 87.5 %. The waiting term is the steady-state value at the present utilization, so a step in load reaches its new latency within the step instead of over the few seconds a real queue takes.

14.3.4 Timeouts, Retries and the Budget

Every refused, shed or timed-out request of generations 0 and 1 is sent again after T_{retry} with probability r, and a third failure is dropped, as shown in Equation (14.4).

$$\begin{aligned} F_g &= \text{shed}_g + \text{overflow}_g + f_{out} S_g \\ R_{g+1}(t + T_{retry}) &= r F_g k_\beta, \qquad g = 0, 1 \\ k_\beta &= \min\Big(1,\ \frac{\beta\, a_0 \Delta t}{r(F_0 + F_1)}\Big) \\ \text{dropped} &= (1 - r)(F_0 + F_1) + F_2 + \text{over budget} \end{aligned} \tag{14.4}$$

where R_{g+1} is the amount of generation $g+1$ that arrives one backoff later, $k_\beta = 1$ unless the retry budget is on, and $\beta = 10\,\%$ caps the retries scheduled in a step at a tenth of the users' requests in it, the figure of the overload chapter of the site-reliability book [9], as is the limit of three attempts. The default $r = 1$ retries every failure. Under the declared tail a 2 s timeout loses 0.01 % of the requests when p_{99} is 1 s, 1 % when it is 2 s and 10 % when it is 4 s.

The request ledgers close by construction, as shown in Equation (14.5).

$$\begin{aligned} \int \lambda_{user}\, \mathrm{d}t &= \text{good} + \text{dropped} + \textstyle\sum_g q_g + \text{pending retries} \\ \int \lambda_{off}\, \mathrm{d}t &= \text{served} + \text{shed} + \text{refused} + \textstyle\sum_g q_g \\ \text{cost} &= \int \left(n_{od}\, c_{od} + n_{spot}\, c_{spot}\right) \mathrm{d}t \end{aligned} \tag{14.5}$$

where good is served minus timed out, n_{od} and n_{spot} count every booting, active, draining and warm server billed at the on-demand and spot prices, and the scorecard's dropped share is dropped over the users' requests. On five disturbed runs, among them the challenge day with a preemption, a crowd and a storm, and case 2 with a slow dependency, a deploy, a forecast miss, a crowd and a 30 % spot preemption under the PI, every ledger closes to a relative 1.1×10^{-14} or better.

14.3.5 An Integer Fleet With a Dead Time

A server is an integer in a pipeline, ordered, booting for T_{boot}, active, and draining for T_{drain} out of rotation after a scale-in. A scale-in cancels booting servers first, the latest ordered first, and then drains active ones. A warm pool, off by default, keeps booted servers out of rotation at full price that join in $T_{warm} = 20\,\text{s}$ and is refilled by cold boots. A share of new servers may be bought on the spot market, and a preemption removes every spot server at once. A deploy restarts the active servers a quarter at a time. Every transition happens at the end of the first step that reaches its due time, so the boot delay is exactly 180 s (480 s for the model-serving fleet) from order to first service. Pumps with minimum run and rest times are integer actuators of the same kind in a water network (Volume 9, Chapter 11).

14.3.6 What the Platform Measures

The autoscaler reads only the platform's metric pipeline. Each 15 s slot accumulates the served work and the offered work in server-seconds, and at the end of each slot the averages over the last $T_m = 60\,\text{s}$ are published as the CPU busy fraction $u_{cpu} = \sum S / \mu \,/ \sum n_{act} \Delta t$, which cannot exceed one, and the requests per server against capacity $u_{req} = \sum \lambda_{off} \Delta t / \mu \,/ \sum n_{act} \Delta t$, which can. The minute's p_{99} is the served-weighted mean of the step values over each clock minute, and a minute above the objective is a breach minute. The counts of ready, booting and draining servers are known at once. The backlog, the true latency, the users' rate apart from their retries and the timeout share are truth columns, prefixed in the export and read only by the scene and the scorecard, and a test fails if the controller module reads any of them.

Table 14.2 Parameters. M: Show Model Dialog, C: Control Card, A: Advanced Control, S: Source. T: Teaching Value, Doc: Platform Documentation, SRE: Site-Reliability Book

Symbol	Meaning	Value	Unit		
λ_{peak}	evening peak of the profile	2400	req/s	M	T
σ, τ_ξ	traffic noise, its time constant	5, 120	%, s	C, S	T
X, m	crowd multiple, forecast miss	4, 25	—, %	A	T
T_{svc}, c	service time, workers per server	40, 4	ms, —	M	T
μ	capacity per server, c/T_{svc}	100	req/s	S	derived
V, c_{99}	variability, p_{99} over mean	1, 3	—	M	T
T_q	balancer queue limit	30	s	M	T
T_{out}, T_{retry}, r	timeout, backoff, share retried	2, 2, 1	s, s, —	M	T
—	attempts per request, budget β	3, 10	—, %	S	SRE
T_{boot}, T_{drain}	boot, drain	180, 30	s	M	T
n_{min}, n_{max}	floor, quota	2, 200	servers	M	T
c_{od}, c_{spot}	on-demand, spot price	0.20, 0.06	$/h	M	T
T_m	metric window	60	s	M	T
—	sync period, tolerance	15, 0.10	s, —	S, A	Doc
T_{cool}	scale-in stabilization window	300	s	C	Doc
—	scale-up limit per period, $\max(n, 4)$	—	servers	S	Doc
u_{SP}	utilization target	70	%	C	T
SLO	p_{99} objective	300	ms	C	T
c_{adm}	admission cap	85	%	C	T
K_c, τ_I, δ	PI gain, integral time, deadband	0.5, 300, 1	—, s, servers	C, A	T
Model-serving fleet, case 8					
c, B	GPUs per server, batch	2, 4	—	M	T
t_0, t_1	batch time $t_0 + t_1 B$	0.10, 0.05	s	M	T
T_{boot}, λ_{peak}	boot, peak	480, 300	s, req/s	M	T
—	SLO, target, cap, price	2, 60, 70, 2.50	s, %, %, $/h	C, M	T

14.3.7 Parameters and Numerics

Table 14.2 gives every parameter. All are teaching values except those read from a cited source, and none is a provider's price, instance type or trace.

The engine integrates in seconds at $\Delta t = 1$ s, the controller runs every 15 s, and the historian logs every 5 s. Every boot, drain, sync period, metric slot, backoff and minute is a whole number of steps. Halving the step to 0.5 s leaves the breach minutes and the reversals of every case unchanged, moves the dropped shares by at most 0.001 points and the costs by at most $0.01, except two day-long runs, the rule in the forecast's place ($90.12 against $90.28) and the PI ($98.80 against $98.92), and no challenge verdict changes. A simulated day costs 64 ms of computing headless and plays in about 24 s at the fastest speed in a browser. At the case 1 design point with the noise off the closed form gives $\rho = 0.6857$ and $p_{99} = 162.2$ ms, against 0.6854 and 162.2 ms from the engine.

14.3.8 Assumptions

The queue is a fluid, so requests are a continuous flow and the waiting time is the steady-state approximation at the present utilization, accurate to the figures above for $V = 1$. The percentile is a declared multiple of the mean with an exponential tail above it, while real services measure their own, and an exponential sojourn time would give $c_{99} = \ln 100 = 4.6$. There is one tier of identical servers behind a balancer that spreads load evenly, with no database, cache or downstream queue, so the slow-dependency event only lowers μ by 40 % for 10 min. A request whose client has timed out is still served, while servers that receive the client's deadline can drop it and shorten a storm. The rule is a written-out copy of the documented algorithm, not its code. Its readiness handling is reduced to booting servers counted at 0 % of the target on a scale-up and at 100 % on a scale-down, which the documentation applies to pods with missing metrics rather than to pods not yet ready. Billing is per second.

14.4 Autoscalers as Implemented

Every automatic mode decides every 15 s from the published metric u, which is u_{cpu} or u_{req} as the learner chooses, and the counts. Its result then passes the platform's two limits and the fleet is reconciled to it at once, which is the integer actuator, whose dead time is the boot.

14.4.1 The Platform's Rule

The rule computes a desired count as shown in Equation (14.6).

$$\begin{aligned} \phi &= u/u_{SP}, \qquad \text{no action if } |\phi - 1| \le 0.1 \\ \phi' &= \begin{cases} u\,n_{act}/n_{rep}\,/\,u_{SP} & \phi > 1 \\ (u\,n_{act} + u_{SP}\,n_{boot})/n_{rep}\,/\,u_{SP} & \phi < 1 \end{cases} \\ n_{des} &= \lceil \phi'\, n_{rep} \rceil \quad \text{unless } \phi' \text{ reverses or } |\phi' - 1| \le 0.1 \end{aligned} \tag{14.6}$$

where the recount ϕ' is made only while servers are booting. The copy reproduces the documented computation on declared inputs. Twenty ready servers at 90 % against a 70 % target give $\lceil 25.71 \rceil = 26$, a reading of 76 % is within the tolerance, twenty ready and four booting at 90 % recount to 1.071 and take no action, at 100 % they give $\lceil 1.1905 \times 24 \rceil = 29$, and at 40 % they give 16. A scale-in then takes the highest recommendation of the last T_{cool} seconds, a scale-out is limited to $n_{rep} + \max(n_{rep}, 4)$ per period, and the result is clipped to the floor and the quota.

As a controller the rule is proportional to the measured load, because $n_{rep}u/u_{SP}$ is the load in units of servers at the target. It has a deadband (the tolerance), a hold (the window), an integer output, a measurement lag of about half the 60 s window plus the publication delay, and a dead time equal to the boot. Counting booting servers as idle on a scale-up is supply-line accounting, so the rule does not order twice for load that servers on order will carry. On the CPU metric a saturated reading limits each decision to a multiple $1/u_{SP}$ of the ready fleet, which is $1/0.7 = 1.43$ at the shipped target.

> **+ DID YOU KNOW? A Cooling Time of at Least One Boot**
>
> A common practice against oscillation is a fixed minimum wait between scaling operations, at least as long as it takes to acquire, boot and configure a machine, and a long wait brings more violations because scale-out slows [3]. The rule of this chapter holds only scale-in, for 300 s against a 180 s boot (Table 14.2).

14.4.2 A Proportional-Integral Autoscaler

The PI option runs the shared velocity-form PID of Section 1.6.2 on the projected utilization $u\,n_{act}/n_{rep}$, the reading with booting servers counted as idle. It is direct acting, so a utilization above target raises the count, and its gain is scheduled so that $K_c = 1$ makes the proportional step equal to the rule's, as shown in Equation (14.7).

$$\begin{aligned} e_k &= \frac{u_k\, n_{act}}{n_{rep}} - u_{SP}, \qquad K_k = \frac{K_c\, n_{rep}}{u_{SP}} \\ s_k &= s_{k-1} + K_k e_k - K_{k-1} e_{k-1} \\ \hat{n}_k &= s_k + I_{k-1}, \qquad n_k^{PI} = \min\bigl(n_{max}, \max(n_{min}, \hat{n}_k)\bigr) \\ I_k &= I_{k-1} + \left[\frac{K_k}{\tau_I} e_k + \frac{n_k^{PI} - \hat{n}_k}{\tau_I}\right]\Delta t_c + \frac{n_k^{app} - n_k^{rec}}{\tau_I}\Delta t_c \end{aligned} \tag{14.7}$$

where $\Delta t_c = 15$ s, the running sum s_k is never clamped, and s_0 and I_0 are set so that the first output equals the starting fleet. The real-valued output becomes an order only when it leaves a deadband of δ servers around the present count (the integer deadband of Section 1.4.8), and the order is then $n^{rec} = \text{round}(n^{PI})$, at least one server away. The last term is the anti-windup against the limits that act after the PI, back-calculation with tracking time τ_I toward the count n^{app} that the cooldown and the

scale-up limit actually applied. The deadband is the PI's own and is not fed back. With no integral action nothing is back-calculated, and a PI that is not driving tracks the applied count for a bumpless switch.

The last term was added in the technical review. Without it, in case 2 at an 80 % target, the PI's order differed from the applied count on 3060 of the 14 400 steps, held by the cooldown or the scale-up limit, and its integral ran on through the hold. When the window let go, the PI removed a server on the rising ramp and the service fell into a retry storm, with 100 breach minutes, 22.3 % dropped and $18.50 over the four hours. Back-calculation against the applied count gives 70 breach minutes, no drops and $14.50 on the same run. Freezing the integral during a hold was also tried and rejected, because it stopped the scale-ins after the crowd and raised the challenge bill under the PI from $90.86 to $111.40.

14.4.3 The Forecast as a Floor

The forecast mode reads the declared diurnal profile at the planned peak, times one plus an adjustable bias, one lead ahead, and orders at least what that traffic needs at the target, as shown in Equation (14.8).

$$\begin{aligned} n_{ff} &= \frac{\lambda_{fc}(t+T_{lead})}{\mu\, u_{SP}}, \qquad n_{des} = \max\left(\lceil n_{ff} \rceil,\, n_{des}^{rule}\right) \\ T_{lead} &= T_{boot} + 15\,\mathrm{s} + T_m/2 = 225\,\mathrm{s} \end{aligned} \tag{14.8}$$

so the servers for a ramp are ordered one boot ahead of it while the rule still answers whatever the forecast misses. The design plan for the app called for a feedforward with a PI trim, and that structure was measured and replaced. On case 6 at an 80 % target with the miss, on five seeds, the floor breached 5, 24, 7, 9 and 12 minutes against 29, 80, 25, 35 and 34 for the trim at $K_c = 0.5$, and trims fast enough to catch the miss ($K_c = 0.8$) breached up to 123 minutes. In the idealized limit of a linear bill, a static utilization constraint and a known boot, the cost minimum orders the forecast's requirement one boot ahead, which the floor already computes. The predictive option below instead penalizes tracking error and count movement over a finite horizon.

14.4.4 Model Predictive Control

The model predictive controller (MPC) uses receding-horizon optimization [19] with tracking and movement penalties used in industrial MPC [20], set up as Section 1.6.3 describes. Nominal cold-provisioning physics gives the 12-state shift register of Equation (14.9).

$$\begin{aligned} x_{j,k+1} &= x_{j+1,k}, \quad j = 0,\ldots,10 \\ x_{11,k+1} &= v_k, \quad y_k = x_{0,k} \end{aligned} \tag{14.9}$$

where $x_{j,k}$ are pipeline counts, v_k is desired replicas, y_k predicted ready count and $T_s = 15\,\mathrm{s}$ the decision interval. Measured ready and booting counts and the controller's own orders initialize the register. Unknown boot ages are spread over 180 s, and newest orders cancel first. An isolated three-server cold addition has zero peak and RMS count residual over 600 one-second samples. Warm joins at 20 s, immediate scale-in and GPU boots at 480 s give RMS residuals of 1.549, 1.639 and 2.121 servers.

Demand is $d_k = u_k \max(1, n_{ready,k})/u_{SP}$, from the selected metric and measured ready count $n_{ready,k}$. Capacity stays at nominal $\mu_0 = 100\,\mathrm{req/s}$ per server. Additive bias compensation, related to offset-free MPC [21], supplies Equation (14.10).

$$\begin{aligned} f(t) &= \lambda_{fc}(t)/(\mu_0 u_{SP}), \quad a_k = 1 - e^{-\Delta t_k/(60\,\mathrm{s})} \\ b_k &= b_{k-1} + a_k\left(d_k - f(t_k) - b_{k-1}\right) \\ r_{i|k} &= \max\left(0, f(t_k + iT_s) + b_k\right), \quad i = 1,\ldots,40 \end{aligned} \tag{14.10}$$

where demand offset b_k starts at $d_0 - f(t_0)$ and Δt_k is elapsed decision time. Preview Off holds $r_{i|k} = d_k$. Forecasts use only the declared daily profile and adjustable bias, without future crowd, miss or noise. Windowed utilization times current ready count approximates demand during fleet changes. The crowd's fleet-to-window ratio reaches 1.25.

Shared LinearMPC/DenseQP solves eleven real counts z_ℓ over 40 steps (10 min), blocked as $(1,1,1,1,2,2,3,4,6,9,10)$, with objective J in Equation (14.11).

$$\begin{aligned} \min_{z} J = &\sum_{i=1}^{40} (y_{i|k} - r_{i|k})^2 + 0.20 \sum_{\ell=1}^{11} (z_\ell - z_{\ell-1})^2 \\ &+ \frac{0.20 T_s}{3600\,\mathrm{s}} \sum_{i=0}^{39} v_{i|k} \\ \text{subject to } & n_{min} \le z_\ell \le n_{max}, \quad |z_\ell - z_{\ell-1}| \le h_k \end{aligned} \tag{14.11}$$

where $v_{i|k}$ holds its block count, $z_0 = n_{rep,k}$ and $h_k = \min(h_{set}, \max(4, \lceil n_{rep,k} \rceil))$, default $h_{set} = 4$. Numerical counts are in servers and currency in dollars, at nominal price \$0.20 per server-hour. Box bounds (default 2–200) and symmetric block-boundary caps are hard. There is no terminal penalty, latency constraint or soft band, and slack is zero. Small default diagonal regularization applies. Only the first count is rounded, capped and clipped before stabilization, the platform scale-out policy and data-plane protections. Cooldown is not predicted. This is a relaxed QP with integer actuation.

Solves occur at initialization and every 15 s, holding commands between decisions. PI tracks the applied count. Nonoptimal, nonfinite or out-of-bounds plans hold that count on first failure, then use PI until recovery. Manual transfer retains the post-stabilization request before actuator caps. The nonlinear plant, instrument packet and frozen predictor stay separate. Backlog and hidden transition times are unavailable. Displayed p_{99} uses the nominal steady web curve without backlog.

14.4.5 Admission Control and the Retry Budget

Admission control is the override. With it on, the balancer admits at most $c_{adm} = 85\,\%$ of what the active servers can serve and refuses the rest at once as a fast error, so the utilization of the admitted flow cannot exceed 0.85 and p_{99} cannot exceed 261 ms once the backlog has drained. The refused requests are failures, so with $r = 1$ they come back as retries. A ramp signal that meters cars onto a freeway short of its breakdown is the same kind of override on another queue (Volume 2, Chapter 8), and Volume 4, Chapter 2 treats a protective override that rests at its clamp until the plant nears a limit. The retry budget acts on the clients instead, as Equation (14.4) states. Neither is a controller of the server count. A manual station sets the count directly, and a helper sets it to what the measured request rate needs at the target. Three models are kept apart. The plant is everything in Section 14.3, the rule carries no model beyond the counts, and the forecast knows only the declared profile and the capacity per server, not the crowd, the noise or the miss.

14.5 Results

The cases are listed in Table 14.3. Each starts from the shipped settings of Table 14.2 (the rule at 70 % on the CPU metric, a 300 s cooldown, admission control and the retry budget off, seed 20261026) at the clock time its row gives, with the fleet sized for the opening traffic at the target, and changes only what the row lists. The figures below come from headless runs of the engine, advanced step by step the way the browser advances it, over the window each sentence gives, and a setting that a sentence mentions was made before the run began. A case repeats bit for bit on its seed, as Section 1.6.1 explains. Table 14.4 collects the comparisons.

14.5.1 The Knee and the Boot by Hand

Case 1 opens at 15:00 in Manual with 28 servers, about 1920 req/s on average over the first ten minutes, a CPU reading averaging 69 % and a minute p_{99} of 151 to 192 ms. Removing servers at

Table 14.3 The Cases and the Challenge

#	Start, change and events	Teaching point
1	15:00, Manual at 28 servers	the knee and the boot
2	06:00, rule through the morning ramp	proportional, band, hold
3	21:30, cooldown 0 s, noise 8 %	flapping
4	19:45, crowd ×4 at 20:00	capacity after the crowd
5	case 4 with admission control	the override
6	06:00, forecast at 80 %, miss +25 % at 08:00	feedforward and its error
7	case 4 with a 60-server quota	the retry storm
8	19:00, GPU fleet, boot 480 s, crowd at 20:00	infeasible without shedding
9	06:00, MPC, Requests, target 70 %	ten-minute plan against the forecast floor
ch	00:00, 30 % spot, preemption at 14:00, crowd at 20:00	the whole day

Table 14.4 Measured Cases. Breach Minutes, Dropped Share of the Users' Requests and Bill Over the Window

Case, window	Setting	Breach (min)	Dropped (%)	Bill ($)
2, 06–10	rule 70 %	0	0	17.42
	rule 80 %	18	0	15.41
	PI 70 %	0	0	16.66
	PI 80 %	70	0	14.50
3, 21:30–01	cooldown 0 s	4	0	15.88
	cooldown 300 s	0	0	16.80
4, 19:45–21:45	as shipped	14	23.7	20.37
5, same	admission 85 %	0	14.0	15.81
	admission, Requests metric	0	4.8	21.37
6, 06–10	forecast 80 %	5	0	17.43
	rule 80 %	26	0	17.49
7, 19:45–23:45	as shipped	130	52.2	36.39
	retry budget	13	15.0	27.19
	admission control	0	8.7	26.91
8, 19–21	as shipped	18	27.4	136.28
	admission 70 %	0	16.4	101.51

15:10 moves along the knee. At 26 servers the p_{99} at 15:25 is 192 ms, at 24 it is 237 ms with no breach, at 23 the reading is 86.4 % and one minute breaches at 315 ms, and at 22 the reading is 90.3 % and five minutes breach, the worst at 466 ms. One server separates a clean quarter hour from a breaching one. Restoring 28 servers at 15:25 takes effect 180 s later, and the first clean minute ends at 15:29, three more breach minutes after the order. A removal acts in the next step and an addition one boot later, which is the asymmetry every autoscaler has to reason about.

14.5.2 The Rule on the Morning Ramp, and the PI

From 06:00 the traffic rises from 624 to 1920 req/s by 09:00. At a 70 % target the rule's reading averages 70.0 % from 07:00 to 09:00 with a peak of 80.7 %, and no minute breaches over the four hours, for $17.42. At 80 % the average is 79.6 %, the peak 91.8 %, and 18 minutes breach for $15.41. The rule chases the ramp from behind, because each decision acts on a reading about half a minute old and serves three minutes later, and the higher target leaves the slope of the knee less room for that lag. A saving of $2.01 buys 18 breach minutes.

The PI at its defaults ($K_c = 0.5$, $\tau_I = 300$ s) holds the 70 % ramp without a breach for $16.66 in 18 actions with none reversed, and at $K_c = 1$ it reverses 8 of 27 actions. At 80 % it is the cheapest, $14.50, and breaches 70 minutes, because a proportional controller with integral action follows a ramp with a standing error and its reading averages 85.3 % from 07:00 to 09:00 against the rule's 79.6 %. On the Requests metric the same PI gives the same 70 minutes, so the saturating CPU metric is not the cause. A target is a utilization, but the objective sits on the steep part of a curve, and a controller's error on a ramp has to be read on that curve.

🕮 **Pause.** Try Exercise 14.2 before reading on.

14.5.3 Flapping and the Cooldown

Case 3 runs the rule from 21:30 to 01:00 on the falling evening shoulder with 8 % noise and no cooldown. It takes 67 actions and reverses 29 of them, removing servers on quiet readings and ordering them back on the next burst at three minutes' notice, and 4 minutes breach for $15.88. A 60 s cooldown still reverses 19 times. The default 300 s takes 18 actions with none reversed and no breach, for $16.80. The hold costs $0.92 and removes the breaches, the reverse of the design plan's expectation that it would cost a breach, because here the reversals themselves caused them. On seeds 7 and 11 the counts without a cooldown are 37 and 28 reversals against 4 and 5 with it. Flapping is a limit cycle of a deadband controller on a noisy measurement with a delayed actuator, and the window is a hold that trades money for calm, the minimum dwell of Section 1.4.7 applied to scale-in only.

> **TRY IT IN THE APP** **3 · Flapping on the Evening Shoulder**
>
> Load the scenario and press Run. Watch the servers on the Day view and the Scaling actions tile. With no cooldown the rule removes servers on quiet readings and orders them back on the next burst. By 01:00 it has taken 67 actions and reversed 29 of them, and 4 minutes have breached.

14.5.4 The Flash Crowd and Admission Control

Case 4 is the crowd of the Introduction. From 36 servers at 20:00, the rule on the CPU metric can at most multiply the ready fleet by $1/0.7$ per boot, because the saturated reading cannot ask for more. The first new server serves at 20:03:45, 108 are active when the crowd ends at 20:12, and the fleet peaks at its 200-server quota at 20:16, sixteen minutes after the crowd began. The breach minutes end at 20:14. A boot of 20 s still leaves 12 breach minutes, because the multiplier per decision and the full queue's stale work matter as much as the boot. The state at 20:06 is shown in Figure 14.2 and listed in Table 14.1. Clients send 9614 requests per second and retries 18 715 more, the balancer refuses 23 129 at the queue limit, the backlog is 30 s of the 52 active servers' capacity, and of the 5200 req/s served only 498 arrive in time.

Pause. Try Exercise 14.1 before reading on.

Case 5 is the same crowd on the same seed with admission control on. The balancer admits at most 85 % of the active capacity and refuses the rest at once, so no minute breaches and the worst minute is 261 ms. Of the users' requests from 20:00 to 20:15, 51.1 % are answered in time against 17.2 % without it, and 14.0 % of the whole window's requests are dropped against 23.7 %, for $15.81 against $20.37. The design plan expected shedding five percent to save the rest, which does not hold for a fourfold crowd. The override also hides the overload from the CPU metric, which reads only what was admitted, so the fleet peaks at 73 servers. On the Requests metric, which counts the refused demand, the fleet reaches its quota in 4.8 minutes, 83.3 % of the crowd's requests are answered in time and 4.8 % are dropped, for $21.37. On seeds 7 and 11 the shipped crowd breaches 14 minutes each and the admitted one none.

> **DID YOU KNOW?** **Refusing Work Drops Less of It**
>
> From the simulator: in the fourfold crowd the shipped rule breaches 14 minutes and drops 23.7 % of the users' requests for $20.37. Admission control at 85 % of the active capacity breaches none, drops 14.0 % and bills $15.81 (Table 14.4, cases 4 and 5).

14.5.5 The Forecast

Case 6 runs the forecast floor at an 80 % target from 06:00, with the traffic climbing 25 % above the forecast from 08:00. No minute breaches before 08:00 against the rule's 17 at the same target, and five breach in all against 26, for $17.43 against $17.49. Without the miss the forecast breaches 1 minute against 18. The floor orders the ramp a lead of 225 s ahead, and the rule above it answers the miss.

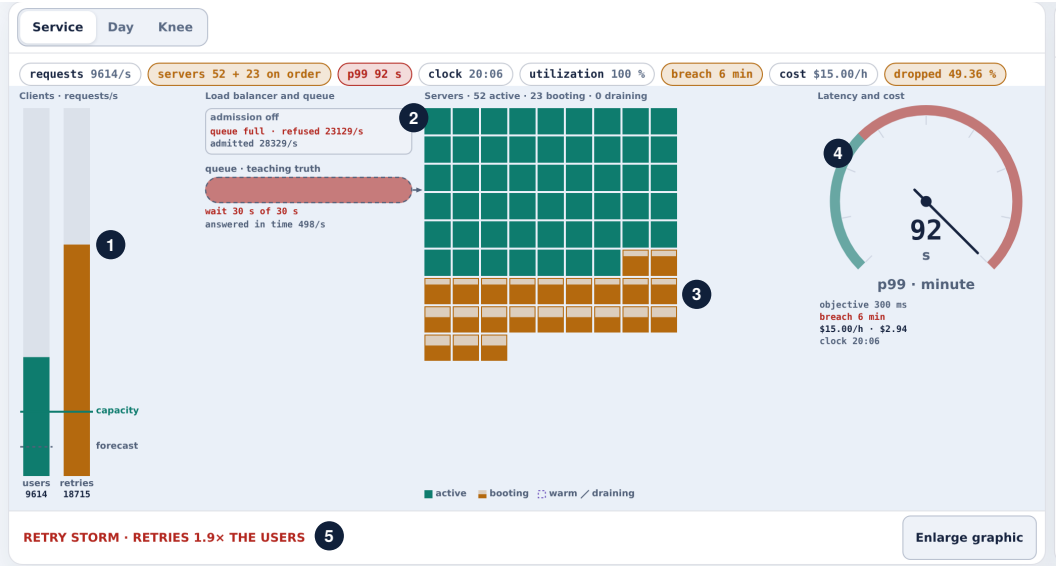


Figure 14.2 The flash crowd of case 4 at 20:06 on the Service view, as shipped (rule at 70 % on the CPU metric, admission control off). The picture is cropped to the Service view, and the header, the control panel, the trends and the Learn panel are cut. (1) The clients' bars show 9614 users and 18 715 retries per second. (2) The balancer refuses 23 129 per second at its queue limit, the queue drawn as teaching truth is full at 30 s of the capacity, and 498 requests per second are answered in time. (3) The rack holds 52 active and 23 booting servers. (4) The gauge reads a minute p_{99} of 92 s against the 300 ms objective and a bill of $2.94 so far. (5) The status strip reads RETRY STORM at 1.9 retries per user. The chip row above the scene reads the utilization pinned at 100 %, 6 breach minutes and 49.36 % of the users' requests dropped since 19:45.

14.5.6 The Retry Storm

Case 7 is the crowd of case 4 with the quota cut to 60 servers, a fleet that can serve 6000 req/s against an evening peak of 2400 req/s. Without a budget, the retries exceed the users in every minute from 20:02 to 22:10, 129 minutes, and 130 minutes breach. The trigger lasted twelve minutes. At 21:00 the users send 2629 requests per second, 44 % of the capacity, yet the minute p_{99} is 91.6 s, because every request now arrives up to three times and the retries run at 1.64 to 1.78 times the users. The service has two states at the same traffic. Healthy, almost nothing fails and the offered load is the users' load. Overloaded, almost everything fails, the offered load is near three times the users' load, the queue stays at its limit, and the overload sustains itself. The loop is a positive feedback, failures cause retries and retries cause failures. This one ends only when the evening traffic has fallen far enough that users and retries together stay below what the fleet serves for long enough to drain the full queue, which happens at 22:10.

🕮 **Pause.** Try Exercise 14.3 before reading on.

The budget breaks the loop. Table 14.5 pairs the shipped case with the same case under the 10 % budget on ten seeds. Without it, nine seeds storm for 107 to 129 minutes and one, seed 3, for 22 minutes, because there the evening noise dipped the users low enough to drain the queue at 20:23. With the budget, no seed has a minute with more retries than users, the largest ratio is 0.10, and the breach minutes fall to 12 or 13, the crowd's own. A sweep of the share of failures retried on the shipped seed shows the gain of the loop. At $r = 0.5$ there is no storm minute, at 0.6 one, at 0.7 fourteen, at 0.8 seventeen, at 0.9 twenty-eight and at 1.0 the full 129. A crowd of 1.5 times the traffic breaches 4 minutes with no storm, while 2, 2.5 and 3 times all storm to 22:10. Admission

Table 14.5 Case 7 on Paired Seeds, 19:45 to 23:45. Storm Minutes Have More Retries Than Users

	No budget			Budget 10 %		
Seed	storm (min)	breach (min)	dropped (%)	storm (min)	breach (min)	dropped (%)
20261026	129	130	52.2	0	13	15.0
1	114	115	46.9	0	13	14.8
2	118	120	48.6	0	12	14.6
3	22	24	16.2	0	12	13.2
4	118	120	49.0	0	12	15.4
5	107	109	45.1	0	13	15.3
7	127	128	51.1	0	12	15.4
11	115	117	48.4	0	13	15.5
23	121	122	48.7	0	12	14.7
99	121	123	49.8	0	12	15.2

control alone also prevents the storm's breaches (none), because a fast refusal keeps the latency of the admitted requests under the objective, although it still drops 8.7 % of the requests and retries still exceed the users for nine minutes.

> **TRY IT IN THE APP 7 · The Retry Storm**
>
> Load the scenario and press Run. From 20:02 the retries outnumber the users, and the service stays in overload until 22:10, two hours after the crowd ended. Switch Retry budget (10 %) on, press Reset and run again: no minute has more retries than users, and 13 minutes breach in all.

14.5.7 A Fleet That Cannot Be Saved Without Shedding

Case 8 serves a model on GPU servers, two GPUs each, that batch four requests in $t_0 + t_1 B = 0.3$ s, so a server carries 26.7 req/s, the unloaded p_{99} is 0.9 s against a 2 s objective, and a server boots in 480 s. Batching trades latency for throughput. At the objective a server answers 16.7, 19.7 and 15.4 req/s in time at batches of 2, 4 and 8, and a batch of 8 breaches 120 minutes of the two hours even without a crowd. With the crowd at 20:00 the first new server serves at 20:08:30, as the crowd begins to fall, 18 minutes breach and 27.4 % of the requests are dropped. A sweep of 168 configurations without admission control (rule, PI and forecast, both metrics, targets from 30 to 90 %, warm pools of 0, 5, 10 and 20 servers) finds none within 5 breach minutes. The best, the rule at 30 % with 20 warm servers, breaches 7 minutes for $387.34 in two hours. With admission control at 70 % the shipped autoscaler breaches none and drops 16.4 %, for $101.51. The objective is infeasible through the crowd for all 168 tested configurations, and the case says so. Only shedding keeps it.

14.5.8 The Day

The graded day starts at midnight and lasts 24 hours, with 30 % of new servers bought spot, a preemption at 14:00 and the crowd at 20:00. Its target is at most 5 breach minutes, at most 1 % of the users' requests dropped and a bill of at most $92. The drop target is 1 % rather than the design plan's 0.5 % because no autoscaler can do better than about 0.6 % once the crowd has begun. An oracle that, under the suggested settings, switches to Manual and orders 130 servers at 20:00:00 for twenty minutes still drops 0.599 % of the day's requests, and 160 servers drop 0.595 %, against 0.038 % for the same day without the crowd. Summed over the first three minutes of the crowd, the users' requests above what the starting fleet can admit come to 0.80 million, 0.57 % of the day's 139.7 million. Table 14.6 compares nine configurations.

The shipped rule fails all three clauses, with 3 breach minutes at the 14:00 preemption and 14 in the crowd, 3.49 % dropped and $94.21. The suggested configuration, the forecast at 75 % on the Requests metric with admission control, passes at 0 breach minutes, 0.82 % and $89.49. What does the work there is the metric and the override, not the forecast. With admission control on, a ramp

Table 14.6 The Day in Nine Configurations (Target: ≤ 5 Breach Minutes, ≤ 1 % Dropped, ≤ $92)

Configuration	Breach (min)	Dropped (%)	Bill ($)	Verdict
shipped: rule 70 %, CPU	17	3.49	94.21	fails
rule 70 % + admission	0	2.01	90.15	fails
PI 70 %, Requests, admission	0	0.87	98.80	fails
forecast 70 %, Requests, admission	0	0.78	95.86	fails
forecast 75 %, Requests, admission (suggested)	0	0.82	89.49	passes
suggested on CPU	0	2.31	84.25	fails
suggested with the rule in place of the forecast	0	0.78	90.12	passes
suggested + retry budget	0	0.95	88.85	passes
suggested + two warm servers	0	0.76	103.59	fails

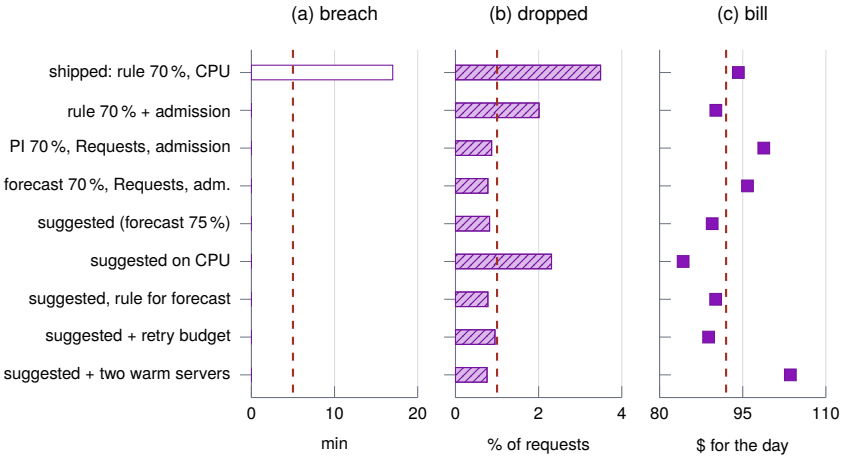


Figure 14.3 The graded day of Table 14.6 in its nine configurations: (a) breach minutes as white bars, (b) the share of the users' requests dropped as shaded bars and (c) the bill as dark squares on an axis that starts at $80, with each target dashed. The table prints the values. Notice that every configuration with admission control keeps the breach target, and that the drop and bill targets then separate them: only the rows on the Requests metric drop less than 1 %, and the warm servers that drop least cost the most.

cannot breach, because the admitted utilization is capped at the 261 ms point, and the rule in the forecast's place also passes, at 0.78 % and $90.12, so the forecast's contribution on this day is $0.63 of the bill, at 0.04 points more of the requests dropped. The same pieces at the shipped 70 % fail on cost, $95.86, and on the CPU metric the override hides the crowd from the rule and 2.31 % is dropped. The warm pool the design plan suggested cuts the drops to 0.76 % but costs $14.10 more. On nine further seeds the shipped rule fails with 14 to 28 breach minutes and 3.46 to 4.54 % dropped, and the suggested configuration passes every time at 0.80 to 0.89 % and $89.78 to $90.92. Figure 14.3 draws the nine configurations against the three targets.

14.5.9 Predictive Orders Against the Forecast Floor

Case 9, "A ten-minute plan against the forecast floor," repeats 06:00–09:00 with Requests and a 70 % target. Table 14.7 matches seed 20261026, instruments, noise and case settings, with MPC weight 0.20 and cap 4 unchanged. Cases 6 and 4 retain targets of 80 and 70 %. Forecast and MPC On receive the same daily schedule without future crowd or miss events. Rule and PI have no preview. Off holds measured demand. The earlier CPU results remain separate.

All five ramp runs have zero breaches and no drops. PI is cheapest at $11.06, against Rule $11.81750, MPC On $11.93333 and Forecast $12.08333. Preview reduces the MPC bill from

Table 14.7 Matched Requests Comparisons, Seed 20261026. MPC Settings Stay Fixed

Case, window	Law	Breach (min)	Dropped (%)	Billed (server h)	Bill ($)	IAE (util·s)	Travel (servers)	Limit (s)	Solves/ fallback	Mean (µs)
9, 06–09	Rule	0	0.000	59.087	11.81750	412.9	22	0	0/0	—
	PI	0	0.000	55.300	11.06000	622.9	18	0	0/0	—
	Forecast	0	0.000	60.417	12.08333	341.9	20	0	0/0	—
	MPC Off	0	0.000	60.233	12.04667	336.6	36	0	720/0	175.9
	MPC On	0	0.000	59.667	11.93333	303.2	27	0	720/0	182.4
6, 06–10	Rule	26	0.000	87.438	17.48750	718.0	33	0	0/0	—
	PI	97	7.262	106.288	21.25750	2095.2	91	0	0/0	—
	Forecast	5	0.000	87.171	17.43417	496.1	21	0	0/0	—
	MPC Off	8	0.000	89.246	17.84917	546.4	51	0	960/0	174.8
	MPC On	11	0.000	88.400	17.68000	510.0	42	0	960/0	181.3
4, 19:45–21:45	Rule	12	19.671	118.292	23.65833	1842.6	327	900	0/0	—
	PI	12	20.444	152.096	30.41917	3020.6	336	720	0/0	—
	Forecast	12	19.671	118.292	23.65833	1842.6	327	900	0/0	—
	MPC Off	13	22.368	289.200	57.84000	5222.6	242	360	480/0	198.3
	MPC On	13	22.492	289.437	57.88750	5319.9	242	390	480/0	199.2

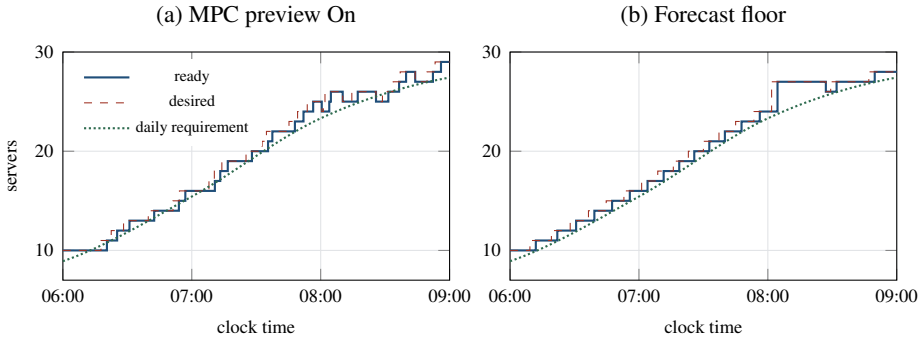


Figure 14.4 Morning server orders under MPC (a) and Forecast (b). Both use Requests, a 70 % target and seed 20261026 through 09:00. Ready counts (solid) and desired orders (dashed) are compared with the daily forecast's capacity requirement (dotted). Neither run breaches, at bills of $11.93333 and $12.08333, while PI costs $11.06 on the same ramp.

$12.04667 to $11.93333, while its IAE falls from 336.6 to 303.2 utilization-seconds. The optimized and floor-based order trajectories differ despite identical zero-breach outcomes, as shown in Figure 14.4. Removing only the linear price term changes one of 720 rounded ramp decisions and adds $0.000833 to the bill, so a material price-optimization benefit is not supported.

MPC loses both disturbed comparisons. On the forecast miss, On breaches 11 minutes for $17.68 against Forecast's 5 for $17.43417, and Off's 8 for $17.84917. Its applied count falls below the forecast's 225 s floor at 164 of 960 decisions. The move cap never binds there. On the crowd, Rule and Forecast give 12 breach minutes and 19.671 % drops for $23.65833, against MPC On's 13, 22.492 % and $57.88750. Its cap binds at 410 of 480 decisions. The count cap and absence of a forecast floor are distinct limitations. Neither observation establishes a unique cause of the losses.

With the unchanged suggested challenge settings, replacing Forecast by MPC gives zero breaches but 1.324 % drops and $121.17363, failing both drop and cost clauses. Forecast passes at zero, 0.816 % and $89.48888. No tuning or challenge requirement is changed. Hours integrate billed on-demand and spot counts before transitions. IAE integrates absolute metric error every 5 s, travel sums applied count changes and limit time counts either count bound. Table timings include QP preparation, excluding initialization. Chrome measured 2900 µs over three startup samples at real

time and 172.1 µs over 2884 warm samples at 3600×, with zero fallbacks. Host and warm-up affect timing.

14.6 Verification, Agreement and Evaluation

Three kinds of evidence support the application, and they are kept apart. Code verification covers the ledgers of Equation (14.5), which close to 1.1×10^{-14} relative, the bill against an independent sum of billed servers over the day to 10^{-12}, the boot delay of exactly 180 and 480 s, the rule against the documentation's computations on six declared inputs, Kingman's term at $c = 1$ to 10^{-12} at three utilizations, every scripted event marked once, and the step halving above. The application now runs 53 checks, including the predictive controller. Agreement with published results is limited to the queueing law, where the declared term is within 3.5 % of the exact M/M/4 wait from 80 % utilization upward. The retry amplification under the budget, a largest ratio of 0.10, is the source's figure by construction [9], not an independent agreement. No traffic trace, price or fleet of a real service was fitted, so no agreement with operating data is claimed. Educational evaluation is limited to play tests made while the app was reviewed, in which a first breach appeared after 23 s of wall time in case 1 and the challenge was passed from its hints in three runs, about 1.5 min. No learning study has been made.

14.7 Technology in Practice

What Measures It

The autoscaler of a real service reads a metric, not a server. On the open-source container platform whose rule this chapter copies, the resource metrics pipeline offers only the minimum CPU and memory metrics needed for automatic scaling, and a fuller set needs a second pipeline that serves custom metrics [22]. The autoscaler can scale on such custom metrics, or on an external metric fetched as a single value [14]. The CPU busy fraction is the chapter's u_{cpu}, and a request rate per server is its u_{req}, the Requests option that in case 5 counts the refused demand the CPU reading hides.

> **TECHNOLOGY** **Sensor: A Metrics Pipeline**
>
> On the open-source container platform the reading is the end of a pipeline. A node agent derives each container's CPU use as a rate over a cumulative counter kept by the kernel, over a window it reports with the value, and a cluster add-on collects these readings every 15 s for the autoscaler [22, 23]. The add-on's own documentation warns against using it as an accurate source of resource usage [23]. A starting server can read misleadingly high while it initializes, so by default the autoscaler ignores a new pod's CPU for its first five minutes unless the sample was taken while the pod was ready, and a pod with no metric is set aside and counted conservatively [14]. The chapter's u_{cpu}, a 60 s mean published every 15 s that cannot exceed one, models this pipeline.

What Moves It

Two autoscalers stand between an order and the capacity it buys. The horizontal autoscaler sets the replica count of a workload, and a cluster autoscaler adds machines when pods cannot be placed and removes machines that stay unneeded [24]. Its documentation discourages node autoscalers that act on CPU use alone, because they may add a node with no pods on it or remove one that carries system-critical pods [24]. The chapter merges the two layers into one fleet of servers that each boot in 180 s, and Figure 14.5 draws the loop on this hardware.

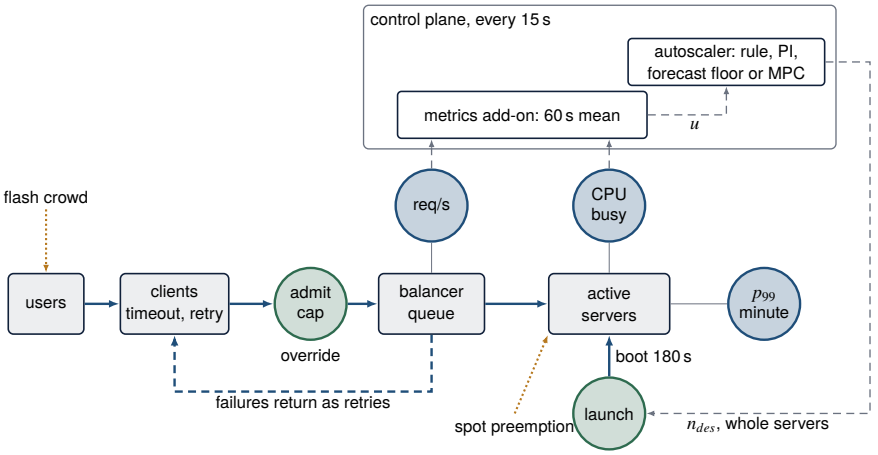


Figure 14.5 The autoscaling loop drawn on the hardware of an open-source container platform, with the chapter's symbols. Sensors are drawn blue, final elements green and disturbances dotted. The metrics add-on turns the CPU busy fraction or the request rate into the published u, the autoscaler orders whole servers, and each order becomes capacity only after a launch and a boot. Notice that the p_{99} the objective is set on is an input of no autoscaler, and that admission control acts at the balancer, outside the count loop.

> **TECHNOLOGY** **Final Element: A Machine Launch**
>
> A replica count is a command, and capacity arrives only when a machine exists to run it. A cluster autoscaler checks every 10 s for pods that the scheduler could not place and asks the provider for a node from a group of identical machines. How long the node takes to appear depends almost entirely on the provider and the speed of provisioning. The autoscaler expects it within 15 minutes, after which it stops counting on it, may try another group and removes a node that never registered. Removal is slow on purpose: a node must stay unneeded for 10 minutes, and non-empty nodes are removed one at a time [24]. The chapter folds all of this into one 180 s boot and a 30 s drain (Table 14.2).

Where the Controller Runs

The horizontal autoscaler is a control loop that runs intermittently inside the cluster's controller manager, by default every 15 s [14], which is the control period of every automatic mode in this chapter. Each autoscaler object holds the structure of the platform's rule: the target, a tolerance of 0.1 by default, a scale-down stabilization window of 300 s, and scaling policies that cap the change in one period [14]. The documentation names the problem the window treats, a replica count that keeps fluctuating with the metric, which it calls thrashing or flapping and likens to hysteresis [14]. The cluster autoscaler under it checks every 10 s [24].

> **IN PRACTICE** **Load Test to the Breaking Point**
>
> Site-reliability practice calls a load test of a server's capacity limits, and of its failure mode under overload, the most important exercise for preventing overload. Components are loaded until they break, both by a gradual ramp and by an impulse, because caching makes the two differ, and are then returned to nominal load to see whether they leave the degraded mode without help. The breaking point feeds capacity planning: at 5000 queries per second per cluster and a peak of 19 000, about six clusters run the service with two spare [9]. Case 1 walks the knee by hand, and case 7 shows a service that does not leave overload when its trigger ends.

What Is Modernizing It

Three developments touch this loop. Scale-up and scale-down behaviors are configured separately, each with its own stabilization window and policies, and a tolerance per direction has been stable since version 1.37 [14], so the hold of case 3 can be set for each direction. Scaling to and from zero replicas is now on by default as a beta feature, for object and external metrics only, because CPU can be measured only on running pods [14]. Warm capacity is kept with low-priority placeholder pods that hold resources until real pods preempt them, which forces the cluster autoscaler to add a node before it is needed [24]. That is the role of the warm pool in case 8 and in the last row of Table 14.6.

14.8 Limitations

The model is a fluid queue with a declared latency term, so it has no individual requests, no burst within a second and no service-time distribution beyond V, and its p_{99} is a multiple of the mean. Bursty traffic would move the knee to the left. With one tier and no downstream queue, a slow database appears only as a lower capacity. The rule's readiness handling is simplified, and the node autoscaler under the servers, placement across zones and startup bursts are absent. The retry storm depends on $r = 1$, on three attempts and on a quota that caps the fleet, and its length on how far the evening traffic falls, so its two hours belong to this case and are not a prediction for a real service. Every price, boot time and objective is a teaching value, the comparisons are seeded teaching results that say nothing about the general optimality of a policy, and a simulated pass is not evidence that a real service would meet a contractual service level.

> **OPEN CHALLENGE** **Predicting the Next Metastable Failure**
>
> Many production systems run in a vulnerable state for efficiency, and a systematic way to build systems robust against unknown metastable failures remains an open problem. Detecting a vulnerable state is difficult because of the sheer size of the systems, and predicting a failure is harder still, since it needs the state, the trigger and its intensity. The defenses in use are often reactions to earlier failures and are not applied systematically. One proposal is to watch a characteristic metric, such as queueing delay or the timeout rate, that the trigger disturbs and that recovers only after the failure ends [10]. In this chapter the share timing out is a truth column that no controller reads.

The MPC fixes web capacity and cold delay even when service time or boot settings change. It omits backlog, retries, immediate scale-in, warm joins, GPU dynamics, drain and warm bills, and spot prices. Measured counts re-anchor its prediction but do not remove these mismatches. Its finite-horizon count-tracking objective neither enforces the latency objective nor minimizes the measured bill.

> **OPEN CHALLENGE** **Forecasting the Crowd**
>
> Existing autoscalers mostly predict the workload from its own past. A survey proposes using social media and other real-time channels to foresee bursts, and notes that a general-purpose predictor of this kind is difficult to design [3]. Progress would be a forecast that sees a crowd one boot ahead. The forecast floor of this chapter knows only the declared daily profile, and in case 4 it breaches as often as the rule (Table 14.7).

14.9 Conclusion

This chapter has treated an autoscaled web service as a plant with an integer actuator, a boot dead time, a saturating metric, a latency knee and a positive feedback through its clients. Queueing doubles the unloaded latency at 83.2 % utilization and the 300 ms objective falls at 87.5 %, within 3.5 % of the exact queue near the knee. The published rule at a 70 % target follows a morning ramp without a breach, yet leaves 14 breach minutes in a fourfold crowd whose servers arrive after it has gone. A 300 s cooldown removes 29 reversals for $0.92. Admission control keeps the objective by refusing work, and on the Requests metric it drops 4.8 % of the crowd window's requests instead of 23.7 %. With the fleet capped, retries hold the service in overload for two hours at 44 % of its capacity, and a 10 % retry budget removes every storm minute on ten of ten seeds. On the graded day the metric and the override do the work, and the forecast saves $0.63. The predictive option costs $11.93 against PI's $11.06 on the three-hour ramp, and loses to the forecast on both disturbed comparisons. The model is a fluid with a declared percentile. Further work would add a downstream dependency with its own queue, request deadlines that cancel abandoned work, a measured traffic trace and a fitted tail. The simulator shows that the capacity a feedback loop orders must be judged against the time it takes to arrive.

KEY POINTS

- Servers ordered by feedback arrive one boot late, so a rule that rides the morning ramp leaves 14 breach minutes in a fourfold crowd.
- The rule is proportional to the measured load, with a deadband, a hold, an integer output and a dead time equal to the boot.
- A 300 s cooldown removes 29 reversals for $0.92.
- Admission control keeps the objective by refusing work.
- With the fleet capped, retries hold the service in overload for two hours, and a 10 % retry budget removes every storm minute on ten of ten seeds.

Where else this applies. The scale-in hold is a minimum dwell of the same kind as the minimum run and rest times of the rack compressors in Chapter 12 (Table 12.4).

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] T. Lorido-Botran, J. Miguel-Alonso, and J. A. Lozano, "A review of auto-scaling techniques for elastic applications in cloud environments," *Journal of Grid Computing*, vol. 12, no. 4, pp. 559–592, 2014.

[3] C. Qu, R. N. Calheiros, and R. Buyya, "Auto-scaling web applications in clouds: a taxonomy and survey," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–33, 2018.

[4] J. L. Hellerstein, Y. Diao, S. Parekh, and D. M. Tilbury, *Feedback Control of Computing Systems*. Wiley, 2004.

[5] A. Gandhi, M. Harchol-Balter, R. Raghunathan, and M. A. Kozuch, "AutoScale: dynamic, robust capacity management for multi-tier data centers," *ACM Transactions on Computer Systems*, vol. 30, no. 4, pp. 1–26, 2012.

[6] J. Dean and L. A. Barroso, "The tail at scale," *Communications of the ACM*, vol. 56, no. 2, pp. 74–80, 2013.

[7] J. F. C. Kingman, "The single server queue in heavy traffic," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 57, no. 4, pp. 902–904, 1961.

[8] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems*. Cambridge University Press, 2013.

[9] B. Beyer, C. Jones, J. Petoff, and N. R. Murphy, Eds., *Site Reliability Engineering*. O'Reilly Media, 2016.

[10] N. Bronson, A. Aghayev, A. Charapko, and T. Zhu, "Metastable failures in distributed systems," in *Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS '21)*, 2021, pp. 221–227.

[11] T. F. Abdelzaher, K. G. Shin, and N. Bhatti, "Performance guarantees for Web server end-systems: a control-theoretical approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 1, pp. 80–96, 2002.

[12] H. C. Lim, S. Babu, J. S. Chase, and S. S. Parekh, "Automated control in cloud computing: challenges and opportunities," in *Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds*, 2009, pp. 13–18.

[13] R. Isaacs, P. Alvaro, R. Majumdar, K.-K. Muniswamy-Reddy, M. Salamati, and S. Soudjani, "Analyzing metastable failures," in *Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS '25)*, 2025, pp. 172–178.

[14] The Kubernetes Authors, "Horizontal pod autoscaling," Kubernetes documentation, https://kubernetes.io/docs/concepts/workloads/autoscaling/horizontal-pod-autoscale/, 2026, page last modified August 3, 2026; accessed October 26, 2026.

[15] H. Sakasegawa, "An approximation formula $L_q \simeq \alpha \cdot \rho^{\beta}/(1-\rho)$," *Annals of the Institute of Statistical Mathematics*, vol. 29, no. 1, pp. 67–75, 1977.

[16] N. Roy, A. Dubey, and A. Gokhale, "Efficient autoscaling in the cloud using predictive models for workload forecasting," in *2011 IEEE 4th International Conference on Cloud Computing*, 2011, pp. 500–507.

[17] P. Mell and T. Grance, "The NIST definition of cloud computing," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication 800-145, 2011.

[18] B. Burns, B. Grant, D. Oppenheimer, E. Brewer, and J. Wilkes, "Borg, Omega, and Kubernetes," *Communications of the ACM*, vol. 59, no. 5, pp. 50–57, 2016.

[19] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[20] 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.

[21] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free model-predictive control," *AIChE Journal*, vol. 49, no. 2, pp. 426–437, 2003.

[22] The Kubernetes Authors, "Resource metrics pipeline," Kubernetes documentation, https://kubernetes.io/docs/tasks/debug/debug-cluster/resource-metrics-pipeline/, 2026, page last modified July 1, 2026; accessed November 22, 2026.

[23] "Kubernetes metrics server," Project README, https://github.com/kubernetes-sigs/metrics-server, 2026, revision of November 2, 2026; accessed November 22, 2026.

[24] "Cluster autoscaler FAQ," Kubernetes autoscaler project, https://github.com/kubernetes/autoscaler/blob/master/cluster-autoscaler/FAQ.md, 2026, revision of September 26, 2026; accessed November 22, 2026.

14.10 Exercises

14.1 ○ Concept check · 2 min

In case 4 the traffic quadruples at 20:00 and the fleet reaches its 200-server quota only at 20:16, after the crowd has passed. What limits how fast the platform's rule on the CPU metric grows the fleet?

(a) The scale-in cooldown holds the fleet for 300 s after every order, so the orders are one window apart

(b) The reading is pinned at 100 %, so a decision asks for at most $1/0.7$ times the ready servers, and the next step waits for a boot

(c) The rule counts booting servers as busy, so it orders twice for the same load and then has to wait for the excess to drain

(d) The metric is a 60 s mean published every 15 s, so the rule cannot see the crowd until the crowd is over

14.2 ○ Concept check · 2 min

On the morning ramp of case 2 at an 80 % target, the PI at its defaults costs \$14.50 and breaches 70 minutes, while the rule costs \$15.41 and breaches 18. Why does the PI breach so much more?

(a) The CPU metric saturates, so the PI cannot see how far above the target the servers run

(b) The PI counts booting servers as busy, so it waits for every boot before it orders again

(c) A PI follows a ramp with a standing error, and its reading averages 85.3 %, close to the 87.5 % where the objective falls

(d) Back-calculation against the cooldown freezes the integral for the whole of the ramp

14.3 Concept check · 1 min

In case 7 at 21:00 the users send 2629 requests per second, 44 % of what the 60-server quota can serve, yet the minute p_{99} is 91.6 s. Which statement explains it?

(a) The 60-server quota leaves too little capacity for the evening users

(b) The boot delay keeps the servers ordered during the crowd from arriving

(c) Admission control refuses most requests, and every refusal is counted as a breach

(d) Failures cause retries and retries cause failures, so the offered load stays near three times the users' and the queue stays full

14.4 Theory and calculation · 6 min

Use Equation (14.3) with the shipped values $c = 4$, $T_{svc} = 40$ ms, $V = 1$ and $c_{99} = 3$, with no backlog.

(a) Find p_{99} at a utilization of 80 %. *Checkpoint:* 213 ms.

(b) Find p_{99} at 85 %, and say why admission control with an 85 % cap keeps every minute under the 300 ms objective once the backlog has drained. *Checkpoint:* 261 ms.

(c) With one worker per server ($c = 1$), find the utilization at which queueing doubles the unloaded latency and the one at which p_{99} reaches 300 ms. *Checkpoint:* 50 % and 60 %.

14.5 Theory and calculation · 7 min

The platform's rule (Equation (14.6)) runs at a 70 % target with the 10 % tolerance.

(a) Twenty ready servers read 90 % and none is booting. Find the desired count. *Checkpoint:* 26.

(b) Twenty ready servers and four booting read 90 %, then 100 %. Find the recount ϕ' and the action in each case. *Checkpoint:* 1.071, no action; 1.190, 29 servers.

(c) In case 4 the CPU reading is pinned at 100 % from 36 ready servers. List the counts the rule can reach, one boot at a time, until the 200-server quota, and estimate when the last step serves if the first full order goes out at 20:01. *Checkpoint:* 52, 75, 108, 155, 200; about 20:16.

14.6 Theory and calculation · 6 min

In case 7 the fleet is capped at 60 servers of 100 req/s, and at 21:00 the users send 2629 req/s. Assume that in the overloaded state almost every attempt fails, that a failure is retried with probability r and that a request makes at most three attempts.

(a) Find the offered load in the overloaded state at $r = 1$, and compare it with the capacity. *Checkpoint:* 7887 req/s against 6000 req/s.

(b) Find the largest offered load with the 10 % retry budget on, as a share of the capacity. *Checkpoint:* 2892 req/s, 48 %.

(c) Find the share r below which the overloaded state cannot hold itself at this traffic, even with every attempt failing. *Checkpoint:* 0.74.

14.7 App challenge · 8 min

Load Scenario 2, *The platform rule and the morning ramp*, at the default seed and change nothing. Run past 10:00 (4 h) at speed 600 (about 24 s), pause, export the CSV file and read columns `breach_min` and `cost_usd` in the row whose `clock_h` reads 10. Reload the scenario, set the Utilization target to 80 % before Run, and repeat. Reload once more, set the Autoscaler to PI and the target to 80 % before Run, and repeat.

(a) Give the breach minutes and the bill of the rule at 70 and at 80 %. *Checkpoint:* 0 and $17.42; 18 and $15.41.

(b) Give the same for the PI at 80 %, and the clock time of the last breach minute in both 80 % runs. *Checkpoint:* 70 and $14.50; 08:37.

(c) What does the rule's higher target buy, and what does it cost? *Checkpoint:* $2.01 for 18 breach minutes.

14.8 App challenge · 8 min

Load Scenario 4, *The flash crowd*, at the default seed and change nothing. Run past 21:45 (2 h) at speed 600 (about 12 s), pause, export the CSV file and read `breach_min`, `dropped_pct` and `cost_usd` in the row whose `clock_h` reads 21.75, and the largest value of `servers_active`. Repeat with Scenario 5, *Admission control: refuse at once, answer the rest*, as loaded, and again with Scenario 5 after setting the Scaling metric to Requests before Run.

(a) Give the four values for case 4. *Checkpoint:* 14, 23.7 %, $20.37, 200 servers.

(b) Give them for case 5 on the CPU metric. *Checkpoint:* 0, 14.0 %, $15.81, 73 servers.

(c) Give them for case 5 on the Requests metric, and explain why the fleet peaks so differently on the two metrics. *Checkpoint:* 0, 4.8 %, $21.37, 200 servers.

14.9 App challenge · 8 min

Load Scenario 7, *The retry storm*, at the default seed and change nothing. Run past 23:45 (4 h) at speed 600 (about 24 s), pause, read the Retries over users tile, export the CSV file and read `breach_min`, `dropped_pct` and `cost_usd` in the row whose `clock_h` reads 23.75. Reload the scenario, switch Retry budget (10 %) on before Run, and repeat.

(a) Give the minutes with more retries than users, the breach minutes, the dropped share and the bill without the budget, and the clock time at which `breach_min` last increases. *Checkpoint:* 129, 130, 52.2 %, $36.39; 22:10.

(b) Give the same with the budget, and its largest retry ratio. *Checkpoint:* 0, 13, 15.0 %, $27.19; 0.10; 20:13.

(c) The crowd ends at 20:12. Explain why the first run recovers two hours later and the second at once. *Checkpoint:* retries hold the offered load above capacity.

14.10 Application area · 4 min

A team serves a language model on GPU servers like those of case 8: each boots in 8 min, and a fourfold crowd lasts twelve minutes. The manager asks for an autoscaler setting that meets the 2 s objective through such a crowd without refusing any request. Using the chapter's measurements, say what to answer and what the real choice is.

14.11 Application area · 3 min

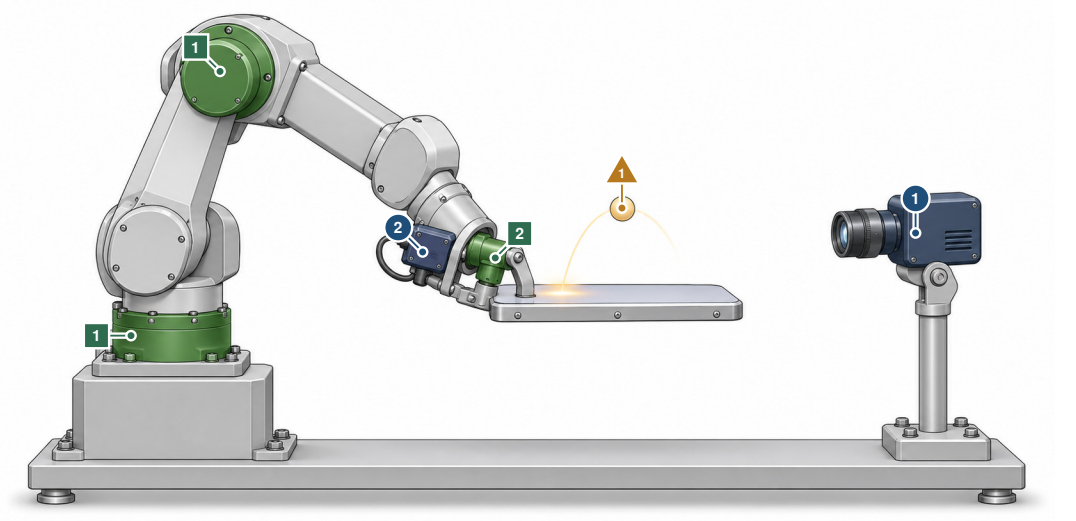
An operator reads the simulator's knee at 83.2 % and its objective at 87.5 % and proposes to run a production service at an 80 % target. Name two assumptions of the model that would move the real knee to the left, and one that makes the simulated retry storm longer than a service with deadlines would see.

14.12 Review the agent · 5 min

An AI assistant advises on the graded day: "The MPC option includes the server price in its objective, so it minimizes the bill. Keep the suggested settings (Requests metric, admission control, 75 % target) and replace the forecast with MPC, and the day will pass for less than the forecast's $89.49." Find the flaw and say how you would check the claim.

CHAPTER 15

A Juggling Robot Arm Whose Paddle Must Be Slowing at Impact



The controller moves 1 the arm's joint drives, which move the paddle up and down
2 the wrist drive that tilts the paddle face

The controller reads 1 the camera that sees the ball, late 2 the impact switch at the paddle mount

What disturbs it 1 a different ball, whose restitution changes

The arm meets the ball while the paddle is rising but already slowing, so notice that the camera on its stand is not needed to keep the ball up: the blind stroke reads no sensor, and only the mirror and tilt laws read the camera. Illustration, not to scale.

IN PLAIN TERMS Keeping a Ball Bouncing on a Paddle

The main idea. A robot arm holds a paddle under a rubber ball and keeps it bouncing to the same height. Each bounce is a short flight between two hits, so the paddle is judged once a bounce, not all the time.

The challenge. The paddle must meet the ball while it moves up but is already slowing. Then a ball that comes back late gets a softer push and falls back into step, with no camera at all. A camera whose pictures come too late cannot make up the lost time.

Why it matters. Walking robots and many machines that repeat one motion are kept steady in the same way, one cycle at a time.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 15.2, the fact in Section 15.4, the two try-its and the fact in Section 15.6, and the key points
Study	50 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the results (Section 15.6), Technology in Practice (Section 15.7), Limitations (Section 15.8) and the open challenge

The control question A ball kept bouncing on a paddle is a hybrid system, and what the juggler holds is a periodic orbit rather than a height. When does a paddle stroke keep the ball up with no sensor at all, and what can feedback through an impact switch or a late camera add to it?

Learning objectives

- Derive the paddle speed at impact that holds a period-1 bounce, and the linearized impact map whose determinant is e^2.
- Compute the window of paddle accelerations at impact in which the orbit is stable, and place a sinusoidal stroke in it.
- Measure a ball's restitution from its bounces on a still paddle, and explain why infinitely many impacts end in finite time.
- Compare the open-loop sinusoid, the once-per-bounce PI law and the mirror law through a change of ball.
- Decide when a sensor's latency leaves a feedback law no orbit to hold.

Prerequisites Sections 1.4.7 (hybrid automata and Zeno solutions), 1.4.9 (maps from one cycle to the next), 1.6.1 (the engine and its seed) and 1.6.4 (scenarios and the challenge). Chapter 12 shows another plant whose switching is part of its dynamics.

The app https://apmonitor.com/apps/paddle-juggler/

See also Volume 7, Chapter 9 for a plant that is self-stable only inside a band of an operating variable. Volume 20, Chapter 11 for the same six-axis arm striking a ball, and Volume 20, Chapter 10 for an arm acting on late camera reports.

15.1 Introduction

Dropped from 50 cm onto a paddle that does not move, a 40 g rubber ball rises to 33, 22, 14 and 9 cm and lies still on the face after 29 impacts in 2.98 s. To send it back to 50 cm every time, the paddle has to meet it while moving up at 0.348 m/s, the speed that replaces what a restitution of 0.80 takes away. A stroke of 6.3 cm each way at 1.655 Hz does that with no sensor at all and holds the apex at 0.502 m with a spread of 1.6 cm over a 30 s run, because the paddle is decelerating at 5.9 m/s^2 when the two meet. A stroke of 3 cm at the same frequency lets the ball die on the paddle, and raised to 11 cm while it runs the stroke throws the ball anywhere from 0.18 to 0.98 m. Each of these runs was made on the simulator of this chapter, a ball juggled by the six-axis arm of a simulated table-tennis robot, and this chapter explains why they end so differently.

The juggler unsettles three beliefs that a first control course tends to leave in place. The first is that the object of control is a setpoint. Here the target is a whole motion repeated every 0.6 s, and its stability is decided between one impact and the next, by a map whose eigenvalues no step response shows. The second is that stability needs a measurement. A paddle that is slowing down when it meets the ball corrects a late bounce with a smaller push and an early one with a larger push, so the orbit pulls itself back with nothing measuring the ball, and with the camera switched off the same stroke carries the ball through a change to a restitution of 0.72 without a drop at six seeds of six. The third is that more gain makes up for a slow sensor. Through a camera 120 ms late, every mirror-law gain from 0.25 to 2.5 lets a 0.20 m bounce die in under 2.4 s, while a blind 3 cm sinusoid at 2.5 Hz keeps it up.

A ball on a paddle belongs to a broad class, mechanical systems whose periodic motions contain impacts, studied as systems with jumps whose solutions mix flows and resets [1, 2], the hybrid automata of Section 1.4.7. A walking machine strikes the ground once a step and its gait is a periodic orbit read on a return map [3, 4], and a robot that throws, bats or rolls an object it never grasps is a manipulator of the same kind [5]. The bouncing ball on a vibrating table is the simplest member, and its period doublings and chaos are a standard example of nonlinear dynamics [6–8]. A press stroke, a cam-driven packaging line and a batch corrected after every run, as in Chapter 9, share its structure, a cycle whose stability is a property of the cycle and a controller whose sample is an event.

However, the robots that juggle are each described around one platform, a linear actuator, a planar juggler or a vision-guided arm [9–12], and the analysis of the impact map is usually done for an ideal paddle that follows its waveform exactly [6, 13]. A student cannot see, on one machine, the map's eigenvalues beside an arm whose drives have limits, a camera that is late and a ball that changes. This chapter shows how much the stability window, the ball's restitution, a sensor's latency and a once-per-bounce controller each decide when a full-sized arm keeps a ball on an orbit, using a browser simulator that logs every bounce with the paddle's speed and acceleration at impact. The impact law, the period-1 orbit, the window of decelerating impacts, the mirror law and blind juggling are all established [6, 10, 12, 14]. The chapter writes the linearized impact map in closed form, with a determinant of e^2 whatever the paddle's waveform, checks the application's numerical Jacobian against it to 6.4×10^{-9}, and measures the same eigenvalues on the running engine with the arm in the loop. It also shows that a smooth periodic stroke always has a slowing branch, so the speeding-up side of the window cannot be held, and that near both edges the bounce-to-bounce spread of the restitution, not the eigenvalue, decides the outcome. The simulator is a 20-state model in which every impact is located inside its step. It drives the collection's six-axis arm module, which moved into the shared framework without changing a bit of its first user's runs, and a sensor boundary gives the laws the impact times and the paddle's own encoders but never the ball.

Section 15.2 traces juggling machines and the bouncing-ball map. Section 15.3 gives the ball, the impact, the arm and the sensors as simulated. Section 15.4 derives the impact map and its window of stable juggling and measures both with the arm in the loop. Section 15.5 gives the sinusoid, the once-per-bounce law and the mirror law as coded. Section 15.6 reports the nine cases and the challenge as measured, each case with the briefing and completion rule of Section 1.6.4. Section 15.7 describes the real impact sensors, arms and controllers, Section 15.8 states the limitations and Section 15.9 draws the conclusions.

15.2 Bouncing Balls, Return Maps and Juggling Machines

The physics came first as a problem in vibration. A ball bouncing on a table driven sinusoidally was analyzed in 1982 as a map from one impact to the next, under the simplification that the table's travel is small against the ball's flight, and the map was found to pass through a sequence of period doublings into chaotic motion as the driving amplitude grows [6]. The same system entered the teaching laboratory when a tabletop experiment reproduced that route to chaos in a physics teaching journal [7], and it serves as a worked example in the standard text on Poincaré maps and their bifurcations [8]. The mechanics of the impact itself, Newton's restitution and the treatment of instantaneous velocity jumps in rigid bodies, belongs to the theory of nonsmooth mechanical systems [2].

Robots took up juggling at the end of the 1980s as a laboratory for manipulation through intermittent contact. A one-joint robot that batted a puck sliding on an inclined plane was controlled by a family of mirror algorithms, in which the robot's motion is a distorted reflection of the puck's [14], and the same group carried the mirror law to planned juggling and catching [9]. In 1993 a vision-guided arm juggled two balls in space with one paddle [15], while a paddle moved on a fixed periodic trajectory was shown to juggle a ball open loop, with the stability coming from the paddle's motion at impact rather than from feedback on the ball [10]. A wider program of dynamic manipulation without grasping, in which a robot moves an object through the object's own dynamics, followed later in the decade [5].

The question of how little sensing a juggler needs was answered next. Sensorless stabilization of bounce juggling by a periodic paddle motion was analyzed and demonstrated [13], and a planar juggler that sensed only the instants of its impacts used them as feedback to strengthen the stability of its rhythm [11]. A robot that never measures the ball at all, a slightly concave paddle on a linear actuator moving on a fixed periodic trajectory, kept a ball bouncing in three dimensions in 2012 [12].

Table 15.1 Signals by Role. Ranges Are Those the Interface Offers

Symbol	Role	Default	Range
h	CV, apex of each flight	—	—
h^*	SP, target apex	0.50 m	0.20 to 0.70 m
x	CV of the tilt law, ball across the face	0	SP 0
z_p	MV, paddle height (Manual)	0	±0.15 m
β	MV, face tilt (Manual, tilt law)	0	±5°
$q_{c,i}$	MV, six joint commands	per law	joint ranges
$x, z, \dot{x}, \dot{z}$	states, each of two balls	per case	—
$q_i, \dot{q}_i$	states, six joints and rates	juggling pose	ranges
$t_k, p_k, \dot{p}_k$	measured, impact switch and encoders	each impact	—
$\hat{x}, \hat{z}$	measured, camera frame	30 Hz, 40 ms	10 to 120 Hz
e_k	hidden, restitution of impact k	$e_0 = 0.80$	—
Δe	DV, a different ball	0	−0.20 to +0.10
β_m	DV, face mounted crooked	0°	±3°
—	DV, camera dropout, 0.5 s	—	event
—	DV, flick across, 5 cm/s	—	event
n	output, consecutive bounces	0	—

HISTORY 1989: A Robot Bats a Puck Up a Slope

In May 1989 a robot that juggled by batting was presented at a robotics conference in Arizona [14]. Its puck slid on an inclined plane, and a bar with a billiard cushion, turning in that plane, struck the puck again and again. An oscillator inside the puck and a grid buried in the plane worked like a large digitizing tablet that gave the puck's position. An observer estimated the puck's velocity and passed it to the planning node at 1 kHz, where the mirror algorithm turned the puck's motion into the bar's reference, and the motor controller issued torque commands at 2 kHz. The controller was designed so that the juggling pattern was an attracting periodic orbit of robot and puck together [14]. The mirror law of Equation (15.13) reads a camera instead of a tablet.

Walking machines rest on the same mathematics. A passive biped walks down a slope on a stable gait with no actuation, its stability read from the step-to-step map [3], and the method of Poincaré sections, extended to systems with impulse effects, was used to prove asymptotically stable walking under feedback [4]. A general framework for hybrid systems with flows and jumps counts the bouncing ball among its examples [1]. The simulator gathers these pieces on one arm, so that the map, the drives and the sensors can be changed one at a time.

15.3 Ball, Paddle and Arm as Simulated

The simulated rig is the six-axis arm of the table-tennis robot app (https://apmonitor.com/apps/table-tennis-robot/), set on a pedestal and holding a paddle face up under a ball, as shown in Figure 15.1. The two applications are one robot with two jobs. The table-tennis robot strikes a ball to send it somewhere, and this one keeps a ball on one orbit. Volume 20, Chapter 11 describes the arm's kinematics and servos from the striking side. The frame has x across the paddle, along the arm's reach, and z up, both measured from the paddle's face at the middle of its stroke, which is the juggling plane. Heights a student reads, the apex and the target, are the ball's lowest point above that plane. The roles of the signals are listed in Table 15.1. Every constant of the ball, the camera and the arm and every threshold is a teaching value, and the application's model dialog and Instructions page say so.

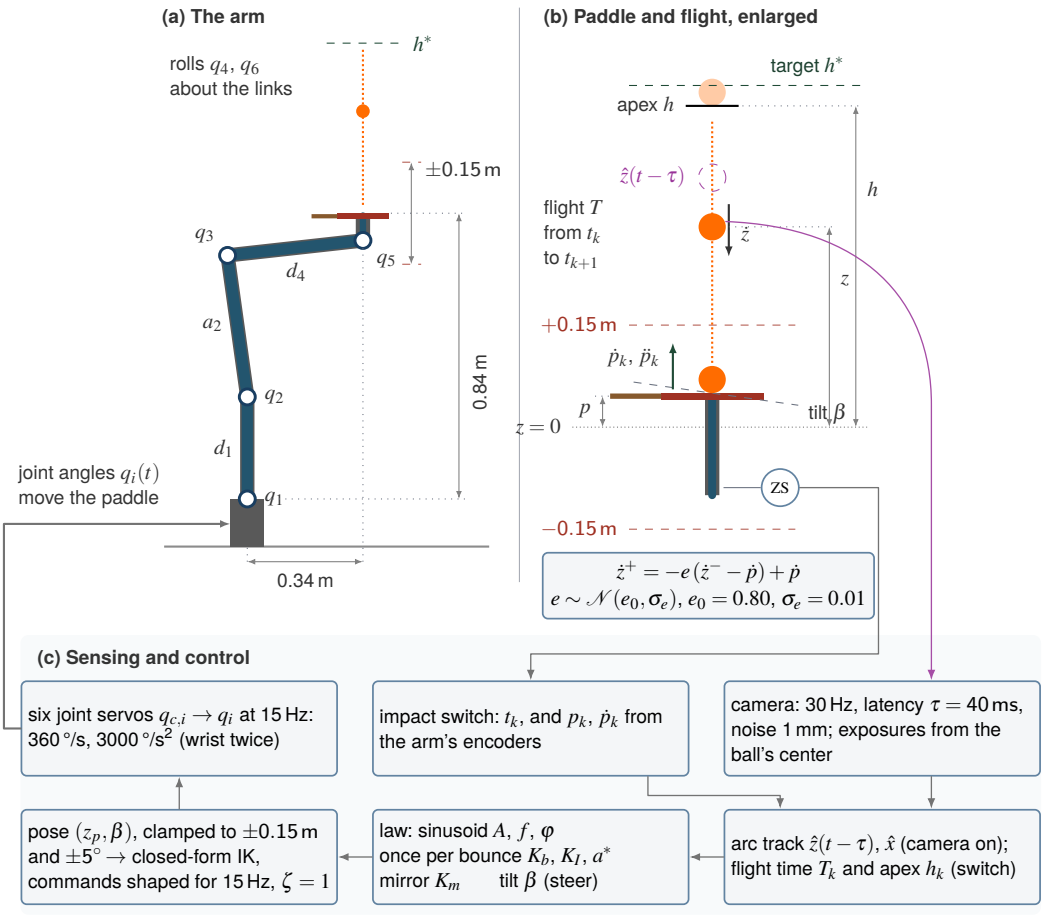


Figure 15.1 The juggling rig, panels (a) and (b) in the layout of the application's Rig view. (a) The six-axis arm on its pedestal at the juggling pose, drawn to scale from its forward kinematics, with the chain lengths d_1, a_2, d_4, the shoulder, elbow and wrist pitch joints q_2, q_3, q_5, the base turn q_1, the paddle's ± 0.15 m stroke and the ball's flight to the target apex h^*. (b) The paddle and the flight enlarged. The ball at height z and speed $\dot{z}$ meets the face at height p moving at $\dot{p}_k$ and accelerating at $\ddot{p}_k$, and the impact law turns $\dot{z}^-$ into $\dot{z}^+$ with a restitution drawn about e_0. Each flight of duration T rises to an apex h against the target h^*, the face can be tilted by β, and the camera sees the ball as $\hat{z}(t-\tau)$, a latency late. The impact switch ZS rings at each impact. (c) The camera and the switch feed the track, a law turns it into a pose, closed-form inverse kinematics and shaped commands turn the pose into joint commands, and six joint servos move the arm. Panels (b) and (c) are not to scale.

15.3.1 Flight

The ball is a solid rubber sphere of radius $r = 20$ mm and mass $m = 40$ g, and between impacts it flies under gravity and a quadratic drag, as shown in Equation (15.1).

$$\ddot{x} = -k_D\,|\mathbf{v}|\,\dot{x}, \qquad \ddot{z} = -g - k_D\,|\mathbf{v}|\,\dot{z}$$
$$k_D = \frac{\rho\, C_D\, \pi r^2}{2m} \tag{15.1}$$

where $\mathbf{v} = (\dot{x}, \dot{z})$, $g = 9.81\,\mathrm{m/s^2}$, $\rho = 1.2\,\mathrm{kg/m^3}$ and $C_D = 0.40$, which give $k_D = 0.0075\,\mathrm{m^{-1}}$. At the 3.13 m/s with which a 0.50 m bounce leaves the paddle, the drag is 0.75 % of gravity. The ball is

dense for a measured reason. A table-tennis ball of 2.7 g, with $k_D = 0.126\,\mathrm{m}^{-1}$, feels about an eighth of its weight in drag at that speed, loses 5.8 % of a 0.50 m rise and returns to the paddle with 88.8 % of the energy it left with. On a still paddle its first six apex ratios read e^2 between 0.56 and 5.7 % low, so the geometric series of case 2 would measure the air as much as the impact. The rubber ball reads e^2 between 0.04 and 0.39 % low.

15.3.2 The Impact, Located Inside Its Step

The face is a flat disc of radius 75 mm with unit normal $\mathbf{n}$, moving at the velocity $\mathbf{u}$ of the paddle's center. An impact keeps the ball's velocity along the face and reverses its normal velocity relative to the face with Newton's restitution [2], as shown in Equation (15.2).

$$\begin{aligned} v_n^{+} &= -e_k\,(v_n^{-} - u_n) + u_n, \qquad v_t^{+} = v_t^{-} \\ e_k &= \min\bigl(0.98,\ \max(0.05,\ e_0 + \Delta e + \sigma_e\,\xi_k)\bigr) \end{aligned} \tag{15.2}$$

where $v_n = \mathbf{v}\cdot\mathbf{n}$ and $u_n = \mathbf{u}\cdot\mathbf{n}$, ξ_k is a standard normal number drawn for each impact from a seeded stream, $e_0 = 0.80$ and $\sigma_e = 0.01$. With a level face the law is $\dot{z}^{+} = -e(\dot{z}^{-} - \dot{p}) + \dot{p}$ in the vertical, and a tilted face sends part of the impulse across, which is how a tilt steers the ball. The paddle's velocity is the arm's tool velocity from its Jacobian. The ball and the face are tested after every step. When the gap between the ball's surface and the face changes sign inside a step of length Δt, the fraction $f = s_0/(s_0 - s_1)$ of the step at which it closes follows from the gaps s_0 and s_1 at the two ends. Ball and face are both interpolated linearly to that fraction, the ball is set on the face, and after Equation (15.2) acts at that instant the ball's flight is integrated over the remaining $(1-f)\Delta t$ with Equation (15.1). A ball that approaches the face slower than 5 mm/s is taken to lie on it, and the juggle is then dropped. A ball whose center crosses the face's plane outside its rim has walked off, and once it falls more than 0.30 m below the juggling plane it is dropped. Either drop latches, stops the run and is reported in the status line.

Over every case as shipped and at half the step, every one of the impacts is located inside the step that holds it, the law holds to 2.7×10^{-15} m/s, and the ball's kinetic energy changes by the paddle's work less the restitution's loss to 10^{-14}. The steps themselves run 0.54 to 1.80 mm into the face before the locator acts, and no logged step leaves the ball in it. Under a paddle shaken at 20 Hz with an amplitude of 1 mm, the steps run 0.889 mm into the face and the smallest logged gap is +0.0001 mm.

15.3.3 The Arm That Holds the Paddle

The paddle is the tool of the collection's servo-level six-axis arm, `shared/src/framework/core/arm.js`. The table-tennis robot wrote the module, and it moved into the shared framework unchanged in arithmetic when this application became its second user. The table-tennis robot's ten scenarios logged bit-identical files before and after the move, shipped and with suggested settings. The juggler commands the arm through the same joint servos and the same command shaping, so the two applications run one arm. The kinematic chain is anthropomorphic and ends in a spherical wrist, as shown in Equation (15.3).

$$\begin{aligned} R &= R_z(q_1)\,R_y(q_2+q_3)\,R_z(q_4)\,R_y(q_5)\,R_z(q_6) \\ \mathbf{n} &= R\,\mathbf{e}_z \\ \mathbf{p}_t &= \mathbf{b} + d_1\mathbf{e}_z + a_2 R_z(q_1) R_y(q_2)\,\mathbf{e}_z \\ &\quad + d_4 R_z(q_1) R_y(q_2{+}q_3)\,\mathbf{e}_z + d_6 R\,\mathbf{e}_z \end{aligned} \tag{15.3}$$

where $\mathbf{p}_t$ is the paddle's center, $d_1 = 0.30\,\mathrm{m}$, $a_2 = 0.42\,\mathrm{m}$, $d_4 = 0.40\,\mathrm{m}$ and $d_6 = 0.08\,\mathrm{m}$, and the base $\mathbf{b}$ stands 0.34 m behind and 0.84 m below the middle of the stroke. That pose was chosen because a face-up paddle is bound by the wrist pitch's $\pm 125^\circ$, and from it every joint stays at least 14.3° inside its range over the ± 0.15 m stroke and $\pm 5^\circ$ of tilt, the wrist pitch at the bottom of the stroke

binding first. At the middle of the stroke the joints sit at $q_2 = -7.9°$, $q_3 = 91.6°$ and $q_5 = -83.7°$, the forearm across and the wrist pointing up. Each joint is a second-order drive with limits, as shown in Equation (15.4).

$$\begin{aligned} \ddot{q}_i &= \mathrm{sat}_{\ddot{q}_{max,i}}\left(\omega_j^2\,(q_{c,i}(t) - q_i) - 2\zeta_j\omega_j\,\dot{q}_i\right) \\ |\dot{q}_i| &\le \dot{q}_{max,i} \end{aligned} \tag{15.4}$$

where $\omega_j = 2\pi f_j$ with $f_j = 15\,\mathrm{Hz}$ and $\zeta_j = 1$ by default, and $q_{c,i}(t)$ is the latest command carried along its own rate between 5 ms updates. Speed and acceleration are capped at 360 °/s and 3000 °/s^2 for the first three axes and at double those values for the wrist. Approaching the end of its range, a drive brakes inside its acceleration limit so as to stop just short of it, and the range is a hard backstop in the plant. With the face level, the arm can accelerate the paddle vertically at 26, 19 and 13 m/s^2 at the bottom, middle and top of the stroke. No case as shipped reaches a range end or a reach limit, and outside the mirror law no drive reaches a speed or acceleration limit. On the stable sinusoid the busiest drive uses 29 % of its speed limit and 43 % of its acceleration limit.

15.3.4 What the Laws Can Know

The impact switch rings at every impact with its time t_k, and the arm's encoders give the paddle's height p_k and speed $\dot{p}_k$ at that instant. It does not say where across the face the ball struck. A blind juggler has nothing else, and two rings give a flight time $T_k = t_{k+1} - t_k$, from which the apex of the flight just ended follows by Equation (15.5).

$$h_k = p_k + \frac{v_k^2}{2g}, \qquad v_k = \frac{p_{k+1} - p_k}{T_k} + \frac{g\,T_k}{2} \tag{15.5}$$

The camera exposes each ball's center at exact multiples of $1/f_{cam}$, with $f_{cam} = 30\,\mathrm{Hz}$, interpolating the ball inside the step that holds the instant, adds 1 mm of Gaussian noise per axis, and delivers each frame $\tau = 40\,\mathrm{ms}$ after its exposure. During a dropout no exposure is delivered. The noise stream draws four numbers at every exposure, whether or not the camera is on, so that a run replays bit for bit. The laws read the camera through an arc track. The frames since the ball's last impact, at most 24, together with the paddle's height at that impact, are fitted by least squares with a ballistic arc in height (gravity known, drag left out) and a straight line across, and until a frame of the new flight arrives the arc leaves the impact at the speed the impact law predicts with $e^* = 0.80$ and the paddle's encoder speed. A law that acts on the camera reads this arc at $t - \tau$, the ball as the latest frame saw it carried along its flight, which the Rig view draws as a ghost ball behind the true one. No controller reads the simulator's truth, which the exported file carries in columns prefixed `truth_`.

15.3.5 Parameters and Numerics

Table 15.2 lists every parameter. The 20 states, each of two balls' position and velocity and the arm's six joint angles and rates, move forward in fixed fourth-order Runge-Kutta steps of 0.5 ms, in which a ball at 3 m/s moves 1.6 mm. The laws and the joint commands run every 5 ms, the camera exposes at its own instants and the historian logs at 100 Hz. Each step executes the laws, integrates, applies the rate clamp and range backstop, resolves impacts, records exposures and logs, in that order, which is the engine step of Section 1.6.1 with the impact resolved inside it. Halving the step to 0.25 ms moves the still paddle's apex ratios by 0.0002 % and the sinusoid's mean apex over 25 s by less than 0.0001 mm. Each case lets the ball go from rest at $t = 0$, its bottom at the target apex (35 cm in case 7) over the middle of the face, or 0.09 m to either side with two balls, the second let go half a bounce later. The arm starts on the law's reference at $t = 0$, or at the middle of the stroke in Manual, and the sinusoid's clock is set so that the stroke is at φ when the ball first arrives. The restitution and the camera noise come from seeded streams, so a run repeats exactly whether its case is opened first, chosen after another case, or restarted with Reset.

Table 15.2 Model Parameters. A Is Adjustable in the Model Dialog, V Stated in the Model Dialog but Fixed, H Hidden in the Code

Symbol	Meaning	Value	
Ball and impact			
e_0, σ_e	restitution, mean and spread	0.80, 0.01	A
k_D	drag factor (C_D 0.40, ρ 1.2 kg/m^3)	0.0075 m^{-1}	A
r, m	ball radius, mass	20 mm, 40 g	V
r_p	paddle face radius	75 mm	V
—	approach at which the ball rests	5 mm/s	V
—	drop height below the plane	0.30 m	H
Arm (the table-tennis robot's)			
$\dot{q}_{max}$	speed limit, axes 1 to 3 (wrist ×2)	360 °/s	A
$\ddot{q}_{max}$	acceleration limit, axes 1 to 3 (wrist ×2)	3000 °/s^2	A
d_1, a_2, d_4, d_6	chain lengths	0.30, 0.42, 0.40, 0.08 m	H
—	ranges, axes 1 to 6 (±°)	170, 100, 150, 200, 125, 200	H
b	base, behind and below the stroke	0.34, 0.84 m	V
—	stroke and tilt limits	±0.15 m, ±5°	V
f_0, ζ_0	drive tuning the commands assume	15 Hz, 1	V
Sensors			
f_{cam}, τ	camera rate, latency	30 Hz, 40 ms	A
σ_{cam}	camera noise per axis	1 mm	A
—	frames in the arc track	24	H
—	seed (restitution, camera)	1	A
Numerics			
Δt, T_c	integration step, control period	0.5, 5 ms	H
—	historian	100 Hz	H

15.3.6 Assumptions

A rigid ball in one plane. The ball does not spin and stays in the arm's vertical plane. A real ball spins after a tilted or off-center hit and drifts out of the plane, and its spin changes the next bounce.

Instantaneous Newton impacts. The restitution does not depend on the impact speed or on where the ball lands, and the face does not grip the ball. Real restitution falls as impacts get faster, and a rubber face grips and spins a ball, so the bounce-to-bounce spread stands in for all of this [2].

A flat face. A drifting ball is steered back only by tilting the face on the camera's estimate. A slightly concave face, as on the blind juggler of [12], would steer it back with no sensor.

A servo-level arm. Positions and orientations follow exact kinematics, while inertia, torque limits and flexibility are left out, and the arm's capacity does not change with what it carries.

Ideal sensors apart from rate, latency and noise. The switch reports every impact exactly, and the camera has no blur, occlusion or lost frames outside a dropout.

15.4 One Bounce at a Time

15.4.1 The Period-1 Orbit

A bounce that rises h above the point of impact leaves at $v = \sqrt{2gh}$ and lasts $T = 2v/g$. To send it back up at the same speed, the paddle must meet it moving up at the speed that replaces the restitution's loss, as shown in Equation (15.6).

$$v = -e\,(-v - u^*) + u^* \quad \Longrightarrow \quad u^* = v\,\frac{1-e}{1+e} \tag{15.6}$$

For $h = 0.50\,\text{m}$ and $e = 0.80$ this is $v = 3.13\,\text{m/s}$, $T = 0.639\,\text{s}$ (1.57 Hz) and $u^* = 0.348\,\text{m/s}$. The flight time fixes the frequency of any stroke that meets every bounce, and the frequency fixes the apex. At 1.655 Hz, the stroke of most cases, a bounce lasts 0.604 s, rises 0.448 m above its impact and needs the paddle at 0.33 m/s, or 0.48 m/s for a ball whose restitution has fallen to 0.72.

15.4.2 The Map and Its Linearization

Let the paddle repeat with period T, and let τ_k be the offset of impact k from its nominal instant and v_k the ball's speed after it. The next impact is where the ball's parabola meets the paddle again, and the impact law there gives the next speed. This is a map from one bounce to the next, and the period-1 orbit is its fixed point, a cyclic steady state in the sense of Section 1.4.9. Linearizing about the period-1 orbit, with the paddle meeting the ball at u^* and accelerating at $a = \ddot{p}(t_k)$, gives Equation (15.7).

$$\begin{bmatrix} \delta\tau_{k+1} \\ \delta v_{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \dfrac{1+e}{g} \\ (1+e)\,a & e^2 + \dfrac{(1+e)^2 a}{g} \end{bmatrix}}_{J} \begin{bmatrix} \delta\tau_k \\ \delta v_k \end{bmatrix} \tag{15.7}$$

The first row says that a faster launch comes back later, by $T/(u^* + v) = (1+e)/g$ per unit of speed. The second says that a late ball meets a paddle whose speed has changed by $a\,\delta\tau$, and that a ball launched faster by δv also arrives faster by $e\,\delta v$ and leaves the next impact faster by $e^2\delta v$ when the paddle does not accelerate. Only the paddle's speed and acceleration at the instant of impact enter, so J is the same for a sinusoid or any other periodic waveform, at any frequency. Its determinant and trace are Equation (15.8).

$$\det J = e^2, \qquad \operatorname{tr} J = 1 + e^2 + \frac{(1+e)^2}{g}\,a \tag{15.8}$$

The eigenvalues solve $\lambda^2 - \operatorname{tr} J\,\lambda + e^2 = 0$. One reaches $+1$ where $\operatorname{tr} J = 1 + e^2$, at $a = 0$, and one reaches -1 where $\operatorname{tr} J = -(1+e^2)$, so the orbit is stable for paddle accelerations at impact in the window of Equation (15.9) [6, 10].

$$-\frac{2g\,(1+e^2)}{(1+e)^2} < \ddot{p}(t_k) < 0 \tag{15.9}$$

At $e = 0.80$ the window runs from $-9.9311\ \mathrm{m/s^2}$ to 0, and at $e = 0.72$ its lower edge moves to $-10.070\ \mathrm{m/s^2}$. Where $|\operatorname{tr} J| < 2e$, from $a = -g$ to $a = -g(1-e)^2/(1+e)^2$, or $-9.81\ \mathrm{m/s^2}$ to $-0.121\ \mathrm{m/s^2}$, the eigenvalues are complex with modulus exactly e, so an error shrinks by a fifth every bounce whatever the stroke. Near the two edges they are real and one of them grows toward 1, to 0.970 at $-9.9\ \mathrm{m/s^2}$ and 0.949 at $-0.05\ \mathrm{m/s^2}$. Outside the window it is 1.977 at $+2\ \mathrm{m/s^2}$.

> **+ DID YOU KNOW? Never With a Rising Arm**
>
> On a planar juggler that sensed only the times of its impacts, sustained stabilization in the laboratory needed the arms' acceleration at impact tuned negative, and it could never be achieved with a positive one [11]. The window of Equation (15.9) ends at the same zero.

The application computes the same Jacobian numerically, by central differences on its own impact map of an ideal sinusoidal paddle, `paddle-juggler/src/poincare.js`, which finds the period-1 orbit by Newton's method and draws the window on the Map view. Over 26 accelerations from -14 to $+4\ \mathrm{m/s^2}$, at 1.566 and 1.655 Hz, its largest eigenvalue modulus agrees with Equation (15.8) to 6.4×10^{-9}, its determinant reads 0.6400 at every one, and its bisected edges are $-9.9311\ \mathrm{m/s^2}$ and 0, as shown in Figure 15.2a. On a finer grid of 386 accelerations the largest difference, 1.7×10^{-5}, falls at $-9.81\ \mathrm{m/s^2}$, where the two eigenvalues meet on the real axis.

15.4.3 The Window With the Arm in the Loop

The ideal map assumes a paddle that follows its waveform exactly. The engine's paddle is the arm's tool, driven through inverse kinematics and six servos, so the eigenvalues were measured on the

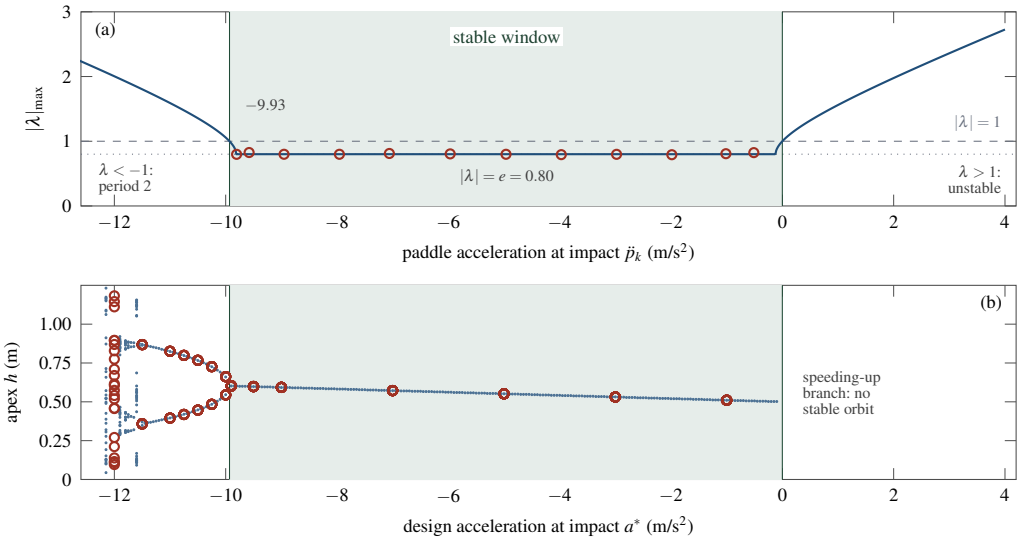


Figure 15.2 The window of stable juggling at $e = 0.80$. (a) The largest eigenvalue modulus of the impact map against the paddle's acceleration at impact, from the application's numerical Jacobian (line), with the window between $-9.93\ \mathrm{m/s^2}$ and 0 shaded, and the eigenvalues measured on the running engine with the arm in the loop (circles), e exact and no drag, from the apex errors after a 2 cm/s kick. (b) The apexes of the last 40 of 300 bounces of the ideal map (dots) and of the engine's flights over the last 15 s of a 40 s run (circles), for strokes at 1.566 Hz designed for each acceleration. Past the lower edge the orbit splits into two apexes, then four, then scatters.

engine as well. For each design acceleration from -0.5 to $-9.7\ \mathrm{m/s^2}$, a sinusoid at 1.566 Hz was run for 40 s with the restitution exact and no drag, the ball was given a 2 cm/s kick after an apex, and the next 15 apex errors were fitted with the second-order recursion $\delta h_{k+1} = c_1\,\delta h_k + c_2\,\delta h_{k-1} + c_0$, whose coefficients are $\mathrm{tr}\,J$ and $-\det J$ for any scalar output of a two-state linear map. The measured eigenvalue moduli lie between 0.79 and 0.83 against the closed form's 0.80, the determinants between 0.63 and 0.68 against 0.64, and the traces within 0.045 of Equation (15.8) at the measured acceleration, as shown in Figure 15.2a. Beyond the edge the arm's orbit splits as the ideal map's does, into apexes of 0.544 and 0.662 m at a design of $-10\ \mathrm{m/s^2}$ against 0.547 and 0.660 m, 0.446 and 0.767 m at -10.5 against 0.447 and 0.766, and 0.395 and 0.825 m at -11 against 0.396 and 0.825, as shown in Figure 15.2b. The ideal map doubles again near $-11.5\ \mathrm{m/s^2}$ and scatters beyond, and at $-12\ \mathrm{m/s^2}$ the engine's apexes range from 0.096 to 1.182 m. The edges hold with the servos in the loop, and at half the step the arm's period-1 and period-2 orbits are unchanged, their accelerations at impact within $0.01\ \mathrm{m/s^2}$.

15.4.4 Why a Slowing Paddle Is Stable, and Why It Must Be

A ball that comes back late meets a paddle that has already slowed. It gets less of a push, leaves lower and comes back sooner, and a ball that comes back early meets a faster paddle and is sent higher and later. Every bounce is pulled back onto the beat, which is the negative feedback hidden in Equation (15.7), and it needs no sensor. Slowing too hard overcorrects, so a late ball comes back early by more than it was late, and the orbit alternates between a high and a low bounce once an eigenvalue passes -1. A riderless bicycle is self-stable in the same way only inside a band of its forward speed (Volume 7, Chapter 9).

🕮 **Pause.** Try Exercise 15.1 before reading on.

Table 15.3 Controller Settings, Defaults and Ranges. Settings Marked adv. Are Shown in the Advanced View

Symbol	Setting	Default	Range
—	law: sinusoid, once per bounce, mirror law or two balls	sinusoid	four
h^*	target apex	0.50 m	0.20–0.70
A, f	stroke amplitude, frequency	6.3 cm, 1.655 Hz	0–15, 1–2.5
φ	stroke phase at first arrival	55°	±180
K_b	bounce-to-bounce gain, (m/s)/m	0.8	0–3
K_I	integral per bounce, (m/s)/m	0.2	0–1.5
a^*	design acceleration at impact	−5 m/s^2	−14 to 4
K_m	mirror gain	1.5	0–2.5
—	camera, tilt law	on, level	off, steer
f_j, ζ_j	joint bandwidth, damping (adv.)	15 Hz, 1	2–30, 0.2–2
z_p, β	manual height and tilt	0, 0°	±15 cm, ±5°
e^*	restitution the laws assume	0.80	fixed
—	mirror target filter	12 Hz, ζ 1	fixed
—	largest commanded impact speed	1.5 m/s	fixed
—	amplitude slew	0.10 m/s	fixed
—	fade after a mode or law change	0.5 s	fixed
—	manual follower	1 m/s, 10 m/s^2	fixed

For a sinusoid $z_p = A\sin\omega t$, the paddle has the speed u^* at two phases of each cycle, one speeding up and one slowing down, with the acceleration at impact $\pm\omega\sqrt{A^2\omega^2 - u^{*2}}$. At 1.655 Hz the window's edges correspond to strokes of 3.17 and 9.71 cm. The speeding-up branch is unstable. A smooth periodic stroke that reaches u^* on its way up reaches it again while slowing, so a ball knocked off the speeding-up orbit settles on the slowing one. With the arm in the loop, the restitution exact and no drag, a stroke designed for +2 m/s^2 at 1.566 Hz is the 4.10 cm stroke also designed for −2 m/s^2, and the ball settles on the slowing branch at −2.01 m/s^2. The planning script that sized this application iterated a paddle on a fresh parabola about each nominal impact, which has no second branch, and its rows (converging from −2 to −8 m/s^2, wandering at 0, lost at +2 and at −10 and below) are reproduced by the application's test of that map.

Pause. Try Exercise 15.2 before reading on.

15.5 The Laws as Coded

Every 5 ms the active law gives the paddle's pose as a function of time near the present, $P(s) = (x_p, z_p, \beta)$, its place across (two balls only), its height and its tilt. The pose is clamped to ±0.15 m in height, ±0.12 m across and ±5° of tilt, turned into joint angles by the arm's closed-form inverse kinematics for the paddle's center and normal (the roll about the normal held at zero), and differentiated by five-point central differences with a 2 ms spacing to give each joint's reference r and its first three derivatives. The command sent to each drive is the servo inverse of Equation (15.10).

$$q_c = r + \frac{2\zeta_0}{\omega_0}\dot{r} + \frac{\ddot{r}}{\omega_0^2}, \qquad \omega_0 = 2\pi\cdot 15\,\text{Hz}, \quad \zeta_0 = 1 \tag{15.10}$$

together with its rate, so that a drive tuned as commissioned follows the law without lag while it stays inside its limits, and one retuned away from 15 Hz falls behind it. Measured with a 1 mm stroke, the shaped tool follows within 0.6 % in amplitude and with less than 0.1° of phase lag up to 6 Hz, while the unshaped servos alone reach their −3 dB point at 9.9 Hz, and drives detuned to 6 Hz under the 15 Hz shaping at 4.4 Hz. This is the table-tennis robot's pipeline. None of the laws uses the framework's velocity-form PID of Section 1.6.2, and the settings are listed in Table 15.3.

15.5.1 The Sinusoid, With No Sensor

The open-loop stroke is $z_p = A\sin\theta(t)$ with $\dot{\theta} = 2\pi f$, its amplitude following the slider at 0.10 m/s, and its phase set so that the stroke is at φ when the ball, let go from rest at the target apex, first arrives

(0° is the middle of the up-stroke and 90° its top). It reads no sensor. On its slowing branch it meets the ball at the design acceleration of Equation (15.11), which the scorecard and the Map view report.

$$a^*_{sine} = -\omega\sqrt{A^2\omega^2 - u^{*2}}, \qquad u^* = \frac{g}{2f}\,\frac{1-e}{1+e} \tag{15.11}$$

For two balls the paddle moves across as $x_p = -0.09\cos(\theta - \varphi)$ and up as $z_p = A\sin(2\theta - \varphi)$, so that it strokes at $2f$ and each ball, 0.18 m from the other, meets it every other stroke. Each ball's bounce then lasts $1/f$, and the design acceleration uses u^* for that bounce with the stroke's $\omega = 4\pi f$.

15.5.2 Once per Bounce, on the Impact Switch

The once-per-bounce law is a discrete-time loop whose sample is an impact, in the spirit of feedback from impact times alone [11]. At each ring of the switch it forms the apex error $e_{h,k} = h^* - h_k$ from Equation (15.5) and sets the paddle's speed for the next impact, as shown in Equation (15.12).

$$\begin{aligned} I_k &= I_{k-1} + K_I\, e_{h,k} \\ \hat{u}_{k+1} &= u^*(h^*) + K_b\, e_{h,k} + I_k \\ u_{k+1} &= \min\bigl(1.5\,\mathrm{m/s},\ \max(0,\ \hat{u}_{k+1})\bigr) \end{aligned} \tag{15.12}$$

where $u^*(h^*)$ is Equation (15.6) for the target with $e^* = 0.80$. The integral is not limited, and only the speed it sets is clamped. The next arrival is predicted from the measured incoming speed with e^*, and a quintic in height carries the paddle from its present position, speed and acceleration to the juggling plane at that instant, arriving at u_{k+1} and decelerating at a^*, and continues on that parabola if the ball is late. The law needs no camera. The restitution of impact k that it reports, $\hat{e}_k = (v_k - \dot{p}_k)/(w_k + \dot{p}_k)$, from the launch speed v_k of the flight after it, the arrival speed $w_k = gT_{k-1} - v_{k-1}$ of the flight before it and the paddle's speed, is the e measured badge of the scene.

15.5.3 The Mirror Law, on the Camera

The mirror law moves the paddle as a scaled reflection of the ball the camera shows [9, 14], with a gain that rises when the seen apex $\hat{h}$ is low, as shown in Equation (15.13).

$$z_p = -\kappa\,\hat{z}(t-\tau), \qquad \kappa = \frac{1-e^*}{1+e^*} - K_m\,\frac{\hat{h}-h^*}{2\,(1+e^*)\,h^*} \tag{15.13}$$

with $\kappa \geq 0$ and $\hat{z}$ held at zero or above. For a camera that sees the ball at once and a paddle that follows the mirror target at once, the bounce-to-bounce eigenvalue of the apex is $1 - K_m$, so a gain of 1 corrects an error in one bounce and 2 is the edge. The target passes through a critically damped 12 Hz filter before the shaping of Equation (15.10), because its jumps at each impact would otherwise hold the drives on their limits. With the camera on, the tilt law turns the face to steer a drifting ball back, $\beta = -(\hat{x}/2T^* + \dot{\hat{x}})/\bigl((1+e^*)(v^*+u^*)\bigr)$, from the track's position and speed across and the target orbit's flight time, launch speed and impact speed.

15.5.4 By Hand, Transfers and the Status Line

In Manual the paddle's height follows its station, a slider, the arrow keys or a tap above or below the paddle, each press 2 cm, at up to 1 m/s and 10 m/s^2, with its joint commands carried along that motion, so a press uses at most 85 % of a drive's acceleration limit and the face stays level to better than 0.0001°. A switch to Manual leaves the stations where the law had the paddle, and a switch to Auto or a change of law runs the new law at once while the offset from the held paddle decays over 0.5 s on a minimum-jerk profile. The status line names DROPPED, WALKED OFF, REACH LIMIT, JOINT LIMIT, CAMERA STALE, TWO BALLS and BLIND, and none of them acts on the arm.

Table 15.4 The Nine Cases and the Challenge

#	Configuration and events	Teaching point
1	Manual, ball let go at 50 cm	reaction is too late
2	still paddle, then $\Delta e = -0.10$	restitution from e^2
3	sinusoid 6.3 cm, 1.655 Hz, first arrival at −135°	the first meeting
4	the same at 55°, amplitude swept on the Map view	the window
5	camera off, e 0.80 to 0.72 at 20 s	juggling blind
6	mirror law, K_m 0.5, tilt steering	latency costs height
7	once per bounce, $K_I = 0$, from 35 cm, e to 0.72 at 12 s	an integral per bounce
8	two balls, 2.2 cm at 3.2 Hz, 18 cm apart	one arm, two orbits
9	mirror law, $h^* = 0.20$ m, camera 120 ms late	honestly infeasible
C	60 s, e to 0.72 at 20 s, dropout at 40 s	everything at once

Table 15.5 Outcomes Read From the Running Engine

Case and condition	Result
1, untouched / press 30–90 ms early	dies after 29 impacts at 2.98 s / next apex 41.6–52.2 cm
2, e 0.80 / 0.70	ratios 0.650 / 0.499, their roots 0.806 / 0.707
3, −135° / 55°	dies at 2.28 s / 49 bounces in 30 s, 0.502 m ± 1.6 cm
4, 3.4–9 cm / 9.75 / 10 cm	hold at 5 seeds / scatters with no spread / 0.11–0.96 m
5, before / after 20 s	0.500–0.503 m / 0.489–0.491 m, none dropped of 6
6, K_m 0.5 / 1 / 1.5 / 2	0.374 / 0.442 / 0.462 / 0.472 m
7, P / PI / none, after 12 s	0.457 / 0.503 / 0.383 m
8, 2.2 cm / 1.5 cm	64 and 63 bounces in 40 s / walks off at 2.76 s
9, K_m 0.25 to 2.5	every gain dies, 1.39–2.38 s
C, shipped / suggested	dropped at 23.2 s, 35 / 99 bounces, 1.6 cm rms

15.5.5 Three Models

The simulation model is everything in Section 15.3. The analysis model is the impact map of an ideal paddle and a drag-free ball, whose linearization is Equation (15.7). The laws' own model is smaller still. They assume $e^* = 0.80$ and a drag-free arc, read the ball only through the switch and the camera, and are never told the true restitution, a change of ball or a crooked face.

15.6 Ten Cases, Measured

The cases are listed in Table 15.4 and their outcomes in Table 15.5. Every number below was measured with the engine stepped headless exactly as the page steps it, at the default seed unless a range of seeds is named, and a setting that a sentence changes was set before pressing Run. The apex is the ball's lowest point at the top of a flight. A bounce counts toward the score when its flight rises at least half the target, 25 cm at the 50 cm target, so the millimeter hops of a dying ball are impacts but not juggling. The application passes its 39 verification checks.

15.6.1 By Hand, and a Still Paddle

From 50 cm the ball meets the paddle at 3.13 m/s, 0.319 s after it is let go. Left alone, the paddle lets the bounces die, 33, 22, 14 and 9 cm, and only the first flight counts. A press of the up key moves the paddle 2 cm, and it sends the next bounce back to 41.6 to 52.2 cm only when it comes 30 to 90 ms before the impact. Pressed 15 ms before, the next apex is 35.9 cm, and pressed at the impact it is the untouched 33.2 cm. A simulated student who presses as the ball is seen to land, with a timing spread of 50 ms, keeps 2.6 counted bounces on average and loses all 40 trials within 2.9 to 8.0 s, while one who presses a steady 50 ms ahead with a spread of 25 ms keeps 43.8 on average and loses none of 40 in 30 s. The window rewards anticipation, which reaction cannot supply.

The still paddle of case 2 measures the ball. Its apexes fall 33.2, 22.1, 14.3, 9.2, 5.9 and 3.8 cm, each 0.64 to 0.67 of the one before with the bounce-to-bounce spread, 0.650 on average, and $\sqrt{0.650} = 0.806$ against the ball's 0.80. With the spread removed the ratios are 0.6375 to 0.6397

against $e^2 = 0.6400$, the small drag taking the rest. The ball lies still after 29 impacts in 2.98 s, a geometric series that fits infinitely many impacts into a finite time if nothing stops it, the Zeno solution of Section 1.4.7. A ball of restitution 0.70 keeps 0.499 of each rise and dies after 19 impacts in 1.88 s. The e measured badge, computed as the blind juggler would from two flight times and the paddle's speed, reads 0.80 to 0.82 over the first ten impacts.

15.6.2 The Sinusoid and the Window

The stroke of case 3 is fixed, and only the phase at which the ball first arrives changes. At −135°, with the paddle on its way down, the bounces die after 28 impacts in 2.28 s. At 55° the ball stays up, 49 bounces in 30 s with the apex at 0.502 m and a spread of 1.6 cm, each impact decelerating at 5.9 m/s^2 and the paddle 5.4 cm above the juggling plane. Every first meeting from −80° to 150° in 5° steps ends on that orbit, even −60°, whose first impact comes while the paddle is still speeding up at 5.0 m/s^2. Only first meetings from 155° round to −85°, with the paddle going down, let the bounces die, in 2.1 to 4.6 s. The orbit is an attractor with a large basin.

> **TRY IT IN THE APP** **3 · The Sinusoid**
>
> Load the scenario and press Run. The ball first meets a paddle on its way down, and the bounces die after 28 impacts in 2.28 s. Press Reset, set Phase at the first arrival to 55° and run again. The same stroke now keeps the ball up, 49 bounces in 30 s with the apex at 0.502 m.

Case 4 sweeps the amplitude at 1.655 Hz on the Map view. From 3.4 cm, slowing at 1.1 m/s^2 at impact, to 9 cm, slowing at 9.0 m/s^2, the orbit holds at all five seeds tried, with apex spreads of 1.1 to 4.9 cm caused by the ball's own spread. Past the lower edge, 9.75 cm (−9.97 m/s^2) scatters even with the spread removed, and 10 cm sends the apexes from 0.11 to 0.96 m. Near both edges the restitution's spread decides before the eigenvalue does. At 9.5 cm the paddle slows at 9.7 m/s^2, still inside the window, and the orbit holds with no spread (an apex spread of 0.7 cm) but scatters at four seeds of five. At 3.3 cm the stroke's top speed, 0.343 m/s, is only 3 % above the 0.334 m/s that a 1.655 Hz bounce needs with the ball's drag, and the orbit holds with no spread but dies at four seeds of five, in 6.7 to 14.6 s. At 3.2 cm, whose top speed is 0.333 m/s, and below, the stroke never reaches that speed and the ball dies with or without the spread. Raised to 11 cm while running, the apexes scatter from 0.18 to 0.98 m within 14 s, and lowered to 3 cm the ball lies on the paddle 3.7 s later.

15.6.3 Blind, Through a Change of Ball

Case 5 switches the camera off, so the only sensor left is the impact switch, and the sinusoid does not read that either. At 20 s the ball is changed for one of restitution 0.72, which needs the paddle at 0.48 m/s instead of 0.33. The orbit moves to a new phase by itself. Over seeds 1 to 6 the paddle's acceleration at impact moves from −5.75 to −5.87 before the change to −4.41 to −4.69 m/s^2 after it, the mean apex from 0.500 to 0.503 to 0.489 to 0.491 m, and the ball is never dropped. The deader ball comes off lower, returns a little early and meets a faster part of the stroke, and the frequency, which fixes the flight time, keeps the apex about 1 cm below where it was. The state of the default seed 9.95 s after the change is shown in Figure 15.3.



Figure 15.3 Juggling blind through a change of ball, case 5 at $t = 29.95$ s on the Rig view. The picture is a composite: the Rig view is cropped from the screen, with the header, the control panel and the trends cut, and the four tiles of the scorecard's This app row, from the Learn panel, are set beside it. The camera is off and the status line reads BLIND, the impact switch only. (1) The arm holds the paddle. (2) The last impact, at 29.906 s, still shows as a burst at the contact and an arrow for the ball's change of speed, 5.9 m/s. (3) The badges read 49 bounces, an apex of 49.1 cm, the paddle slowing at 4.8 m/s^2 and rising at 0.48 m/s at impact, a flight of 605 ms, and a restitution of 0.72 measured from the flight times alone for the impact at 29.30 s, against the 0.726 drawn for it. (4) Under 49 consecutive bounces, the count panel's 16 apex bars are all flights after the change at 20 s, the first 6.0 cm and the fifth 4.2 cm below the target outside the 3 cm band, the rest inside it. (5) The scorecard's tiles give an apex error of 1.8 cm rms over 46 flights, the present stroke's design acceleration of −4.6 m/s^2 for the deader ball, and a smallest gap at the log of 0.008 mm.

> **TRY IT IN THE APP** **5 · Blind**
>
> Load the scenario, whose camera is off, and press Run. Watch the apex strip and the badges across the change of ball at 20 s. Nothing measures the ball, yet the deader ball meets a faster part of the stroke, its apex settles about 1 cm lower, and it is never dropped.

15.6.4 The Mirror Law and Its Latency

Case 6 closes the loop through the camera, 40 ms late. The apex settles short of the target by an amount the gain shrinks, 0.374, 0.442, 0.462 and 0.472 m at gains of 0.5, 1, 1.5 and 2, with spreads of 1.0 to 1.3 cm, tighter than the sinusoid's 1.6 cm. With no latency the gains of 0.5 and 1.5 reach 0.455 and 0.485 m, so the latency costs energy, and at 80 ms the gain of 0.5 loses the ball in 11.9 s. Latency does not bring the bounce loop's oscillation to a lower gain. Measured as the lag-one correlation of the apex errors over six seeds, the bounce-to-bounce eigenvalue is -0.50 at a gain of 2 and -0.89 at 2.5 through the 40 ms camera and the target filter, and -0.73 and -0.99 with no latency, against the ideal -1 and -1.5, so at both latencies it nears -1 only at 2.5, where the spread grows to 2.1 cm. The mirror law strikes while the paddle is still speeding up, at 1.0 to 1.6 m/s^2 on average over the gains, outside the open-loop window, where only the feedback holds the orbit. Reversing with the ball at each bounce, it puts a drive on its acceleration limit for 8.3, 16.4, 18.0 and 18.6 ms a bounce at the four gains. Flicked across at 5 cm/s, the ball is steered back by the tilt law within 3.0 cm of the middle, and with the face held level it walks off the edge 2.0 s later.

15.6.5 An Integral That Runs Once per Bounce

Case 7 starts the ball at 35 cm on the once-per-bounce law with the camera off. The proportional part brings the apex to the target in four bounces, 0.390, 0.406, 0.449 and 0.502 m, and the PI runs exactly once per ring of the switch, 48 times for 48 rings. After the change of ball at 12 s the nominal speed u^* is too slow, and the proportional part settles 4 cm short, at 0.457 m, because it needs the error to act. With an integral of 0.2 per bounce the apex returns to 0.503 m, at the cost of an overshoot to 0.569 m on the climb from 35 cm, and with no gain at all it falls to 0.383 m. After the change the deader ball arrives early on the planned quintic and meets the paddle at about $-0.06\ \mathrm{m/s^2}$, near the window's upper edge, so the feedback rather than the stroke's own deceleration holds the orbit.

15.6.6 Two Balls, and a Camera Too Late

Two balls 18 cm apart share one arm in case 8, the paddle crossing between them with up to $9.5\ \mathrm{m/s^2}$ across at its turns and stroking 2.2 cm at 3.2 Hz. In 40 s they bounce 64 and 63 times at mean apexes of 0.492 and 0.493 m, spreads of 2.1 and 2.0 cm, every impact slowing, at 5.41 and $5.45\ \mathrm{m/s^2}$ on average against the $5.67\ \mathrm{m/s^2}$ the stroke was designed for. At 1.5 cm the balls fall off the beat and one walks off the edge after 8 impacts, 2.76 s into the run.

Case 9 is infeasible by design. A 0.20 m bounce lasts 0.404 s, and a camera 120 ms late shows each frame after 30 % of a bounce has passed. Every mirror gain from 0.25 to 2.5 in steps of 0.25 drops the ball, in 1.39 to 2.38 s, each time by the bounces shrinking until the ball lies on the paddle. The paddle rises for a ball that left its apex 120 ms earlier, meets it late and slow, and gives back less than the impact took. With the 40 ms camera the same law holds a 0.172 m apex, at 80 ms it lasts 9.10 s and at 100 ms 2.94 s. Without the camera a 3 cm sinusoid at 2.5 Hz keeps the ball up indefinitely at 0.223 m, decelerating at $6.5\ \mathrm{m/s^2}$ at impact. Information that arrives after it was needed cannot be bought back with gain. A robot that picks parts off a conveyor meets the same latency and moves each late camera report on by the belt travel since its capture (Volume 20, Chapter 10).

Pause. Try Exercise 15.3 before reading on.

15.6.7 A Minute of Juggling

The challenge juggles for 60 s at a 0.50 m target through a change of ball from 0.80 to 0.72 at 20 s and a camera dropout of 0.5 s at 40 s. The target is no drop and an apex within 3 cm rms of the target over every flight from 2 s, with the model and the disturbances as given, and the score is the count of consecutive bounces. Changing a model parameter other than the seed, a disturbance or the target during the challenge forfeits that attempt, and the run pauses at 60 s with its verdict. The configurations are compared in Table 15.6. As shipped, a 4.1 cm sinusoid at 1.60 Hz slows at only $2.1\ \mathrm{m/s^2}$ at impact. The deader ball needs the paddle at 0.50 m/s at that frequency, the stroke never moves faster than 0.41 m/s, and the ball is dropped about 3 s after the change at every seed from 1 to 10. The suggested stroke, 6.3 cm at 1.655 Hz, slows at $5.9\ \mathrm{m/s^2}$, has the margin for the deader ball, never reads the camera and passes at all ten seeds. The once-per-bounce law passes too, and the comparison shows what feedback adds here. Its integral holds the mean apex after the change at 0.497 to 0.500 m, where the blind stroke settles at 0.489 to 0.490 m, but its strokes then meet the deader ball at -0.05 to $+0.24\ \mathrm{m/s^2}$, with no deceleration to damp an error, and its spread grows to 1.5 to 2.4 cm against 1.2 to 1.6 cm. The bias that feedback removes is smaller than the spread it lets in. The suggested minute plays in 30.0 s of wall time at the top speed of 2× in a browser. Figure 15.4 draws the five configurations against the 3 cm target.

> **DID YOU KNOW? Feedback That Scores Worse**
>
> From the simulator: over the challenge minute at seed 1, the blind 6.3 cm sinusoid holds the apex to 1.6 cm rms and the once-per-bounce law to 2.5 cm, while the mirror law on the camera keeps 98 bounces and still fails at 7.6 cm (Table 15.6).

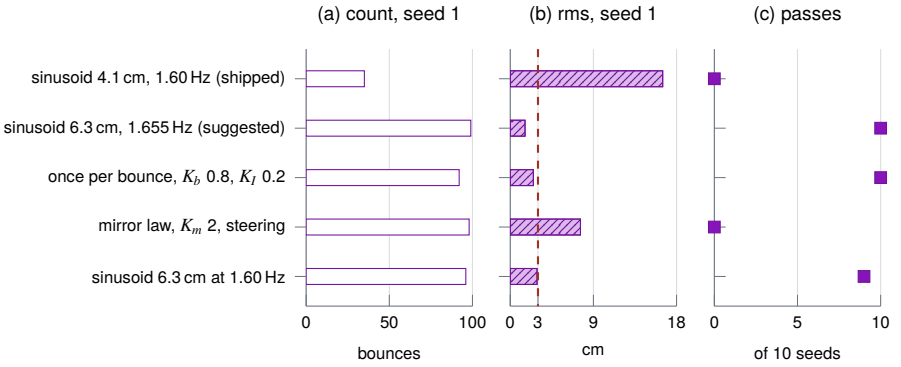


Figure 15.4 The challenge minute of Table 15.6 in its five configurations: (a) the count of consecutive bounces at seed 1 as white bars, (b) the apex error at seed 1 as shaded bars against the 3 cm rms target (dashed) and (c) the passes over ten seeds as dark squares. The table prints the values. Notice that the mirror law keeps almost as many bounces as the suggested sinusoid yet fails on its error, and that the once-per-bounce feedback passes with a larger error than the blind stroke.

Table 15.6 The Challenge Over 60 s, Seeds 1 to 10

Configuration	Count seed 1	Rms seed 1 (cm)	Rms, 10 seeds (cm)	Passes
sinusoid 4.1 cm, 1.60 Hz (shipped)	35[a]	16.5	16.508–21.450	0 of 10
sinusoid 6.3 cm, 1.655 Hz (suggested)	99	1.6	1.630–2.236	10 of 10
once per bounce, K_b 0.8, K_I 0.2	92	2.5	2.252–2.690	10 of 10
mirror law, K_m 2, steering	98	7.6	7.645–7.942	0 of 10
sinusoid 6.3 cm at 1.60 Hz	96	2.9	2.742–3.065	9 of 10

[a]Dropped at 23.2 s (23.1 to 23.4 s over the seeds, after 35 or 36 counted bounces).

15.7 Technology in Practice

What Measures It

Juggling robots have sensed the ball in three ways. The batting robot of 1989 read its puck through an oscillator inside it and a grid in its plane [14], a three-axis arm juggled two balls on a real-time stereo camera system [15], and the blind jugglers sense no more than their impacts [11, 12]. To judge a blind robot its builders still added instruments. A microphone timed each impact and lit a small array of lights that a video camera above the paddle could see, the ball's place on the paddle came from those frames in an offline analysis, and the motor encoder gave the paddle's position and speed at impact [12]. The impact switch and the arm's encoders of this chapter, and its camera, are these two kinds of measurement (Table 15.1).

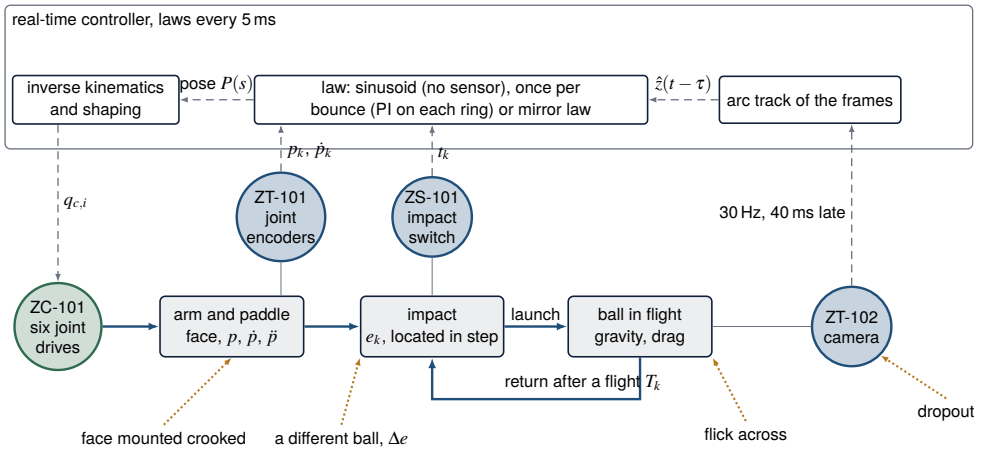


Figure 15.5 The juggling loop drawn on real hardware with the chapter's symbols. Sensors are drawn blue, the final element green and disturbances dotted. The impact switch, in practice a vibration sensor on the paddle, gives the impact times t_k, the joint encoders give the paddle's height and speed at each impact, and the camera's late frames feed the arc track. Only the mirror law reads the camera, only the once-per-bounce law the switch, and the sinusoid reads nothing.

> **TECHNOLOGY** **Sensor: Accelerometers and Microphones That Hear the Impact**
>
> A real impact switch is a vibration sensor. On one planar juggler two accelerometers at the tops of its arms recorded the small high-frequency vibrations of each impact, and their raw signals were band-pass filtered to remove the slow accelerations of the arms' own motion and the noise above [11]. A blind paddle robot heard its impacts through contact microphones on the paddle and read the paddle's position and speed at impact from its motor encoder [12, 16]. The sensor fails by missing an impact or by reporting a false one. Those microphones did not detect every impact, so a ball state was trusted only after a consistent series of impact times [16], and the planar juggler smoothed its arm motion to prevent false detections [11]. The chapter's switch rings at every impact.

What Moves It

A blind juggler can move its paddle along one axis. One such robot drove a concave aluminum paddle with an industrial linear motor, sized to juggle balls with a restitution of at least 0.7 to apex heights of at least 1.2 m, and a servo controller tracked the paddle's trajectory [12]. A juggling arm moves the paddle through all its joints at once. Figure 15.5 draws the loop of this chapter on such an arm, its drives modeled by Equation (15.4).

> **TECHNOLOGY** **Final Element: An Arm Driven Near Its Torque Limits**
>
> One learning system for a cable-driven juggling arm interpolated via-points in joint space with cubic splines and tracked them with a PD controller and gravity compensation, at accelerations of up to eight times gravity [17]. Tight box constraints on the joint trajectory kept the arm from hitting its joint limits. Rigid-body simulators could not represent the effects near the torque limits or the cable dynamics at high acceleration, and the wear ruled out running for days [17]. The arm of this chapter has speed and acceleration limits but no torque limits or flexibility.

Where the Controller Runs

The juggling controllers of the literature run on dedicated real-time computers. In the batting robot of 1989 the observer reported to the planning node at 1 kHz and the motor controller issued

torque commands at 2 kHz [14]. The planar blind juggler generated its code from a block-diagram simulation tool, ran it on a dedicated target computer under a real-time operating system, and sampled at 1 kHz [11]. The blind paddle robot split the work, a state machine on a real-time control system timing each stroke and passing profile parameters to the motor's servo controller [12]. The laws of this chapter run every 5 ms with the joint commands (Table 15.2), but the once-per-bounce law acts only when the switch rings, so its sample is the bounce.

> **IN PRACTICE** **Timing the Deceleration From Impact Statistics**
>
> A blind paddle robot was brought into service in steps. The servo controller of its linear motor was tuned by hand until the tracking was acceptable, and a real-time control system timed the stroke and gave the servo its profile parameters [12]. Around each nominal impact the paddle decelerated for a set interval, first sized for a static error in the restitution and then fixed from the statistics of the measured impact times. An interval of 0.05 s worked well, and a longer one added vertical robustness, because larger timing errors still met a decelerating paddle [12]. The once-per-bounce law of this chapter also keeps decelerating on its parabola when the ball is late.

What Is Modernizing It

Three developments touch this loop. The first is learning on the real arm. A cable-driven arm learned to juggle two balls from 56 minutes of experience with a binary reward, and its final policy juggled for up to 33 minutes, about 4500 repeated catches [17]. The second is learning a correction on top of an idealized plan. A residual learned from the task error gave stable three-, four- and five-ball juggling on two arms, open loop with respect to the balls, and converged from the second attempt [18]. The third is control that runs on events rather than on a clock. In event-triggered control the inputs are updated only when events occur, while a controller that executes periodically can load shared computing and communication more than the task needs [19]. The once-per-bounce law of Section 15.5 is a controller of that kind, with the impact as its event.

15.8 Limitations

The ball is a rigid, spinless sphere in one plane with a restitution that does not depend on speed, and its impacts are instantaneous and frictionless across the face, so spin, grip and the fall of restitution with impact speed are represented only by the bounce-to-bounce spread. A ball whose restitution jumps at one instant would not occur on a real paddle, but the step separates the margin of the stroke from the slow wear that a real ball shows. The face is flat, so a drifting ball is steered back only through the camera. The arm moves through exact kinematics and speed- and acceleration-limited drives with no rigid-body dynamics, and its stroke and tilt limits are teaching values chosen for wrist margin rather than taken from a particular robot.

> **OPEN CHALLENGE** **Trusting a Learned Juggler in the Loop**
>
> Closed-loop policies suit robot juggling, because the contacts between hand and ball are not consistent and feedback could recover from the irregularities. A learned closed-loop policy is hard to trust, however. Noisy ball observations or outliers might make it unstable and drive the arm into its joint limits, and its builders knew of no verification process for complex closed-loop networks, so they ran an open-loop stroke whose safety rests on the tracking controller and the joint constraints [17]. Progress would be a guarantee that a closed-loop policy is stable for every possible observation, edge cases included.

The analysis is local. Equation (15.7) describes small errors about the period-1 orbit, and the basin of that orbit, how far a bounce can stray and still return, was measured on the engine only through the phases of case 3. The once-per-bounce integral has no anti-windup, the mirror law's eigenvalue of $1 - K_m$ holds only for an ideal camera and paddle, and the camera has a fixed rate, latency and Gaussian noise with no blur or misread frames. Every threshold, the 5 mm/s at which a ball comes to

rest, the stroke and tilt limits and the challenge's 3 cm rms, is a teaching value of a simulator and not a measured property of a juggling robot.

15.9 Conclusion

A ball on a paddle is controlled one bounce at a time, and this chapter put that map on a six-axis arm with limited drives, an impact switch and a late camera. The linearized impact map has a determinant of e^2 for any stroke and a trace that falls with the paddle's acceleration at impact, so a paddle slowing between 0 and 9.93 m/s^2 holds the orbit with no sensor, and the eigenvalues measured with the arm in the loop, 0.79 to 0.83, match the closed form's 0.80. A 6.3 cm stroke at 1.655 Hz keeps the ball at 0.502 m within 1.6 cm and carries it blind through a change of ball with the apex 1 cm lower, while near either edge of the window the ball's own spread decides. The mirror law on a 40 ms camera settles 3 to 13 cm short as its gain falls, a camera 120 ms late defeats every gain, and a once-per-bounce integral removes the bias of a deader ball but lets in more spread than it removes. The model is planar, its impacts ideal and its arm without dynamics. A fuller model would add spin and a speed-dependent restitution, a concave face, rigid-body arm dynamics, and a basin-of-attraction map beside the eigenvalue window. The juggler is a machine whose stability is designed into its stroke before any sensor is added.

KEY POINTS

- A paddle that is slowing down when it meets the ball holds the bounce with no sensor, because a late ball gets a smaller push and comes back sooner.
- The linearized impact map has a determinant of e^2 for any stroke, and the orbit is stable for paddle accelerations at impact between −9.93 m/s^2 and zero.
- A 6.3 cm stroke at 1.655 Hz keeps the ball at 0.502 m within 1.6 cm and carries it blind through a change of ball.
- A camera 120 ms late defeats every mirror gain, and a once-per-bounce integral removes the bias of a deader ball but lets in more spread than it removes.

Where else this applies. A correction that waits for its measurements also runs the filling line of Chapter 9: with one group in four weighed instead of every group, the per-head correction at K 0.10 passes at none of 12 seeds instead of all 12 (Table 9.6).

References

[1] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Systems*, vol. 29, no. 2, pp. 28–93, 2009.

[2] B. Brogliato, *Nonsmooth Mechanics: Models, Dynamics and Control*, ser. Communications and Control Engineering. Cham: Springer International Publishing, 2016.

[3] T. McGeer, "Passive dynamic walking," *The International Journal of Robotics Research*, vol. 9, no. 2, pp. 62–82, 1990.

[4] J. W. Grizzle, G. Abba, and F. Plestan, "Asymptotically stable walking for biped robots: Analysis via systems with impulse effects," *IEEE Transactions on Automatic Control*, vol. 46, no. 1, pp. 51–64, 2001.

[5] K. M. Lynch and M. T. Mason, "Dynamic nonprehensile manipulation: Controllability, planning, and experiments," *The International Journal of Robotics Research*, vol. 18, no. 1, pp. 64–92, 1999.

[6] P. J. Holmes, "The dynamics of repeated impacts with a sinusoidally vibrating table," *Journal of Sound and Vibration*, vol. 84, no. 2, pp. 173–189, 1982.

[7] N. B. Tufillaro and A. M. Albano, "Chaotic dynamics of a bouncing ball," *American Journal of Physics*, vol. 54, no. 10, pp. 939–944, 1986.

[8] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, ser. Applied Mathematical Sciences. New York: Springer, 1983, vol. 42.

[9] M. Buehler, D. E. Koditschek, and P. J. Kindlmann, "Planning and control of robotic juggling and catching tasks," *The International Journal of Robotics Research*, vol. 13, no. 2, pp. 101–118, 1994.

[10] S. Schaal and C. G. Atkeson, "Open loop stable control strategies for robot juggling," in *Proceedings IEEE International Conference on Robotics and Automation*. IEEE Computer Society Press, 1993, pp. 913–918.

[11] R. Ronsse, P. Lefèvre, and R. Sepulchre, "Rhythmic feedback control of a blind planar juggler," *IEEE Transactions on Robotics*, vol. 23, no. 4, pp. 790–802, 2007.

[12] P. Reist and R. D'Andrea, "Design and analysis of a blind juggling robot," *IEEE Transactions on Robotics*, vol. 28, no. 6, pp. 1228–1243, 2012.

[13] R. Ronsse, P. Lefèvre, and R. Sepulchre, "Sensorless stabilization of bounce juggling," *IEEE Transactions on Robotics*, vol. 22, no. 1, pp. 147–159, 2006.

[14] M. Buhler, D. E. Koditschek, and P. J. Kindlmann, "A family of robot control strategies for intermittent dynamical environments," *IEEE Control Systems Magazine*, vol. 10, no. 2, pp. 16–22, 1990.

[15] A. A. Rizzi and D. E. Koditschek, "Further progress in robot juggling: The spatial two-juggle," in *Proceedings IEEE International Conference on Robotics and Automation*. IEEE Computer Society Press, 1993, pp. 919–924.

[16] P. S. Reist, "Sensorless juggling machines and chaos for control," Doctoral thesis, ETH Zurich, 2015.

[17] K. Ploeger, M. Lutter, and J. Peters, "High acceleration reinforcement learning for real-world juggling with binary rewards," in *4th Conference on Robot Learning (CoRL 2020)*, 2020, arXiv:2010.13483.

[18] K. Ploeger, A. Kshirsagar, and J. Peters, "Task-error residual learning for real-robot five-ball juggling," arXiv preprint arXiv:2606.16978, version 2, 2026.

[19] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1992–1997, 2015.

15.10 Exercises

15.1 Concept check · 2 min

The 6.3 cm sinusoid at 1.655 Hz keeps the ball bouncing with the camera off and reads no other sensor. Why does a bounce that comes back a little late not drift further off the beat?

(a) A late ball meets a paddle that has already slowed, gets a smaller push and comes back sooner

(b) A late ball meets a paddle that is still speeding up, gets a larger push and is sent back on the beat

(c) The impact switch reports the late impact, and the stroke shifts its phase to meet the next bounce

(d) The air drag takes more energy from a high bounce, so every apex returns to the target by itself

15.2 Concept check · 1 min

With the arm in the loop, the restitution exact and no drag, a sinusoid at 1.566 Hz is designed to meet the ball while the paddle is speeding up at 2 m/s^2. The closed form gives an eigenvalue of 1.977 there. What does the engine show?

(a) The ball is lost within a few bounces, as the eigenvalue predicts

(b) The ball settles on the same stroke's slowing branch, at about −2 m/s^2

(c) The orbit splits into a high and a low apex, as it does past the lower edge

(d) The ball holds the speeding-up orbit, because the servos add the damping it lacks

15.3 Concept check · 2 min

In Scenario 9 the mirror law drops a 0.20 m bounce through a camera 120 ms late at every gain from 0.25 to 2.5. Which change keeps the ball up?

(a) Add an integral term to the mirror gain, so the apex error is removed over several bounces

(b) Lower the target apex further, so each bounce needs less energy from the paddle

(c) Switch the camera off and run a 3 cm sinusoid at 2.5 Hz, which needs no measurement

(d) Raise the camera's frame rate from 30 to 120 Hz while its latency stays at 120 ms

15.4 Theory and calculation · 6 min

A drag-free ball with a constant restitution $e = 0.80$ is juggled to an apex $h = 0.50\,\mathrm{m}$ above the point of impact, with $g = 9.81\,\mathrm{m/s^2}$.

(a) Find the launch speed v, the flight time T and the upward paddle speed u^* at impact that hold the period-1 orbit (Equation (15.6)). *Checkpoint:* $v = 3.13\,\mathrm{m/s}$, $T = 0.639\,\mathrm{s}$, $u^* = 0.348\,\mathrm{m/s}$.

(b) The same ball is let go from rest 0.50 m above a still paddle. Each flight lasts e times the one before. Sum the fall and every later flight to find when the bounces end. *Checkpoint:* $0.319 + 2ev/\big(g(1-e)\big) = 2.87\,\mathrm{s}$.

(c) Case 2 measures an average apex ratio of 0.650 and a ball at rest after 29 impacts in 2.98 s. Repeat (b) with the effective restitution $\sqrt{0.650}$, and say what kind of solution of a hybrid system the still paddle shows. *Checkpoint:* $e = 0.806$ gives 2.97 s. A Zeno solution.

15.5 Theory and calculation · 7 min

Use the linearized impact map of Equation (15.7) with $g = 9.81\,\mathrm{m/s^2}$.

(a) At $e = 0.80$ and a paddle acceleration at impact $a = -5\,\mathrm{m/s^2}$, find $\mathrm{tr}\,J$ and the eigenvalues. How fast does an error shrink? *Checkpoint:* $\mathrm{tr}\,J = -0.011$, a complex pair of modulus 0.800.

(b) Repeat for $a = +2\,\mathrm{m/s^2}$. *Checkpoint:* $\mathrm{tr}\,J = 2.301$, eigenvalues 1.977 and 0.324.

(c) Find the lower edge of the window, Equation (15.9), for $e = 0.80$ and for a deader ball with $e = 0.72$. *Checkpoint:* $-9.931\,\mathrm{m/s^2}$ and $-10.070\,\mathrm{m/s^2}$.

15.6 Theory and calculation · 8 min

A sinusoidal stroke $z_p = A\sin\omega t$ meets a bounce lasting $1/f$, and on its slowing branch it meets the ball at the design acceleration of Equation (15.11). Take $g = 9.81\,\mathrm{m/s^2}$ and no drag.

(a) At $f = 1.655\,\mathrm{Hz}$, find the paddle speed u^* the bounce needs for $e = 0.80$ and for $e = 0.72$. *Checkpoint:* 0.329 m/s and 0.482 m/s.

(b) Find the design acceleration of the 6.3 cm stroke at 1.655 Hz for both balls. *Checkpoint:* $-5.89\,\mathrm{m/s^2}$ and $-4.61\,\mathrm{m/s^2}$.

(c) The challenge ships a 4.1 cm stroke at 1.60 Hz and changes the ball to $e = 0.72$ at 20 s. Compare the stroke's top speed with the speed the deader ball needs, and say what happens. *Checkpoint:* Top speed 0.412 m/s against 0.499 m/s: no impact can hold the orbit.

15.7 App challenge · 7 min

Open Scenario 2, "2 · A still paddle", at the default seed, and run at 1×. The run pauses by itself when the ball comes to rest.

(a) Run it as loaded. Read the first six apexes from the Orbit view's table, divide each by the one before (start from the 50 cm release), and average the six ratios. Estimate e. Tolerance 0.005 on the ratio. *Checkpoint:* 33.2, 22.1, 14.3, 9.2, 5.9, 3.8 cm. Mean ratio 0.650, $e \approx 0.806$.

(b) Read the clock and the impact count on the Sensors card when the run pauses. Then press Reset, set A different ball to -0.10 and run again. Read the same two values and the ratio of the second apex to the first. Tolerance 0.01 s. *Checkpoint:* 2.98 s and 29 impacts. Then 1.88 s, 19 impacts, ratio about 0.50.

15.8 App challenge · 8 min

Open Scenario 4, "4 · The window", at the default seed, and run at 1×. Each part sets Stroke amplitude before pressing Run.

(a) Set Stroke amplitude to 9.75 cm and run to 30 s. Read the Design acceleration at impact tile, compare it with the window's lower edge, and read the lowest and highest apex between 10 and

30 s on the Map view or in the exported column `truth_apex_m`. Tolerance 0.02 m. *Checkpoint:* $-10.0\ \mathrm{m/s^2}$, outside the edge at $-9.93\ \mathrm{m/s^2}$. Apexes from about 0.13 to 0.94 m.

(b) Reload the scenario, set Stroke amplitude to 3.3 cm and run. Read the tile and the time at which the ball comes to rest, then explain why a stroke inside the window loses the ball. Tolerance 0.1 s. *Checkpoint:* $-1.0\ \mathrm{m/s^2}$, dropped at 14.6 s after 21 counted bounces.

15.9 App challenge · 9 min

Open Scenario 7, "7 · Once per bounce", at the default seed, and run at 2× (two seconds of plant time per wall second). The ball is changed for one of restitution 0.72 at 12 s. In each part run to 30 s, export the CSV file and average the column `truth_apex_m` over the rows from 20 to 30 s. Tolerance 0.005 m.

(a) Run it as loaded, with Integral per bounce at 0. *Checkpoint:* 0.459 m.

(b) Press Reset, set Integral per bounce to 0.2 and run again. Also read the highest apex before 12 s. *Checkpoint:* 0.504 m, highest apex 0.569 m.

(c) Press Reset, set Bounce-to-bounce gain and Integral per bounce to 0, and run again. Explain the three results. *Checkpoint:* 0.383 m.

15.10 Application area · 3 min

An engineer measures a ball's restitution by dropping it from 0.50 m onto a still paddle and reading the apex ratios, as in case 2. Use Equation (15.1).

(a) Find the drag as a fraction of gravity at the 3.13 m/s of the first bounce, for the chapter's rubber ball with $k_D = 0.0075\ \mathrm{m^{-1}}$ and for a table-tennis ball with $k_D = 0.126\ \mathrm{m^{-1}}$. *Checkpoint:* 0.75 % and 12.6 %.

(b) What does that do to the measurement, and what should the engineer do? *Checkpoint:* The table-tennis ball reads e^2 up to 5.7 % low.

15.11 Application area · 4 min

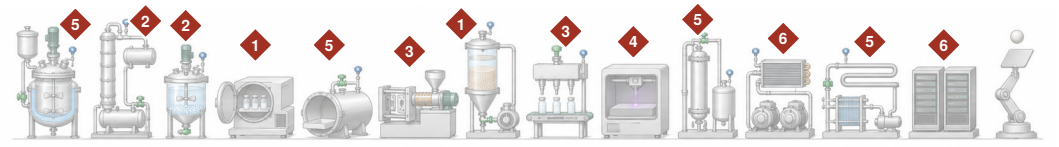
A team builds a blind juggling machine for a trade-show booth, a paddle on a fixed sinusoid with no camera, to run all day with balls that wear. From the chapter's assumptions and limitations, name three things the real machine must handle that the simulator leaves out or idealizes, and say how the stroke should be chosen.

15.12 Review the agent · 5 min

For the challenge, an AI assistant writes: "In case 6 the mirror law at a gain of 2 holds the apex with a spread of 1.0 to 1.3 cm, tighter than the sinusoid's 1.6 cm. Set Law to Mirror law, Mirror gain to 2 and Paddle tilt law to Steer. The apex error will be about 1.3 cm rms, well inside the 3 cm target, and the camera feedback will also absorb the deader ball." Find the flaw and say how you would check the claim in the app.

CHAPTER 16

Summary and Outlook



Open gaps 1 end quality 2 trajectories 3 sparse results 4 learning 5 supervisors 6 coupling

Notice that the gap of verified supervisors joins four plants, the reactor, the autoclave, the capture train and the pasteurizer, more than any other gap. Illustration, not to scale.

IN PLAIN TERMS **What Fourteen Batches, Runs and Switches Taught**

The main idea. Each chapter of this book ran a plant that starts and stops: a batch judged once at its end, a run repeated many times, or a machine that jumps between modes. Nothing settled at a steady value that a controller could hold.

The challenge. What decided the result was the path the batch followed, the reading that ended each step and the simple guards around every switch. A controller that held its target well still made a poor batch when it was told to follow the wrong path.

Why it matters. Judging the end result on what was really made tells a good batch from one that only looked good on the screen.

Fourteen chapters put batches, repeated runs and switched plants in front of a controller. In each one the result was judged on something no loop held at a setpoint: a dose finished without a runaway, a drum at its purity, a vial dry and intact, a part of the right mass, a bounce that came back to its apex. Table 1.3 of Chapter 1 listed what each chapter judges at the end, the strategies it compares and what keeps its switching safe, and left the results to the chapters. This chapter adds the results, draws a short guide for choosing a strategy, looks at where practice is heading, reports how the volume itself was made with AI agents and what they got wrong, and lists the problems that remain open.

16.1 What the Chapters Showed

Tables 16.1 and 16.2 compare the chapters. The first holds the plants judged at the end of one batch, where the strategy is a trajectory, a guard that ends a phase or a correction between shots. The second holds the plants that repeat or switch, where the strategy is a correction from one run to the next, a learned profile or a supervisor over a switched plant. Each row names what is judged, a strategy that worked with the device that kept its switching safe, one that failed, and the measured result with the table or section that reports it. Figure 16.1 sets one like pair of numbers from each row side by side.

The columns of the two tables make five points.

The path decided the product, not the tracking. On the crystallizer the best-tracked curve, 0.142 °C at its worst, made 96.3 % fines, and the cubic curve with a larger deviation made 13.8 %. On the batch column the open-loop ramp collected 15.73 mol against the composition loop's 15.29 mol without any feedback. The semi-batch reactor held its temperature within 0.2 K while it stored enough reagent to rupture its disc. In each case the loop did what it was asked. The question that decided the batch was what it was asked to follow (Section 1.4.3).

Table 16.1 Batches Judged at the End, Chapters 2 to 8

Chapter	Judged at the end	What worked	What failed	Measured result
2, semi-batch reactor	The whole dose fed, with the MTSR below its limit if cooling fails.	Dosing limited on the calorimeter's MTSR. A feed cut with hysteresis and a latched interlock.	A fixed dose at 30 °C, held within 0.2 K while 4398 mol accumulate.	Circulation lost at 2:30, the fixed dose is lost at 3:55 and the limited dose peaks at 74.6 °C. Challenge 22.76 against 1.61 K (Table 2.5).
3, batch column	Product at a cumulative purity of 0.990 in 180 min.	An open-loop reflux ramp searched on the model. Tracking that opens the composition loop without a bump.	The same ramp on a charge of 0.42. Constant composition behind an 8 min analyzer.	15.10, 15.29 and 15.73 mol for constant reflux, composition and ramp (Table 3.5). Wrong charge 1.26, trimmed 4.74, re-optimized 10.69 mol (Table 3.7).
4, crystallizer	Mean size and fines of the filtered crystals.	A cubic cooling curve, slow first. A cascade that states the wall limit.	Curves that tracked better. A cubic batch shortened to 150 min.	Fines 96.3, 99.3 and 13.8 % at peak deviations 0.142, 0.637 and 0.237 °C (Table 4.4). At 150 min, 68.2 % (Table 4.7).
5, freeze-dryer	Every vial group dry, none collapsed.	An endpoint-aware shelf with a learned resistance and the Pirani endpoint.	The shipped recipe and a probe-limited shelf. The learned resistance with a gauge factor of 1.7, not 1.6.	Passes 0, 0 and 6 of 6 seeds. Stoppered at 24.74 against 26.12 h for a tuned recipe and 24.47 h reading the truth (Table 5.6). Every group collapsed at 1.7 (Table 5.8).
6, autoclave	Cure of the slowest part, exotherm of the hottest core.	Lead/lag selection, the max loop through a low selector, and the observer's exotherm feedforward.	The air's clock alone. A thick spar ramped at 3 °C/min without the feedforward.	Spar peak 243.9 against 184.8 °C with the feedforward, at the same cycle (Table 6.6). Challenge passed in 289.8 min (Table 6.7).
7, molding machine	Mass of every part.	Switchover by tracking. A correction from the scale with 40 MPa of authority.	The fixed recipe. Switchover at 50.5 % fill (short) or 103.6 % (flash).	Parts in tolerance 9, 26 and 30 of 30 (Table 7.6). Handover step at most 0.44 % in 20 shots (Table 7.5).
8, particle coater	Four layer thicknesses and the carbide spread.	An inferential endpoint corrected by a load cell. Ramped gas changes and a timeout.	A timer on a cold carbide step. The estimate alone at a coating efficiency of 0.72.	Carbide 26.2 by timer, 35.0 µm by estimate (Table 8.4). Gap 28.35 closed to −0.79 µm for 191 to 219 min (Table 8.6).

An endpoint is only as good as the estimate behind it. A clock cannot see a cold carbide step, and the coater's estimate cannot see a coating efficiency outside its model, which left the batch 28.35 µm thin with every setpoint held. A load cell closed that gap and was itself blind to a wrong charge ticket. The freeze-dryer's learned resistance held its margin over a wide range and failed outright on a gauge factor 6 % high. The guard that worked in each case read a second, independent measurement, as Table 1.1 anticipated.

A correction between runs pays only for what persists. The molding correction took the campaign from 9 to 30 parts in tolerance, and at its default gain it widened the shot-to-shot spread by 15 %, close to the 17 % its loop gain predicts (Section 7.5). On the vial filler, with 5.9 weighed vials of a head between a fill and its result, the plain per-head law spread the heads to 47 to 48 mg, where

Table 16.2 Runs That Repeat and Plants That Switch, Chapters 9 to 15

Chapter	Judged at the end	What worked	What failed	Measured result
9, vial filler	Fill of every vial, with results in flight.	A record-based estimate per head, or the plain law detuned. A validity window and a step limit.	Pooled and per-head corrections at a gain of 0.50.	Accepted 566 and 519 against 2325 (record-based) and 2324 (detuned), with 2342 for ideal information (Table 9.6).
10, powder bed fusion	Melt pool width along every layer.	Learning from the last layer with a Q-filter, feedback underneath, the profile kept at a new outline.	More feedback gain. Learning without the filter.	Defects 4.7 fixed, about 2.5 with feedback, 0 from layer 3 with learning. No filter, 24.8 by layer 25. Kept 13.2 against 23.2 reset, in percent (Table 10.4).
11, antibody capture	Antibody per liter of resin and hour, antibody lost.	A UV threshold or a learned capacity. Switch only when the next column is ready.	A 25-minute clock. Loading on through a long cleaning hold.	38.32 and 37.98 against 37.23 g/L/h, recovery 98.57 against 90.42 % (Table 11.6). Loss 9.05 against 0.05 % paused (Table 11.5).
12, refrigeration pack	Goods temperature, energy and compressor starts.	A scheduler that desynchronizes the cases, floating pressures, a 6 K superheat. Band, minimum run and rest.	Free-running thermostats, in step after 65.2 min. A 2 K superheat.	Summer day 9 trips and 383 starts against 0 and 108 (Table 12.7). Three hours 45 starts free against 8 scheduled (Table 12.8).
13, pasteurizer	Hold time and temperature of every parcel.	A cascade with feedforward under a diversion supervisor, return after a flush, conditional integration.	No flush before the return. Two thermometers drifting together.	Challenge 4 diversions and 1554 L against none (Table 13.6). No flush, 8 diversions (Table 13.5). 128 parcels pass unseen (Table 13.4).
14, autoscaled service	Latency percentile, dropped requests and cost.	Admission control, a retry budget, a forecast floor, a cooldown.	The platform's rule in a fourfold crowd. Retries with the fleet capped.	Crowd 14 breach minutes and 23.7 % dropped, 0 and 4.8 % admitted on requests (Table 14.4). Storm 129 to 0 minutes (Table 14.5).
15, juggling paddle	A period-one bounce at its apex.	An open-loop stroke whose paddle slows at impact, 6.3 cm at 1.655 Hz.	The shipped stroke of 4.1 cm. A mirror law behind a camera 120 ms late.	Challenge passed on 0 and 10 of 10 seeds, 1.6 cm rms (Table 15.6). Every gain dies at 120 ms (Table 15.5).

the record-based estimate held 4.6 to 6.8 mg (Table 9.5). Learning on the powder bed removed defects that repeated at every corner and diverged without its filter. Section 1.4.5 gives each result its formula: the noise widening $2/(2-\lambda)$, the delay polynomial and the monotone convergence factor.

Switching was kept safe by devices outside the controller. Every chapter that switched relied on a device that had no model in it: a feed cut with hysteresis and a latch, tracking at a switchover, a flush before the return to forward flow, a minimum run and rest, a cooldown, a retry budget, a column that is ready before the switch. Removing one was measurable. The pasteurizer without its flush diverted 8 times instead of once. The autoclave's predictive controller met its limits only after the selector's protections were kept above it (Section 6.7.7).

A stable cycle can be the goal or the fault. The juggler's stroke makes the period-one bounce an attracting cycle with no sensor at all, and the measured eigenvalues, 0.79 to 0.83, match the closed form's 0.80 (Section 15.4). The refrigeration thermostats find a stable cycle too, all four in step

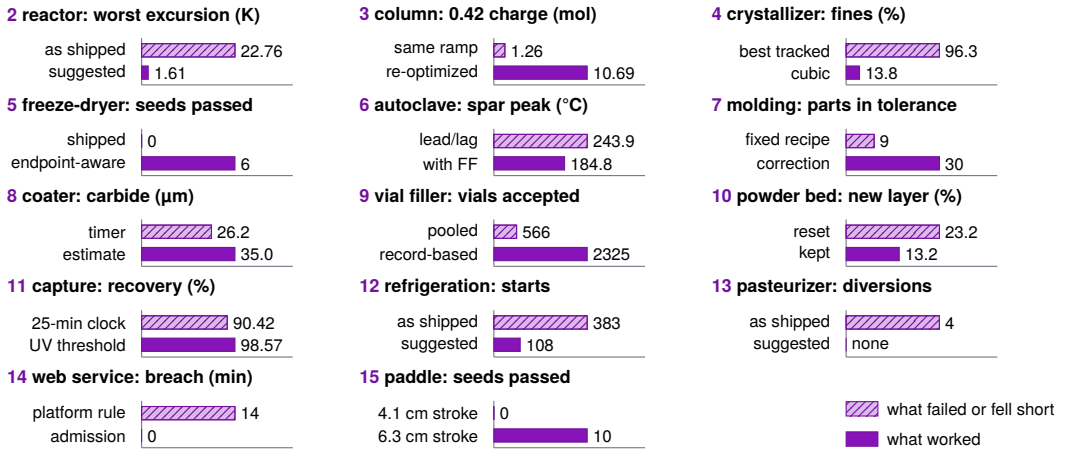


Figure 16.1 One like pair of numbers from each row of Tables 16.1 and 16.2, digit for digit, each panel on its own scale. The hatched bar is the policy or setting that failed or fell short, the solid bar the one that worked. Higher is better for the column, the freeze-dryer, the molding machine, the coater, the vial filler, the capture train and the paddle, lower for the other seven. The pasteurizer's "none" is drawn at zero. Notice that every better number came with a change of path, endpoint, correction or guard, and none with a tighter loop alone.

through one suction pressure, and that cycle is the fault. Section 1.4.9 reads both the same way, as the fixed point of a cycle-to-cycle map.

16.2 Choosing a Method

The chapters suggest an order of questions for a new batch, cycle or switched plant. Each names the chapter where its answer decided the result.

1. *What is judged at the end, and what during the run measures it?* Write the end result as a function of the path, as the dose of Section 1.4.1 is, before choosing a loop. If nothing in the run measures it, plan an estimate and a check on it (Chapter 8).
2. *Which variable carries the hazard?* Limit that variable, not the one the trend shows. The reactor's hazard was the unreacted reagent, and the temperature showed none of it (Chapter 2).
3. *Is the trajectory known well enough to run open loop?* A ramp searched on a correct model beat feedback, and on a wrong charge it collapsed to 1.26 mol. Re-optimize when the charge changes and trim only small errors (Chapter 3).
4. *What ends each phase?* A clock is blind to a slow batch, an event to a missing trigger and an estimate to its own model. Put a time limit behind each trigger (Chapter 7, Chapter 8).
5. *Does the disturbance repeat within a run, between runs, or not at all?* Learn what repeats along the run (Chapter 10), correct between runs what persists (Chapter 7), and leave the rest to feedback, which a correction can only make worse.
6. *How many results are in flight?* Count the runs between an action and its measurement, and either estimate from the record of each run or detune the gain (Chapter 9).
7. *Where does the plant switch, and what stops it chattering or bumping?* Choose a band, a dwell, a latch, tracking or a persistence time for each switch (Section 1.4.7), and check that a shared utility does not pull independent switches into step (Chapter 12).

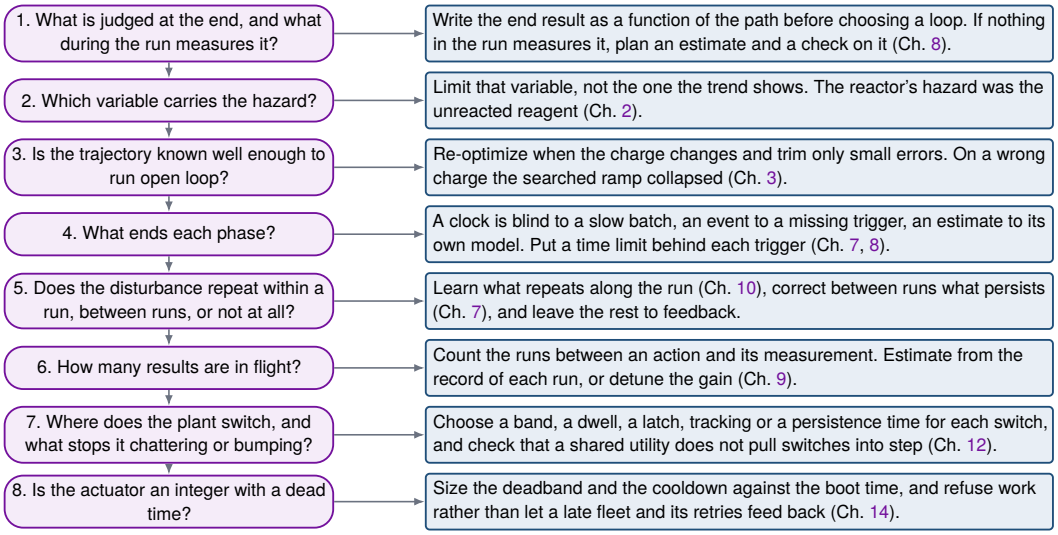


Figure 16.2 The eight questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that what is judged and what carries the hazard are settled before any loop is chosen, and that the last two questions are about switches rather than loops.

8. *Is the actuator an integer with a dead time?* Size the deadband and the cooldown against the boot time, and refuse work rather than let a late fleet and its retries feed back (Chapter 14).

A predictive controller earned its place where it planned around a limit the rules met by cutting. On warm cooling water it dosed the reactor with no cut where the one-step optimizer cut 21 times, for a batch 5.8 min longer (Table 2.6). On four freeze-dryer shelves it never lost the chamber pressure, which the endpoint-aware policy lost for 36.7 min, and moved the shelf 31 K against 482 K (Table 5.9). Elsewhere it traded one measure for another or lost. The crystallizer's batch fell from 187.0 to 126.9 min while fines rose from 13.8 to 15.6 % (Table 4.9). On the batch column it made 14.2914 mol against 14.3393 mol for a ramp searched on the run's model, with a reflux travel of 173.47 against 5.81 (Table 3.8). On the refrigeration pack it used 21.5184 kWh in three hours against the scheduler's 21.5574 kWh, with 24 rack starts against 8 (Table 12.8), and on the web service it billed $11.93 against the PI loop's $11.06 on the morning ramp (Table 14.7). Volume 9 treats that choice in full. Figure 16.2 draws the eight questions in order.

16.3 Where Practice Is Heading

Batch optimization moves to measurements. A review of batch control names the main difficulty as run-end outputs, such as final quality, that cannot be measured during the run, with process models that tend to be poor. It places the research push in measurement-based optimization that uses data from the current and previous batches [1]. The batch column and the coater of this volume show both halves of that statement. A ramp searched on the model was the best policy until the charge changed, and an estimate of thickness was right until the efficiency moved outside it. In the editor's view the recipe written once and run unchanged will give way to recipes that are re-planned from each batch's own measurements, with the safety layer left as it is.

Corrections between runs spread beyond the wafer fab. Run-to-run control was developed for semiconductor processes, combining response surfaces, statistical process control and feedback, and one review proposed a control chart on the output that acts as a deadband [2]. Run-to-run control for

batch processes has its own survey [3], and learning control, repetitive control and run-to-run control have been surveyed together [4]. A survey of iterative learning control treats stability, performance, learning transients and robustness, and describes a field still active after two decades [5]. In additive manufacturing, learning schemes now use data from previous layers and previously printed parts and are designed for plants that vary from one iteration to the next [6]. A review of metal additive manufacturing found that factory-ready machines had not yet incorporated monitoring that detects discontinuities during the build [7]. The powder bed of Chapter 10 shows why both matter: learning needs a measurement of every layer, and it fails where the repetition breaks.

Some batches become continuous. An article on modernizing pharmaceutical manufacturing states that continuous processing can address agility, flexibility, cost and robustness, supported by quality by design and process analytical technology, and that the US Food and Drug Administration supports its implementation with science- and risk-based approaches [8]. The antibody capture of Chapter 11 is already a cyclic schedule fed without a stop. In the editor's view the freeze-dryer and the vial filler will remain batch and lot steps for longer, and the work there is the estimate of what no probe measures and the record that ties each result to its vial.

Switching is designed, not tuned. A tutorial on hybrid dynamical systems sets out models, stability and control for plants that combine flows and jumps, with emphasis on robustness to perturbations and measurement error [9]. Synchronized display cases have been modeled as a hybrid system whose synchronized state is a periodic trajectory, and the analysis found that more measurement uncertainty makes synchronization less prevalent [10]. Predictive hybrid control has been applied to a supermarket refrigeration benchmark [11], and the autoscaling of elastic cloud services has a review of its own techniques [12]. In the editor's view the devices of this volume (band, dwell, latch, tracking, flush, budget) will stay, because they are simple enough to verify, and optimization will be placed above them, as the autoclave's predictive controller had to be.

Instruments that reach the product. A white paper on monitoring practice in pharmaceutical freeze-drying finds thermal-conductivity gauges and capacitance manometers both commonly used, though often not in the most useful way. It names wireless product temperature, laser absorption spectroscopy in the vapor duct and heat flux measurement as tools that may become best practice. Wireless sensors had appeared in the few years before it, avoid the lead wires that make probing awkward in manufacturing, and serve at laboratory and production scale [13]. A review of crystallization control finds most in situ image analysis quantitative only at low solid concentrations, typically at most 5 to 7 percent, and expects research on systems that work at higher ones [14]. The freeze-dryer of Chapter 5 reads two probe vials and two gauges, and the crystallizer of Chapter 4 reads an analyzer of the liquid. In the editor's view both would gain most from a measurement of the product itself.

Feedforward from the part's own geometry. Adjusting the laser power once per layer is a common strategy in powder bed fusion, and some commercial machines close photodiode feedback within a layer. A study of 2025 notes that such feedback stays reactive, cannot anticipate known geometry such as overhangs, and complicates certifying parts for critical applications when it changes parameters in real time. Its feedforward on each scan vector, built on a light finite-difference thermal model and a reduced melt pool model and tested on a complex part, cut geometric inaccuracy in key dimensions by 62 percent and porosity by 16.5 percent on average [15]. The learning of Chapter 10 reaches the corners by another road. It learns them from the last layer, where such a model predicts them before the first.

Clients that limit their own retries. A handbook of site reliability engineering describes a retry budget of up to three attempts per request and a budget per client that allows a retry only while

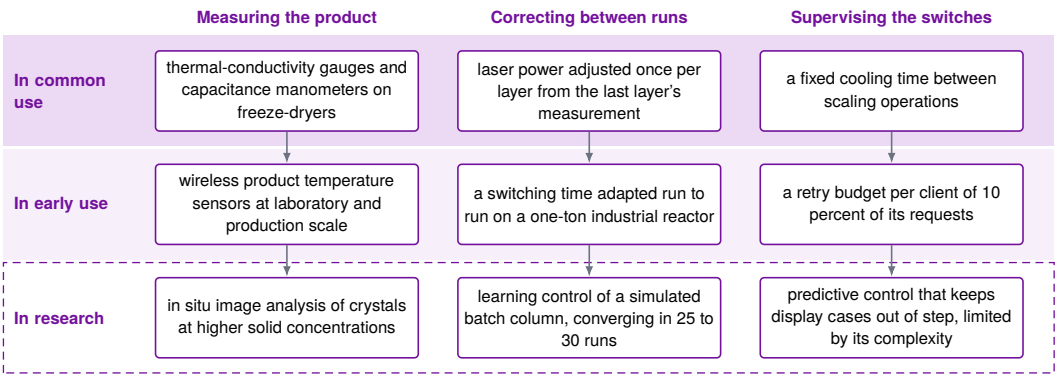


Figure 16.3 Developments in three threads. The top lane holds what the cited source calls commonly used, a common strategy or a common solution adopted in practice, the middle lane what it reports in service at production scale or at one industrial plant, and the bottom lane what it reports from simulation or names as a need, in the editor's reading. Sources by thread: measuring the product [13, 14], correcting between runs [15, 18], supervising the switches [16, 17, 19]. No date is drawn, because the sources give none for these steps.

retries are under 10 percent of its requests. In its worst case, a datacenter that rejects most requests, the limit per request alone lets the load grow to just below three times, and the budget per client holds the growth to about 1.1 times [16]. A survey of autoscalers finds a fixed cooling time between scaling operations a common solution adopted in practice, set to at least the time to acquire and boot a machine [17]. The web service of Chapter 14 carries both devices. Its retry budget of 10 % left no storm minute on any seed, and its cooldown of 300 s reversed no action on the evening shoulder. Figure 16.3 sets these developments and those above in three threads.

> ⧖ HISTORY **2017: Run-to-Run Optimization on a One-Ton Reactor**
>
> The history boxes of Chapter 1 end in 1995, with run by run control in chip making. A chapter of 2017 reported the idea on a one-ton industrial reactor. Each batch ran isothermal and then adiabatic, and the time of the switch was adapted from one batch to the next. The reaction time fell significantly after only two batches, without off-specification product, and ended 35 percent shorter than the isothermal policy of industrial practice. A back-off from the final temperature limit was kept for safety, and the policy became standard practice for the industrial partner [18]. A review of 2018 found the quality loop of injection molding the least developed of its three loops, its learning and predictive solutions good but not widely used [20]. In the editor's reading the chapters of this volume stand between those two reports.

16.4 Engineering With AI Agents

An AI agent is a language model that works toward a goal in a loop. It plans, calls tools such as a compiler, a simulator or a browser, reads the results and continues [21]. In the editor's view, agents can now carry out much of the work that used to fill the hours of a control project. They derive a model from balances or fit one to plant data. They write the simulator and its tests, code the controller, tune it against stated criteria, script commissioning checks, write the documentation and watch running loops for degradation. The apps and chapters of this series were produced that way under the editor's direction, and the account that follows this section reports what the agents got wrong in this volume and how it was caught.

In the editor's view, this changes where an engineer's value lies. When a model, a controller and a report can each be produced in minutes, producing them is no longer the scarce skill. The work moves to five things.

Quality control. An agent's output arrives as a claim: this model fits, this tuning settles in four minutes, this code implements the equation. The engineer checks claims against the plant or the simulator. A reported number is measured again on a fresh run at the stated seed. The code is read for what it clamps, what it initializes and what it reads, because a display that shows the controller's model in place of the plant looks correct until the two differ. A tuning is tried on a disturbance the agent did not test. The most useful checks are independent of the author, made by a second agent or a person who has not seen the first answer, and they are checks that can fail.

Supervision. The engineer sets the scope of each task and decides what an agent may touch. Reading a historian and proposing a tuning is one level of access. Writing a setpoint, downloading a configuration or changing an interlock is another, and it passes through the plant's management of change with a named approver. Supervision also fixes the order of work: a baseline before a change, one change at a time and a comparison that shows what moved.

Creativity. An agent answers the question it is given. Choosing the question remains the engineer's task: which loop is worth improving, whether a disturbance calls for feedforward, a cascade or a change to the process, which step test the operation can tolerate, and which comparison would show that a predictive controller earns its place over a retuned PID loop. The experiment that separates two explanations is designed by someone who can imagine both.

Process insight. A model contains what its author put in. The equipment does more. A valve sticks, the dead time of an analyzer grows when its sample flow drops, a thermowell lags, a heat exchanger fouls and an operator intervenes at the same hour each night. The engineer who knows the unit recognizes a result that is too good, asks what the model left out and knows which measurement to distrust first.

Accountability. Responsibility does not transfer to a tool. Functional safety standards assign the activities of the safety lifecycle to persons and organizations of stated competence, and they require an assessment that includes someone independent of the design team [22]. An agent may draft the hazard analysis or the test procedure. A person owns the decision to accept it, to start up and to keep running.

These five are the professional form of the habit introduced in Section 1.7: check before trusting, and know enough to check.

16.4.1 How This Volume Was Made

Each of the fourteen apps was built and tested by agents. An agent then wrote the paper on it, keeping verification notes that trace every number to a run, and a referee report was written on each paper. At least nine of the fourteen reports are self-reviews by the agent that wrote the paper. Six apps of the later rounds passed usability, design and technical-accuracy reviews before their papers. Defects went into a log for each app with a severity, a cause, a suggested fix and a status. For this book one agent per chapter reran the numbers its exercises use. The editor set the scope of each round and the rules on what each agent could change, and statements that runs contradicted were left for the editor to decide.

The defect logs. The fourteen logs written beside the papers hold 104 entries with a severity. Of these, 32 are wrong behavior or output of an app, 49 are documentation, 13 are cosmetic and 10 are other kinds, such as a test allowance or a scan period. The other eight apps account for 91 entries. The six that had a technical review before their papers account for 13. Their reviews before the paper had already logged and fixed findings in each app's own record. Five cases show the errors that matter for this volume's topic.

A batch graded off specification when it was run well. The batch column's scorecard compared the receiver's true average with the specification at a tolerance of 10^{-6}. The tail cut closed the drum on the analyzer's running average, which lags the column, so a well-run batch landed 3.2×10^{-4} below 0.990 and the policy the app ships as the winner, 15.735 mol at 0.98968, was graded a failure. The app's own tests already used a tolerance of 10^{-3}. The paper agent found it while writing the paper, and one shared test now grades all five places that judge the drum.

Switches that were not bumpless. On the same column the composition loop kept integrating during the total-reflux startup, when the plant could not answer it, and collection opened at a reflux ratio of zero. The distillate fell to 0.9462. With the loop tracking the ratio it will open on, the dip is to 0.9896. On the crystallizer, supersaturation control entered with a proportional kick of 5.5 °C on its setpoint, and the result depended on how the learner reached it. With 6 g of seed two routes finished at 285 and 293 min. Treated as a transfer, the batch moves the valve 66 % instead of 153 %. The crystallizer's log records that its entries were found by reading the source line by line against the numbers the app quotes.

A headline that hid where it came from. The freeze-dryer's endpoint-aware policy stoppers 1.38 h before a recipe tuned on the same batch. The referee asked where the time came from, and 1.07 h of it lies in secondary drying, the weaker half of the model (Section 5.6). The paper and the app now say so. The vial filler's first draft claimed that the record-based estimate has an advantage when the gain must be high. The lot-step test gave 2093 against 2094 accepted, and the claim was rewritten. The molding paper stated the variance law $2\sigma^2/(2-k)$ before it was measured, and a 195-part run was added to test it.

A predictive controller less safe than the selector. The first autoclave MPC kept only the max loop above its plan and set its ceilings without a margin for the reaction heat. A technical review ran 46 worst cases and found it worse than the lead/lag selector with feedforward on the selector's limit in 12 of them, with one thermocouple 3.15 °C over for 4.8 min. The cause was the model's heat of reaction. The fix was chosen by a rule written before any candidate ran and amended once after the first results. It keeps the feedforward and the lead clamp above the plan, adds a reaction margin to the ceilings and moves the optimized recipe from 281.0 to 291.3 min (Table 6.9). A headline of 281.0 min would otherwise have credited the MPC with minutes it earned by planning to the limit on its own model.

Controllers that read the plant's truth. In three apps a controller took values it could not measure on a real plant, and a fourth log documents one more. The reactor's calorimeter, its heat transfer law, the MTSR limiter and the optimizer took fifteen constants from the run's own parameters, so a weaker jacket entered in Parameters looked safe. The limiter dosed in 354.0 min with the true MTSR at 80.51 °C, against 222.1 min and 104.09 °C when the same jacket came from the disturbance slider. The autoclave's cure observer knew a learner's changed resin, and with it a thick part peaked at 202.6 °C instead of the 215.2 °C it reaches on the declared values. The juggler's impact switch reported where the ball truly hit, up to 4.9 mm from the camera's estimate. The capture train's surge hold reads the true tank volume, documented and left as a plant-side interlock. The reactor and autoclave fixes moved no shipped run, and the juggler's moved two cases. The reactor and autoclave entries record the probe behind the shared section's advice to read code for what it reads. A hidden value is changed, and the question is whether the controller moves while the measurements do not.

The book round. The fourteen chapter records list 207 rows of numeric verification, each a stated value or group of values set beside its reproduction. No reported number failed to reproduce, apart from one sweep. On the coater three last digits of a spread sweep differ when the setpoint is set before

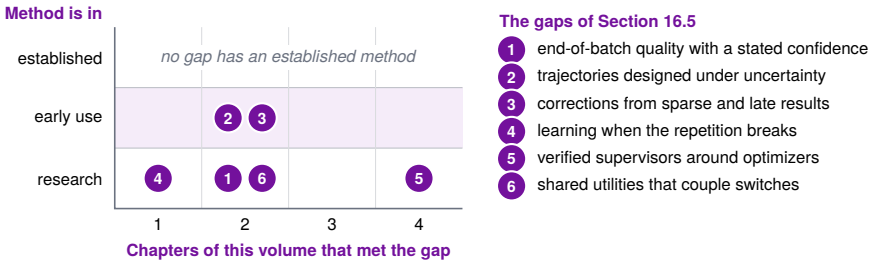


Figure 16.4 The six gaps of this section. Across is the number of chapters that each item names as evidence: two for gap 1 (Chapters 8 and 5), two for gap 2 (3 and 4), two for gap 3 (9 and 7), one for gap 4 (10), four for gap 5 (6, 13, 2 and 11) and two for gap 6 (12 and 14). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Notice that the gap met by the most chapters, the verification of the logic around an optimizer, has the least mature method.

Run rather than during the opening purge, a path dependence the exercise now states as a range. The book round's defect list holds 13 entries for this volume, five of them check counts that grew after the papers. The molding paper called both loops direct acting, against the convention of Section 1.6.2, and the autoclave paper used "maximum selector" loosely. Both are corrected in the chapters and have since been fixed in the papers. The agent writing Chapter 1 first stated that a deadband around an integer actuator turns the PI loop into a thermostat with a band of $\pm 7.5\,^{\circ}\mathrm{C}$. The simulation gave $\pm 2.6\,^{\circ}\mathrm{C}$, because the integral moves too, and the formula was dropped.

Every error above was caught by a check that could fail: a second measurement, a rerun along another path, a decomposition of a headline, a sweep of worst cases, a hidden value moved to see what else moved. Most sat at a boundary that this volume is about, an endpoint, a switch or a correction from a record. Four working rules follow. Judge the end result on the truth and the estimate separately. Test every switch along every route into it. Decompose every comparative claim before it is printed. Keep the protections above any optimizer, and let a person decide what changes.

16.5 Gaps in Batch, Run-to-Run and Hybrid Control

Six gaps belong to this volume's topic, and the gaps common to agent-era practice follow them in Section 16.5.1. In the editor's view none is closed. Each item gives the chapters' evidence, then the state of the literature with its sources. Why each gap is hard and what progress would look like are the editor's view. Figure 16.4 places the gaps by the chapters that met them and by the maturity of the method available, and the same numbers mark the plants in the picture that opens this chapter.

1. *End-of-batch quality with a stated confidence.* The coater's load cell closed one gap and could not see a wrong charge, and the freeze-dryer's learned resistance failed on a gauge error of 6 %. A chapter on batch control warns that estimates of infrequently measured product properties can never completely replace the measurements themselves [18]. A retrospective of a coated particle fuel program asks for measurement techniques deployable in an industrial setting, since its buffer density is measured by interrupting the coating and gives only an average [23]. The gap is hard because the error that matters comes from a disturbance no instrument in the batch sees. An estimator that states, during the batch, which disturbance it cannot see and how large an error it would miss is not available.

2. *Trajectories designed under uncertainty.* The column's ramp collapsed on a wrong charge, and the crystallizer's predictive batch bought an hour with more fines. Optimization on a model computes inputs optimal for the model, often fails to reach plant optimality, and in practice gives suboptimal and often infeasible operation. In crystallization and polymerization an off-specification batch

often cannot be corrected at all [18]. The gap is hard because each batch is one draw from the spread and cannot be repeated to average it out. A routine method that searches a recipe over the spread of charges and kinetics, and reports what the spread costs, remains an open engineering problem.

3. *Run-to-run correction with sparse and delayed results.* With one group of vials in four weighed, no correction passed the challenge, not even ideal information. The molding correction widened the spread of what did not persist by 15 %. A review of injection molding finds the loop that feeds part quality back the least developed of three, and its learning and predictive solutions good but not widely used for lack of standardization [20]. The gap is hard because the weighing plan, the delay and the gain are set by different people for different reasons. Design rules that tie them to the fraction of product released untested are thin.

4. *Learning when the repetition breaks.* A learned profile kept at a new outline left 13.2 % of the first layer defective, and one reset left 23.2 %. A study of powder bed control finds feedback reactive to geometry it could know in advance, and real-time changes of parameters a complication for certifying parts [15]. The gap is hard because what was learned belongs to the shape it was learned on. Methods to transfer it to a new shape, a new machine or a new powder lot, with a guarantee during the first layers, are needed.

5. *Verified supervisors around optimizers.* The autoclave's first predictive controller was less safe than the selector in 12 of 46 worst cases, and the pasteurizer's flush, the reactor's latch and the capture train's ready signal were all checked by simulation alone. The Open challenge box at the end of this section treats this gap.

6. *Shared utilities that couple switches.* Four thermostats locked in step through one suction pressure, and a capped web fleet held its clients in a retry storm for two hours. A thesis that kept two display cases apart by predictive control found the method limited by its growing complexity and computation time on larger systems [19]. A workshop paper on metastable failures, in which a bad state persists after its trigger is removed, calls a systematic approach to robustness against unknown ones an open problem [24]. The gap is hard because the coupling appears only when many units act together. Design rules that predict it from the plant's structure, before it appears in operation, are needed for plants with many switched units on one header or service.

OPEN CHALLENGE Verifying the Logic Around an Optimizer

Gap 5 runs through four chapters, each with protections checked by simulation alone. The run-to-run reactor policy kept a back-off from its temperature limit for disturbances during the run that the method cannot handle [18]. A juggling robot was given an open-loop stroke because no verification process for complex closed-loop network policies was known to exist [25]. Hybrid systems theory states robust stability for plants that flow and jump [9]. In the editor's view the logic is hard to verify because interlocks, switches and an optimizer act on one plant at once. Progress would be a check of the combined logic, run before commissioning, that fails when a protection is removed.

16.5.1 Gaps Common to Agent-Era Practice

The final chapter of each volume lists gaps specific to its topic. Seven more are common to every topic in the series. In the editor's view none of them is closed.

1. *Verifying agent-written control code and models before they reach a plant.* Tests show that a controller behaves on the cases tested. For controllers learned from data, guarantees of stability and constraint satisfaction rest on assumptions about the model and its uncertainty, and open problems remain [26]. In the editor's view, code written by an agent has no comparable body of method yet. Practice needs a routine path from an agent's draft to code a plant can trust:

independent derivation, tests of what the code reads and clamps, and trials with hardware in the loop.

2. *Trustworthy models from sparse or poor plant data.* Plant records are short, taken in closed loop, noisy and marked by unmeasured disturbances. A fitting routine returns a model from any record. What is missing is a dependable statement, delivered with the model, of where it can be trusted, and a method that says so when the data cannot identify a parameter.

3. *Safety cases and standards for agents in commissioning and operation.* Functional safety standards define a lifecycle of specified requirements, verification and assigned human roles [22]. In the editor's view, how an agent that writes configuration, runs commissioning checks or advises an operator is qualified, limited and audited under them is not yet settled.

4. *Security of agent access to control systems.* An agent with credentials to a control system is a new path into it. Industrial security standards state the requirements a control system must meet at each security level [27]. An agent that reads documents, messages or web pages can also be steered by text placed there. In the editor's view, least privilege, read-only access by default and human approval of every write are a starting point and not a solution.

5. *Benchmarks and reproducibility.* An agent asked twice may deliver two different controllers, and a result reported without its plant, seed and scoring cannot be checked. Control needs shared benchmark plants with fixed disturbances and published measures, on which agent-built and conventional designs are compared like for like. The apps of this series, with a default seed and an exported record of every run, are a start.

6. *Transfer from simulation to hardware.* A controller developed on a simulator inherits what the simulator omits. The robotics literature documents the gap between simulated and real performance and the methods used to narrow it, such as randomizing the simulated plant and adapting on the real one [28]. Process plants add slow dynamics, rare upsets and little freedom to experiment, so each transfer still needs a staged plan from the simulator to a bench rig to the unit.

7. *Keeping engineers' judgment sharp when agents do the routine work.* Automation removes the routine work through which skill is built and leaves the person the hard cases, an irony described for process operators in 1983 [29]. The same now applies to engineers. An engineer who has never tuned a loop by hand or closed a balance on paper is poorly placed to review an agent that has. Deliberate practice, of the kind the exercises in this book provide, has to be designed into education and into work.

KEY POINTS

- The path a batch followed decided its product, not how closely the loop tracked that path.
- An endpoint is only as good as the estimate behind it. The guard that worked read a second, independent measurement.
- A correction between runs pays only for what persists, and learning fails where the repetition breaks.
- Simple devices outside the controller kept every switch safe, and an optimizer earned its place only with the protections kept above it.

References

[1] D. Bonvin, B. Srinivasan, and D. Hunkeler, "Control and optimization of batch processes," *IEEE Control Systems Magazine*, vol. 26, no. 6, pp. 34–45, 2006.

[2] E. Del Castillo and A. M. Hurwitz, "Run-to-run process control: Literature review and extensions," *Journal of Quality Technology*, vol. 29, no. 2, pp. 184–196, 1997.

[3] K. Liu, Y. Chen, T. Zhang, S. Tian, and X. Zhang, "A survey of run-to-run control for batch processes," *ISA Transactions*, vol. 83, pp. 107–125, 2018.

[4] Y. Wang, F. Gao, and F. J. Doyle, "Survey on iterative learning control, repetitive control, and run-to-run control," *Journal of Process Control*, vol. 19, no. 10, pp. 1589–1600, 2009.

[5] D. A. Bristow, M. Tharayil, and A. G. Alleyne, "A survey of iterative learning control," *IEEE Control Systems Magazine*, vol. 26, no. 3, pp. 96–114, 2006.

[6] Z. Afkhami, D. J. Hoelzle, and K. Barton, "Robust higher-order spatial iterative learning control for additive manufacturing systems," *IEEE Transactions on Control Systems Technology*, vol. 31, no. 4, pp. 1692–1707, 2023.

[7] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, "Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing," *Materials & Design*, vol. 95, pp. 431–445, 2016.

[8] S. L. Lee, T. F. O'Connor, X. Yang, C. N. Cruz, S. Chatterjee, R. D. Madurawe, C. M. V. Moore, L. X. Yu, and J. Woodcock, "Modernizing pharmaceutical manufacturing: From batch to continuous production," *Journal of Pharmaceutical Innovation*, vol. 10, no. 3, pp. 191–199, 2015.

[9] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Systems Magazine*, vol. 29, no. 2, pp. 28–93, 2009.

[10] J. Leth, R. Wisniewski, J. G. Rasmussen, and H. Schiøler, "Stochastic analysis of synchronization in a supermarket refrigeration system," *Electronic Proceedings in Theoretical Computer Science*, vol. 174, pp. 40–50, 2015.

[11] N. L. Ricker, "Predictive hybrid control of the supermarket refrigeration benchmark process," *Control Engineering Practice*, vol. 18, no. 6, pp. 608–617, 2010.

[12] T. Lorido-Botran, J. Miguel-Alonso, and J. A. Lozano, "A review of auto-scaling techniques for elastic applications in cloud environments," *Journal of Grid Computing*, vol. 12, no. 4, pp. 559–592, 2014.

[13] S. Nail, S. Tchessalov, E. Shalaev, A. Ganguly, E. Renzi, F. Dimarco, L. Wegiel, S. Ferris, W. Kessler, M. Pikal, G. Sacha, A. Alexeenko, T. N. Thompson, C. Reiter, J. Searles, and P. Coiteux, "Recommended best practices for process monitoring instrumentation in pharmaceutical freeze drying—2017," *AAPS PharmSciTech*, vol. 18, no. 7, pp. 2379–2393, 2017.

[14] Z. K. Nagy and R. D. Braatz, "Advances and new directions in crystallization control," *Annual Review of Chemical and Biomolecular Engineering*, vol. 3, no. 1, pp. 55–75, 2012.

[15] N. Kirschbaum, N. Wood, C.-E. Kim, T. U. Tumkur, and C. Okwudire, "Vector-level feedforward control of LPBF melt pool area using a physics-based thermal model," *Additive Manufacturing*, vol. 112, p. 104981, 2025.

[16] B. Beyer, C. Jones, J. Petoff, and N. R. Murphy, Eds., *Site Reliability Engineering*. O'Reilly Media, 2016.

[17] C. Qu, R. N. Calheiros, and R. Buyya, "Auto-scaling web applications in clouds: a taxonomy and survey," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–33, 2018.

[18] D. Bonvin and G. François, "Control and optimization of batch chemical processes," in *Coulson and Richardson's Chemical Engineering*. Elsevier, 2017, pp. 441–503.

[19] L. F. S. Larsen, "Model based control of refrigeration systems," Ph.D. dissertation, Department of Control Engineering, Aalborg University, Aalborg, Denmark, 2006, iSBN 87-90664-29-9. [Online]. Available: https://vbn.aau.dk/ws/files/549479206/thesis.pdf

[20] O. Ogorodnyk and K. Martinsen, "Monitoring and control for thermoplastics injection molding: A review," *Procedia CIRP*, vol. 67, pp. 380–385, 2018.

[21] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024.

[22] International Electrotechnical Commission, "Functional safety – Safety instrumented systems for the process industry sector – Part 1: Framework, definitions, system, hardware and application programming requirements," IEC 61511-1:2016, Edition 2.0, 2016.

[23] P. A. Demkowicz, D. A. Petti, J. D. Stempien, J. Palmer, W. F. Skerjanc, D. M. Scates, G. Hawkes, T. Gerczak, and G. Helmreich, "Advanced gas reactor fuel development and qualification program retrospective and lessons learned," Idaho National Laboratory, Idaho Falls, ID, Report INL/RTP-25-06381, Revision 0, 2026, oSTI 17429367.

[24] N. Bronson, A. Aghayev, A. Charapko, and T. Zhu, "Metastable failures in distributed systems," in *Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS '21)*, 2021, pp. 221–227.

[25] K. Ploeger, M. Lutter, and J. Peters, "High acceleration reinforcement learning for real-world juggling with binary rewards," in *4th Conference on Robot Learning (CoRL 2020)*, 2020, arXiv:2010.13483.

[26] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411–444, 2022.

[27] International Electrotechnical Commission, "Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels," IEC 62443-3-3:2013, Edition 1.0, 2013.

[28] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: A survey," in *Proc. 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 737–744.

[29] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

16.6 Exercises

16.1 Discussion · 8 min

On the batch column of Chapter 3 a ramp searched on the model collected 15.73 mol, constant-composition control 15.29 mol and the best constant reflux 15.10 mol. An agent proposes to ship the open-loop ramp on every charge because it won. Argue for and against, and say what you would require before the ramp runs without feedback.

16.2 Discussion · 8 min

An agent tuning the correction between runs on a new filling line proposes a gain of 0.50 because "a faster correction removes drift sooner." Using Chapter 7 and Chapter 9, explain when a higher run-to-run gain helps and when it hurts, and list what you would ask the agent to measure first.

16.3 Discussion · 7 min

The coater of Chapter 8 ended its layers on a timer, on an integrated estimate and on the estimate corrected by a load cell. For each guard, name a disturbance it handles and one it misses, using the chapter's results, and propose a guard for a new batch that has no load cell.

16.4 Discussion · 8 min

An agent replaces the autoclave's lead/lag selector with a predictive controller and reports the optimized recipe 8.8 min faster. Using Section 16.4.1 and Chapter 6, describe the review you would run before accepting it, and say where the protections should sit relative to the optimizer.

16.5 Discussion · 6 min

The freeze-dryer's endpoint-aware policy stoppered 1.38 h before a recipe tuned on the same batch, and 1.07 h of that came from secondary drying. Explain why such a decomposition belongs in any comparison an agent writes, and propose two other checks of a comparative claim, using examples from Section 16.4.1.

16.6 Discussion · 7 min

In the reactor of Chapter 2, a weaker jacket entered in Parameters let the MTSR limiter dose in 354.0 min with the true MTSR at 80.51 °C, while the same jacket from the disturbance slider gave 222.1 min and 104.09 °C. Explain what this difference reveals, why no shipped run showed it, and design a test that would catch the same defect in another batch app.

Answers to Selected Exercises

Short answers are given for the odd-numbered exercises. Numeric answers from an app are for its default seed and settings unless the exercise says otherwise.

Chapter 1. Introduction

1.1 (a) 18.97 min. (b) 1.09 min. (c) 94.9 min.

1.3 (a) $\tau_c = 4$ min, lag 6 °C. (b) 1.5 elements.

1.5 (a) -0.432. (b) -0.5. (c) 0.559. (d) 0.4 per batch, $\lambda < 1.33$.

1.7 (a) 0.5 at zero frequency and 0.9875 at the shortest period. (b) -0.5: a constant error halves and changes sign each batch.

1.9 (a). The thermostat's duty rises to 0.67 by itself, while the fixed duty settles at 73.1 to 74.9 °C.

Chapter 2. Semi-Batch Reactor Dosing as the Safety Layer

2.1 (c). MTSR is the batch temperature plus the adiabatic rise of the unreacted B, and the cold batch holds most of what it has been fed.

2.3 (a). The heat release is spread out, so the temperature shows no symptom while the unreacted B climbs to 765 mol.

2.5 (a) 248 mol and 76.6 °C, against 234 mol and 76.2 °C from the engine. (b) About 64 000 mol, so the estimate does not apply: the reaction cannot keep up and nearly all the feed accumulates. (c) $\Delta T_{ad,acc} = RT_K^2/E_r = 8.0$ K at 67 °C.

2.7 (a) The batch is lost at 3 h 55 min, with at most 4398 mol unreacted. (b) 74.6 °C, class 2 throughout, 276 kg fed against 750 kg on the fixed profile. The outcome was set by the B already in the batch when the feed stopped.

2.9 MPC: dose 3 h 38 min, complete 3 h 52 min, highest T 71.9 °C, no cut. Optimizer: dose 3 h 28 min, complete 3 h 46 min, highest T 73.2 °C, 21 cuts. The optimizer finishes 6 min sooner by letting the batch run above the cut.

2.11 (a) 127.9 °C, just below T_{D24}. (b) About 1376 kg. The liquid volume grows from 4.08 m^3 to 5.19 with the shipped quench and to 5.61 m^3, 94 % of the vessel, with the larger one.

Chapter 3. Reflux Policy on a Batch Distillation Column

3.1 (b). The objective exists only at the terminal time: product moles in the receiver, graded on the cumulative average of the drum.

3.3 (d). With an instantaneous rule no purity above 0.990 earns anything, so holding the distillate exactly at the specification is optimal by construction.

3.5 (a) 0.7072 mol/min. (b) 0.1010 mol/min and 15.15 mol. (c) 429 and 412 kJ/mol.

3.7 (a) 9.16 mol, 13.99 mol of slop, tail cut at 81.8 min, stop at 161 min. (b) 15.10 mol at 0.9895, 0.06 mol of slop. At 6 the receiver reaches the specification just as the clock runs out.

3.9 (a) 1.26 mol, tail cut at 36.5 min. (b) 3.10 mol. (c) 10.69 mol on a ramp from 3.40 to 17.83. Measure the charge and re-plan before collecting.

3.11 The divider cannot follow a reflux demand that climbs to the limit, and a 5 min lag costs product in the loop and lets off-specification material into the drum at the cut. Run a searched reflux staircase open loop and check the drum by laboratory assay, or infer the composition from a tray temperature.

Chapter 4. Batch Crystallizer Trajectory Control

4.1 (b). The cubic batch makes 373.4 µm at 13.8 % fines with the larger peak deviation and integrated error.

4.3 (d). Primary nucleation goes as σ^5, so the height of an excursion matters far more than its length.

4.5 (a) 1.02 µm/min. (b) 3.973×10^{-3} kg/(kg °C) and 0.0032 °C/min. (c) Linear 0.156 °C/min (48 times), natural 0.724 °C/min (224 times). Cubic from 0 to 0.467 °C/min.

4.7 (a) 13.434 °C min, 159 µm, 96.3 %, 179 min. (b) 15.065 °C min, 373 µm, 13.8 %, 187 min. (c) The cubic curve matches the cooling demand to the growing crystal surface, and an accelerating setpoint is harder for a PI loop to follow.

4.9 (a) 127 min, 372 µm, 15.6 %, 277.6 %. (b) 187 min, 373 µm, 13.8 %, 66.4 %. (c) The scorecard misses valve movement between log samples.

4.11 (a) About 62.9 % and 68.3 %, where the app measures 62.2 and 67.7 %. (b) The batch is unreachable once the utility is at 18 °C: the valve is pinned for 133 min and the run stops at 300 min at 65.0 %. Nothing looks abnormal on the trend.

Chapter 5. Hottest and Slowest Vials in Freeze-Drying

5.1 (a). The outer edge peaks 1.1 K under the limit and is dry at 16.4 h, the center stays 2.3 K under and is dry at 19.6 h.

5.3 (c). $\hat{p}_w$ is $0.7/0.6 = 1.17$ times the true vapor, $\hat{r}$ settles at 0.52, and every group collapses, 2.12 K over for 404 K min.

5.5 (a) 0.161 Pa. (b) 0.654 kg/h. (c) 18.1 h for four shelves and 13.6 h for three. Four shelves cannot meet 18 h at 10 Pa on arithmetic alone.

5.7 (a) 21.2 h, 23.6 h, 25.1 h, 50.0 h, closest 3.7 K under. (b) 25.4 h and 35.4 h. The last 14.9 h of primary drying dried nothing.

5.9 (a) 21.16 h, pressure held. (b) 21.81 h, pressure lost for 37 min. (c) 21.11 h, pressure held. The gain is the pressure setpoint, not the horizon.

5.11 Only 0.31 h of the gain is in primary drying, the part calibrated on measured data. The other 1.07 h rests on secondary-drying kinetics borrowed from another formulation and a teaching bound water. Require a secondary-drying calibration for the product, a check of the Pirani law and duct, and pilot runs.

Chapter 6. Lead/Lag Control of an Autoclave Cure

6.1 (d). The core passes 195 °C while the hottest thermocouple reads 187.3 °C, and its heat leaves only through half the laminate.

6.3 (b). After 6 h at 180 °C the cure stands at 0.8992, because the diffusion rate falls below the kinetic rate.

6.5 (a) 185.5 K and 216.4 K, a sixth higher. (b) $\hat{s} = 0.370$ K/s and $\dot{\alpha} = 2.0 \times 10^{-3}\,\mathrm{s}^{-1}$. (c) $K_{ff}\widehat{\Delta T} = 262.7$ K exceeds the cap, so $T_F = 185 - 80 = 105$ °C. The bound is about 10.5 times the true rise.

6.7 (a) 463 min, peak 186.1 °C, α 0.896, 4 min outside the band. (b) 439 min, 202.2 °C, 0.906, 124 min. (c) The lagging spar sets the ramp and the hold clock, and the leading skin, held by the max loop, sets the limit.

6.9 (a) Fails, 438 min and 78 min outside the band. (b) Passes, 290 min, 189.3 °C, 0.878. (c) 180 °C for 30 min, passes in 301 min at 192.6 °C. The feedforward keeps the spar's exotherm away from the limit, so the optimizer's 2 °C margin holds at 185 °C.

6.11 (a) 148.5 min, 33.9 % of the cycle, and 4.97 cycles a day against 3.29. (b) Any two of these: edges, corners and ply drops of a thick part, which run hotter than a one-dimensional part, the flow field with its slow positions and radiation, a specific heat that rises with temperature, and resin flow and compaction.

Chapter 7. Injection Molding Switchover and Shot Weight

7.1 (c). The part is decided in the pack, and the extra flow resistance lowers the cavity pressure while the gate is open.

7.3 (a). For an uncorrelated disturbance the integral correction raises the variance by $2/(2-k)$ at every gain.

7.5 (a) 24.8 MPa. (b) 4, 8, 12 and 15 MPa, first in tolerance the sixth part (23.871 g predicted). (c) 23.967 g, inside the tolerance but 0.13 g light. The measured 25.4 MPa implies a sensitivity of about 0.0133 g/MPa in the thicker melt.

7.7 (a) SHORT, 23.412 g, 98.1 %. (b) OK, 24.038 g. (c) FLASH, 24.560 g, 93.8 MPa logged. (d) OK, 24.150 g, 56.6 MPa.

7.9 (a) Not met, 9 of 30. (b) Not met, 26 of 30, the correction pinned at 15.0 MPa. (c) Met, 30 of 30, 30.6 MPa in force on the last part. The drift and the thicker resin need more than 15 MPa, and the correction peaks at 34.2 MPa.

7.11 About 45 MPa/g. The default, $k = 0.55$, is stable but close to the bound, and a smaller gain or the record-based form is safer.

Chapter 8. Inferential Control of TRISO Fuel Coating

8.1 (c). The load cell weighs the coater, so its reading follows the bed mass m_b through a filter. Everything else in the list is computed.

8.3 (a). A miscount scales the surface area and the mass gain together, so the correction, which compares gains, sees nothing to correct.

8.5 (a) $N = 280\,300$, $A_b = 0.220\,\mathrm{m}^2$. (b) $H = 7.16\,\mathrm{mm}$, $U_{ms} = 0.228\,\mathrm{m/s}$. (c) 18.4 slpm.

8.7 (a) 26.22 µm, a 59.5 min layer, rejected on the SiC mean. (b) 35.03 µm, an 81.1 min layer, accepted in 139 min. (c) 0.734, close to the rate ratio of 0.740. The measured temperature T_m in the kinetic branch slows the estimate.

8.9 (a) 35.11 µm and 2.89 µm, rejected on SiC variability. (b) 2.06 µm at 1.50, rejected; 1.86 to 1.87 µm at 1.55, accepted. (c) The endpoint holds the mean. Circulation sets the spread, and it is fully developed at $R_{good} = 2.00$.

8.11 The load cell's mass gain against the modeled gain; the load cell's opening reading against the ticket; and the bed pressure drop, which collapses when the bed slumps.

Chapter 9. Vial Filling With Late Checkweigher Results

9.1 (b). The pooled law holds the average, so head 3 ends 10.4 mg under while heads 1, 2 and 4 end 6.7, 3.4 and 0.9 mg over.

9.3 (d). With two purge shots per head the first fills are 9.0 mg under to 2.2 mg over.

9.5 (a) $\sigma_e^2 = \sigma_p^2 + 2\sigma_r^2$, $\sigma_e = 5.31\,\mu\mathrm{L}$, $\rho_1 = -\sigma_r^2/(\sigma_p^2 + 2\sigma_r^2) = -0.356$. (b) 5.30 mg. (c) 6.86 mg.

9.7 (a) −10.4 mg on head 3, +6.7, +3.4, +0.9 mg on the others, 32 underfills accepted. (b) −0.7 mg, 3 accepted, 18 held. (c) One command against four.

9.9 (a) 1598, 20.9 mg, 92 under and 18 over, 62.0 vials/min. (b) 2329, 13.3 mg, none outside, 55.8 vials/min. (c) 2393, 12.9 mg, 6 underfills accepted.

9.11 (a) 11.7 mg, a command 11.4 µL lower. (b) A true underfill of about 10 mg with 158 underfilled vials accepted.

Chapter 10. Iterative Learning in Powder Bed Fusion

10.1 (a). By the time the photodiode sees the pool widen and the reduced power arrives, the corner is mostly over.

10.3 (c). The learned profile is wrong only where the outline changed. It still cuts power at the three corners that stayed and for the heat the part has gathered.

10.5 (a) 10.41 J, 141.6 K. (b) 48.7 K, factor 1.073. (c) 149.8 µm against the 150 µm of scenario 1, and about 121 µm on a straight edge.

10.7 (a) 4.10 %. (b) 2.57 % at 1 and 2.93 % at 2. (c) 8.25 % at 3 and 2.66 % at 1. The loop turns spatter into power in proportion to its gain.

10.9 (a) Layer 4 at 5.34 µm; layer 25 at 17.03 µm and 24.79 %. (b) 2.83 µm with no defects after layer 2; 4.70 µm with defects through layer 8 and small ones on layers 19 and 24. (c) 1.65 µm. A one-bin lead is closer to the true lag, so the short wavelengths are not overcompensated.

10.11 Keep feedback underneath the learning, at a lower gain than it would have alone. Keep the learned profile when the outline changes, or index it by feature, rather than discard it. Always run a Q-filter, chosen as insurance against an unknown lag.

Chapter 11. Antibody Capture on Resin of Unknown Capacity

11.1 (d). The lag catches 78.3 instead of 21.8 g, and the loss moves from 0.06 to 0.09 %.

11.3 (b). The stale baseline reads background as breakthrough: 20 switches instead of 14 and 0.47 L/g of buffer instead of 0.34, at 0.04 % lost.

11.5 (a) 55.25 and 45.5 g/L, 27.8 g per column. (b) At most 1.18 g/min against 2.1 g/min delivered, so the surge tank fills after about 209 min, before the harvest is through. (c) 4.

11.7 (a) 0.06 %, 21.8 g, 12, 0.30 L/g. (b) 0.09 %, 78.3 g, 11, 0.28 L/g. The lag caught 56.5 g more while about 0.1 g more was lost. The deeper switch bought resin and buffer.

11.9 (a) 38.4 g/L/h with 4.62 % lost, not met. (b) 38.3 g/L/h with 0.71 %, met. (c) 38.7 g/L/h with 3.95 %, not met. The threshold and the pause are both needed, and the pause matters more.

11.11 (a) About 0.53 g in the bed and 0.45 g in the line. (b) About 3.1 %, close to the 2.95 % measured in the wash. A wash into the loading zone recovers it.

Chapter 12. Thermostat Synchrony in a Refrigeration Pack

12.1 (d). The cases are coupled through the one suction pressure: an opening raises T_s and cuts the boiling heat of the cases already cooling, which holds back the case ahead.

12.3 (b). The goods exchange heat with the air through a two-hour lag, and the air never leaves its ordinary cycling.

12.5 (a) Above: $a(\Delta T - SH_{mss})$. Below: GSH_{mss}/λ_{mss} with $a\lambda_{mss} = -\ln(1 - SH_{mss}/\Delta T)$. (b) $\Delta T = 8.14$ K, $\rho = 10.2$. (c) About 13, and 54 s against 18 s.

12.7 (a) 5.98 K (±0.05); 0.74 to 7.82 K (±0.05). (b) 18.9 min (±0.1). (c) 3.99 K (±0.05). Below the line the coil's gain and lag rise under an unchanged loop.

12.9 (a) Floating 43.2 kWh and 27 starts, fixed 45.4 kWh and 58 starts (±0.1 kWh, exact counts). (b) About 4.9 %. (c) Case 2, the meat case at 1 °C, at about 78 % duty.

12.11 (a) 4.56, 7.27 and 10.91 kW; about 37 % of the compressor power. (b) P_{sat} is 2.00, 6.66 and 10.19 bar, so the share is $\sqrt{4.66/8.19} = 0.75$. Lower head pressure starves the valves.

Chapter 13. Milk Pasteurizer Holding and Flow Diversion

13.1 (c). The recipe is a temperature and a time, and $t_{tube} = V_T/Q$ falls below 15 s above 10 800 L/h.

13.3 (a). With it on, the hot-water setpoint bottoms at 71.5 °C instead of 70.3 °C, and one diversion is the whole cost.

13.5 (a) 10.12 kW/K, 5.67, 0.850, 1.38, 0.251. (b) $a_T = 0.994$, $g = 0.211$, 0.950 and 0.0486. (c) 0.39 °C.

13.7 (a) 1 diversion, 576 L. (b) 142 parcels, none diverted. (c) Parcels already in the tube at 330 s.

13.9 (a) 4 diversions, 1554 L, 74.2 °C s, safe but failed. (b) No diversion, 34.9 °C s, passed. (c) 7 diversions, 2084 L.

13.11 (a) 857 s, 62.5 %. (b) 77.0 °C and 81.2 °C. (c) About 29 % of clean, near 82 min, and the approach.

Chapter 14. Cloud Autoscaling With Integer Servers

14.1 (b). A busy fraction cannot read above one, so each decision can multiply the ready fleet by at most $1/u_{SP} = 1.43$, once per boot.

14.3 (d). The service has two states at the same traffic, and the clients' retries hold it in the overloaded one.

14.5 (a) $\lceil 25.71 \rceil = 26$. (b) At 90 %, $\phi' = 1.071$, within the tolerance, so no action. At 100 %, $\phi' = 1.190$ and $\lceil 1.190 \times 24 \rceil = 29$. (c) 52, 75, 108, 155 and the quota of 200, with the last order serving at about 20:16.

14.7 (a) Rule 70 %: 0 breach minutes, \$17.42. Rule 80 %: 18, \$15.41. (b) PI 80 %: 70, \$14.50, and the last breach minute closes at 08:37 in both. (c) \$2.01 saved for 18 breach minutes. Breach counts exact, bills $\pm$\$0.01.

14.9 (a) 129 storm minutes, 130 breach minutes, 52.2 %, \$36.39, last breach minute at 22:10. (b) 0, 13, 15.0 %, \$27.19, largest ratio 0.10, last breach at 20:13. (c) Without the budget the retries keep the offered load above what 60 servers serve until the evening traffic falls. Shares ± 0.1 point, bills $\pm$\$0.01, counts exact.

14.11 Bursty traffic ($V > 1$) and a heavier tail than the declared $c_{99} = 3$ (an exponential sojourn gives 4.6) both move the knee left. The model serves requests whose clients have timed out, which servers that receive the deadline would drop.

Chapter 15. Robot Paddle Juggling

15.1 (a). The paddle is slowing at impact, so the push it gives falls with the delay and pulls each bounce back onto the beat.

15.3 (c). Without the camera a 3 cm sinusoid at 2.5 Hz keeps the ball up indefinitely, at 0.223 m.

15.5 (a) -0.011, $\lambda = -0.006 \pm 0.800\,j$, modulus 0.80: an error shrinks by a fifth every bounce. (b) 1.977 and 0.324: unstable. (c) -9.931 m/s^2 and -10.070 m/s^2.

15.7 (a) Mean ratio 0.650, $e \approx 0.806$. (b) 2.98 s and 29 impacts, then 1.88 s and 19 impacts with ratios near $0.70^2 = 0.49$.

15.9 (a) 0.459 m. (b) 0.504 m, with an overshoot to 0.569 m on the climb. (c) 0.383 m. The proportional law needs an error to act, and the integral removes it.

15.11 Spin and a ball that leaves the plane, restitution that falls with impact speed and wear, and a ball that drifts across the face. Choose a concave face and a stroke well inside the window, slowing near the middle of it.

Chapter 16. Summary and Outlook

16.1 A good answer grants that the ramp wins on a correct model, by 0.44 mol over the composition loop, and notes that the comparison judges the drum's cumulative purity. It then cites the wrong charge of 0.42, on which the ramp collected 1.26 mol, a proportional trim 4.74 mol and a re-optimized ramp 10.69 mol. It requires a measurement or estimate of the charge before each batch, re-optimization when it differs, a trim or a composition loop as a backstop for small errors, and a check of the analyzer lag before relying on composition feedback, since an 8 min analyzer left 1.66 mol.

16.3 A good answer says the timer handles a nominal plant but gave 26.2 µm of carbide on a cold step, the estimate handled the cold step (35.0 µm) but left a gap of 28.35 µm at an efficiency of 0.72, and the load cell closed that gap to -0.79 µm but missed a 20 % charge error. It proposes an independent second measurement or a check of the estimate against an absolute quantity, a time limit behind the estimate, and a record of which disturbances the guard cannot see.

16.5 A good answer says a headline credits a whole policy with a gain that may come from one part of it, here the weaker half of the model, so the reader cannot judge how far it transfers. It proposes checks such as a comparison against a detuned or retuned rival given the same knob (the detuned filler law at 2324 against 2325, or the freeze-dryer rule at 15 Pa that tied the predictive controller) and a measurement of a stated law before it is printed (the molding variance law, tested over 195 parts).

Glossary

Each term is given as this volume uses it, in one plain sentence, followed by the chapter where it is first used in earnest. The two indexes give every page.

accumulation Reagent fed into a reactor that has not yet reacted, which could release its heat all at once if the cooling failed. *Chapter 2*.

admission control Refusing part of the incoming requests at the door, so that the servers can answer the rest in time. *Chapter 14*.

anti-windup A guard that stops the integral of a controller from growing while its actuator sits at a limit. *Chapter 1*.

autoclave A heated pressure vessel in which a composite part, sealed in a vacuum bag, is cured under heat and pressure. *Chapter 6*.

autoscaler Software that adds or removes servers of a service by a rule that watches a measured load. *Chapter 14*.

back-off A margin kept between a limit and the value the controller aims for, so that an error in the model does not break the limit. *Chapter 1*.

batch A quantity of product made in one run that starts, follows a recipe and ends, and is judged after the end. *Chapter 1*.

batch distillation Boiling a charged mixture in a column and drawing off its lighter part over time, so the composition keeps changing. *Chapter 3*.

breakthrough The moment product starts to leave a loaded column because the resin near its outlet is nearly full. *Chapter 11*.

bumpless transfer Handing an actuator from one controller to another so that the new one starts from the output already in use. *Chapter 1*.

capacitance gauge A pressure gauge that reads the true pressure of any gas from the bending of a thin diaphragm. *Chapter 5*.

cascade control An arrangement in which one controller sets the setpoint of another, faster one, as a reactor loop sets a jacket loop. *Chapter 2*.

checkweigher A scale on a filling line that weighs filled containers and reports each result some time after the fill. *Chapter 9*.

cluster autoscaler Software that adds machines when new copies of a service cannot be placed and removes machines that stay unneeded. *Chapter 14*.

coefficient of restitution The share of its speed a ball keeps when it bounces off a surface, from zero for none to one for all. *Chapter 15*.

collapse temperature The temperature above which a drying product loses its structure, the limit of a freeze-drying cycle. *Chapter 5*.

contactor A large relay that switches a heater or a motor fully on or fully off. *Chapter 1*.

cooldown A wait after one scaling action during which no further action of the same kind is allowed. *Chapter 14*.

cure The chemical hardening of a resin as its molecules join, driven by heat and timed by the recipe. *Chapter 6*.

cyclic steady state The repeating pattern that a repeated operation settles into, the same from one cycle to the next. *Chapter 1*.

deadband A range around the present value inside which a controller makes no change. *Chapter 1*.

distributed control system (DCS) A plant-wide control system with a supervisory level over local controllers that run the loops. *Chapter 1*.

dose In the kettle example, a measure of the heat treatment a product has received, gathered over the whole batch. *Chapter 1*.

dwell time A minimum time a switched device must stay in one state before it may switch again. *Chapter 1*.

endpoint constraint A requirement that holds once, at the end of a batch, such as a minimum dose or a purity. *Chapter 1*.

estimate A value computed from a model and measurements for something no instrument reads directly. *Chapter 1*.

EWMA controller A run-to-run controller that corrects the next setting from an exponentially weighted moving average of past errors. *Chapter 1*.

exotherm Heat released by a reaction, which can raise a temperature faster than cooling can remove it. *Chapter 6*.

feedforward Setting an actuator from a model of what the plan needs, before an error appears, so feedback corrects only what the model misses. *Chapter 1*.

fines Very small crystals, made by a burst of new crystals, that are hard to filter and lower the product quality. *Chapter 4*.

flapping A count of servers that keeps rising and

falling because the scaling rule reacts to every swing of the load. *Chapter 14*.

flow diversion device A valve on a pasteurizer that sends milk back for reheating whenever it may not have been held hot long enough. *Chapter 13*.

freeze-drying Removing water from a frozen product by letting the ice turn straight to vapor under a vacuum. *Chapter 5*.

funnel experiment A classroom demonstration that adjusting a process after every result, when only noise is present, makes the results more scattered. *Chapter 1*.

gate freeze The moment the narrow entry of a mold cavity solidifies, after which pressure on the screw no longer reaches the part. *Chapter 7*.

guard The condition on the way out of a recipe phase that says when the next phase begins. *Chapter 1*.

holding tube A length of pipe that keeps hot milk at its pasteurizing temperature for the legal holding time. *Chapter 13*.

HTST pasteurization High-temperature, short-time pasteurization: heating milk briefly in a continuous flow to kill harmful organisms. *Chapter 13*.

hybrid automaton A model made of modes, each with its own smooth motion, joined by conditions that switch between them. *Chapter 1*.

hysteresis Switching on at one value and off at another, so that a small swing cannot make a device switch again at once. *Chapter 1*.

impact map A rule that gives the state of a bouncing ball at one impact from its state at the impact before. *Chapter 15*.

injection molding Forcing molten plastic into a closed mold, holding it under pressure and cooling it into a solid part. *Chapter 7*.

integer actuator An actuator that comes in whole units, such as heating elements, compressors, servers or columns. *Chapter 1*.

interlock Logic that forces a safe action, such as cutting a feed, when a dangerous condition is detected. *Chapter 2*.

iterative learning control Control that improves the whole input profile of the next run from the error measured along the last run. *Chapter 1*.

latch Logic that keeps a protective action in force until its cause is cleared and someone resets it. *Chapter 1*.

lead In learning control, the shift in time or distance by which the last error is moved forward to cover the delay of the plant. *Chapter 1*.

lead/lag control Control that follows the slowest of several readings for the program and the hottest for the limit. *Chapter 6*.

load cell A weighing sensor of strain gauges bonded to a metal part, whose resistance changes with the force on it. *Chapter 1*.

melt pool The small spot of liquid metal that a laser makes as it scans across a bed of metal powder. *Chapter 10*.

metrology delay The number of runs that pass before the measured result of a run reaches the controller. *Chapter 1*.

minimum run and rest times Rules that keep a compressor on, or off, for a least time after each start or stop to limit wear. *Chapter 12*.

mirror law A juggling rule that moves the paddle as a mirror image of the falling ball, scaled to send it back up. *Chapter 15*.

model predictive control (MPC) Control that computes the best inputs over a future horizon at each step and applies only the first of them. *Chapter 1*.

monotone convergence Learning in which the error of every run is smaller than that of the run before, at every frequency. *Chapter 1*.

MTSR Maximum temperature of the synthesis reaction: the temperature a batch would reach if cooling failed and all unreacted reagent reacted. *Chapter 2*.

nucleation The birth of new crystals in a solution, as opposed to the growth of those already present. *Chapter 4*.

path constraint A limit that must hold at every instant of a batch, such as a maximum temperature. *Chapter 1*.

peristaltic pump A pump that doses liquid by rollers squeezing a flexible tube, so the liquid touches only the tube. *Chapter 9*.

phase One step of a recipe, with its own setpoints, its own controller and a condition that ends it. *Chapter 1*.

PI and PID control Feedback that acts on the present error, its accumulated sum and, for PID, its rate of change. *Chapter 1*.

Pirani gauge A pressure gauge that reads from heat loss through the gas, so its reading depends on which gas is present. *Chapter 5*.

primary drying The long phase of freeze-drying in which the ice is removed by turning it to vapor. *Chapter 5*.

process analytical technology Measuring quality during manufacture, rather than only at the end, so that the process can be controlled to it. *Chapter 4*.

programmable logic controller (PLC) A rugged industrial computer that runs logic, timing, counting and control loops for a machine or a unit.

Chapter 1.

Protein A A protein fixed to a chromatography resin that holds antibodies from a harvest while impurities are washed away. *Chapter 11*.

Q-filter A low-pass filter in a learning update that stops the update from learning errors it cannot correct. *Chapter 1*.

quantization Rounding a continuous request to the nearest available step, such as a whole number of elements. *Chapter 1*.

recipe The plan of a batch: its phases, their setpoints and the conditions that end each phase. *Chapter 1*.

reflux ratio The share of the condensed vapor at the top of a column that is returned down the column rather than drawn off. *Chapter 3*.

regenerator A heat exchanger in which hot pasteurized milk warms the cold raw milk coming in. *Chapter 13*.

retry storm Extra load created when clients that timed out send their requests again, which slows the service further. *Chapter 14*.

run-to-run control Control that adjusts the recipe settings of the next run from the measured result of earlier runs. *Chapter 1*.

secondary drying The last phase of freeze-drying, in which water bound to the dried product is driven off at a higher temperature. *Chapter 5*.

selector Logic that passes on the highest or the lowest of several signals, so the most demanding one is in charge. *Chapter 6*.

semi-batch reactor A reactor charged with one reagent and fed another over time, so that the feed rate limits the heat released. *Chapter 2*.

sequencer The logic that steps a machine or a batch through the phases of its recipe. *Chapter 7*.

shrinking-horizon control Predictive control that plans to the fixed end of a batch, so its horizon shortens as the batch proceeds. *Chapter 1*.

solenoid valve An on-off valve moved by an electromagnet coil against a return spring. *Chapter 1*.

split range One controller output divided between two final elements, such as a heating valve and a cooling valve. *Chapter 2*.

spouted bed A bed of particles kept moving by a jet of gas from below, as in a particle coater. *Chapter 8*.

staging Switching fixed-speed units such as compressors on or off in a set order to follow the load. *Chapter 12*.

superheat How far a vapor is heated above its boiling temperature, kept small but positive at the outlet of a cooling coil. *Chapter 12*.

supersaturation How far the concentration of a solution lies above what it can hold at rest, the driving force of crystal growth. *Chapter 4*.

switchover The moment an injection molding machine changes from filling the mold at speed to packing it under pressure. *Chapter 7*.

tare The weight of an empty container, subtracted from the filled weight to give the net fill. *Chapter 9*.

thermocouple A temperature sensor made of two different metals joined at a tip, which produce a small voltage that depends on temperature. *Chapter 1*.

thermostat An on-off controller that switches a heater or a cooler at two set temperatures. *Chapter 1*.

time-proportioning Giving an on-off actuator a fraction of each fixed period, so that its average output follows a continuous demand. *Chapter 1*.

trajectory A setpoint that changes along the batch, such as a heating ramp or a cooling curve. *Chapter 1*.

TRISO particle A tiny nuclear fuel kernel wrapped in four coating layers that keep the fission products inside. *Chapter 8*.

vacuum bag A sealed film over a composite layup, drawn to a vacuum so that the pressure holds the layers together. *Chapter 6*.

Zeno behavior Switching infinitely often in a finite time, which a model with no hysteresis band or dwell can predict. *Chapter 1*.

Index of Control Topics

adaptive switching, 268
admission control, 340
AI agents
  engineering practice, 383
  open problems, 387
  reproduction, 385
  truth audit, 385
  verification record, 384
  while studying, 17
analyzer
  bias, 70, 100, 103
  lag, 61, 266
Arrhenius rate, 199

back-off, 36
batch control, 1, 2, 141
  cooling trajectory, 83
  design procedure, 380
  endpoint, 111
  open problems, 49, 386
  population of units, 111
  reflux policy, 55
  semi-batch dosing, 23
  strategies, 11
  trajectory, 377
batch variability, 207, 215
bumpless transfer, 16, 94, 152, 170, 365, 385
  by tracking, 180

cascade control, 33, 317
  inner-loop speed, 34
  wall constraint, 94
conditional integration, 34, 317
constant-composition control, 57
coupled oscillators, 284
cure observer, 152, 165
cycle-to-cycle control, 172
cyclic scheduling, 269
cyclic steady state, 11, 269, 379

data export, 17
dead time, 330
  latency, 248
dynamic optimization, 55, 78

endpoint constraint, 5, 118
endpoint detection, 120, 378, 382
endpoint-aware policy, 120
event location, 359
event-based control, 365
EWMA controller, 6, 180, 230, 268

feasibility, 369
feedforward
  forecast floor, 339
  from an observer, 152
  model-based, 317
  trajectory, 5, 66
feedforward control
  learned profile, 248
Fenske equation, 60
finite-volume method, 263
flapping, 342

gain scheduling, 205
grid search, 66, 153
gross-error check, 211
guaranteed ramp, 151

heat exchanger
  counterflow effectiveness, 313
hunting, 294
hybrid automaton, 9, 269, 317
hybrid systems, 179, 283, 354, 382
  safe switching, 379, 382, 387
hysteresis, 9, 293

ideal-information bound, 120, 230, 268
impact dynamics, 359
inferential control, 194
integer actuator, 10, 293, 330
integral action
  once per event, 365
interlock, 318
iterative learning control, 3, 8, 105, 172, 244, 255, 381
  broken repetition, 252
  lead, 248
  learning gain, 251
  monotone convergence, 248
  with feedback, 252

Kalman filter, 16, 154

latch, 318
lead/lag control, 141, 151
limit cycle, 295

measurement
  unrepresentative, 126, 135
measurement lag, 336
method of moments, 88
metrology delay, 187, 214, 221
minimum dwell, 342
mirror law, 357, 365
model mismatch, 128, 230
model predictive control, 35, 339
  batch, 381
  in the apps, 16
  mixed-integer, 298
  safety, 385
  shelf and pressure, 121
  shrinking horizon, 72, 95, 153
  soft constraints, 36, 154
  successive linearization, 37
moving horizon estimation, 16

offset-free control, 36, 121, 154
open-loop stability, 362, 372
operator splitting, 148
outlier rejection, 229
override control, 141, 268, 340
  external reset, 152

parcel model, 312
path constraint, 5, 34, 83, 118
period doubling, 356
periodic orbit, 354
  period-1, 361
persistence, 317
phase sequencing, 179, 206
phase transition, 4
PID control
  anti-windup, 15, 94, 338
  velocity form, 15, 33, 64, 151, 179, 205, 248, 268
plausibility check, 320
Poincaré map, 362
population balance, 88, 198
process gain
  operating point, 291
  time-varying, 63
product history, 309
proportional control
  deadband, 338
purity credit, 64

Q-filter, 8, 248
quantization, 10
queueing
  latency knee, 335

recipe, 4, 47, 76, 104, 133, 151, 164, 196, 206
recipe optimization, 146, 153
recycle, 316
repetitive process control, 244
residence time, 313
return map, 11, 362
run-to-run control, 6, 49, 77, 170, 187, 221, 378, 383
  authority, 184
  integral form, 230
  noise amplification, 180, 229

safety interlock, 35, 49
saturating measurement, 336, 347
scenarios, 16, 180
  completion rules, 17
scorecard, 17
selector, 151
sensor bias, 184, 227, 268
sensor latency, 360
sensor redundancy, 320, 324
simulation engine, 14
simulator framework, 14
soft constraint, 122
soft sensor, 194, 216, 268
  correction by a slow measurement, 204
split-range control, 29, 146, 164
state estimation, 34, 203
  learned parameter, 120, 134, 268
superheat control, 295, 301
supersaturation control, 94, 105
supervisory control, 13, 309
supply-line accounting, 338
switching control, 9, 170, 268
synchronization, 283
  desynchronization, 294, 304
  order parameter, 294
  through a shared utility, 291

terminal objective, 56
trajectory tracking, 5, 83
truth log, 315

windup
  cascade, 317

Zeno behavior, 367

Index of Applications

ABS, 174
accumulation-limited dosing, 34
additive manufacturing, 243
aerodynamic drag, 358
affinity chromatography, 262
aseptic filling, 221, 224
aseptic intervention, 227, 237
autoclave, 142
    pressure, 148
autoclave cure, 141
autoscaling, 330

back-pressure valve, 314
balance tank, 310
ball
    rubber, 358
batch distillation, 55, 57
batch kettle, 3
boot delay, 336, 348
bouncing ball, 354
bracketed release, 227
breakthrough, 261, 263

camera, 360
capacitance gauge, 118, 134
capture train, 261
checkweigher, 12, 224, 236
chemical vapor deposition, 193, 215
cloud computing, 330, 332
coated particle fuel, 193
coefficient of restitution, 359
collapse temperature, 118
composite manufacturing, 141
compressor rack, 286
compressor staging, 293
condenser, 289
    vapor path, 117, 126
continuous bioprocessing, 261, 277
Coriolis flowmeter, 47
criticality classes, 30
crystal size distribution, 89, 105
crystallizer
    seeded batch cooling, 83
cut rule, 61, 77

dead volume, 272
defrost, 294
DiBenedetto equation, 147
display case, 286, 303
distillation column
    batch, 55
diurnal traffic profile, 334
dose records, 228, 238
downskin, 247
dried-layer resistance, 116
dynamic binding capacity, 261

encrustation, 89
exotherm, 142
expansion valve
    electronic, 285, 302
    thermostatic, 285

fill giveaway, 227
fill-weight control, 229
flash, 176
flash crowd, 334, 349
floating head pressure, 294, 303
floating suction pressure, 294
flow diversion valve, 310, 324
fouling, 265
    milk, 314, 325
freeze-drying, 111, 133
frost, 288
frozen skin, 175

gate freeze, 176
gelation, 148
glass transition, 118, 147
Gordon–Taylor equation, 118

hanging drop, 226
heat accumulation, 247
high-pressure cut-out, 289
holding time, 310, 325
holding tube, 310
holdup, 63
horizontal pod autoscaler, 333
HTST pasteurizer, 310

impact switch, 360, 371
injection molding, 170

juggling
  two balls, 369

Karl Fischer titration, 118
keyhole mode, 245
Kubernetes, 333

lack of fusion, 247
laminate
  thick, 147
Langmuir isotherm
  kinetic, 265
laser powder bed fusion, 243, 254
load balancer, 333
lyophilization, 111, 114

maximum temperature of the synthesis reaction (MTSR), 30
melt
  compressibility, 175
  power-law viscosity, 175
melt pool, 245
melt pool monitoring, 245
metastable failure, 333, 349
metastable zone, 85
metering pump, 47
milk pasteurizer, 309
minimum spouting velocity, 198
minimum stable superheat, 288
mold
  cavity, 174
mold cooling, 183
multi-column chromatography, 262, 277

nitrogen bleed, 117

overhang, 247

parcel record, 314
part length, 183
part scale, 176
part weight, 176
pasteurization, 309, 311
periodic counter-current chromatography, 261
peristaltic pump, 222, 237
photodiode, 247, 253
Pirani gauge, 118, 133
plastics processing, 170
porosity index, 148
prepreg, 142
pressure transducer, 176, 186
process analytical technology, 86
programmed cooling, 86
Protein A chromatography, 261
PVT law, 175

qualification, 245, 254, 255
quench, 30
queueing, 335

R134a, 289
raw-contact risk, 315
reaction calorimetry, 34
reagent accumulation, 28
reciprocating screw, 171, 186
recovery flush, 317
recycled resin, 171
reference weight, 234
reflux ratio, 57, 76
refractometer, 75
refrigeration pack, 283
regenerator, 310
relative volatility, 59
residence time, 271
resin aging, 265, 278
resin utilization, 270
restocking, 298
retry budget, 336
retry storm, 336
robot arm
  six-axis, 359, 371
robot juggling, 354, 372
rupture disc, 30

secondary drying, 118
seeding temperature, 99
semi-batch reactor, 23
server fleet, 333, 349
service-level objective, 333
short shot, 176
shrinkage, 176
silicon carbide, 195
sinusoidal stroke, 364
solubility curve, 86
spatter, 247
spot server, 336
spouted bed, 196
stoppering, 224
sublimation, 116
suck-back, 222
suction accumulator, 289
supermarket refrigeration, 283, 286
supersaturation, 85, 86

surge tank, 266
switchover
    trigger, 172

T2 Laboratories, 26
tail latency, 333
thermal runaway, 24, 26
thermocouple, 12, 148, 163
    product probe, 118
thermoset kinetics, 147
thermostatically controlled load, 284
three-way valve, 104
time to maximum rate, 28
timing pump, 314
titer assay, 266
total-reflux startup, 65
TRISO fuel, 193
tubing wear, 226

UV detector, 266, 276

vacuum bag, 142
velocity-to-pressure switchover, 171, 187
vial
    heat transfer coefficient, 117
    heterogeneity, 111
vial filling, 221
viscosity
    of the bulk, 226

walking robot, 357
warm pool, 336
web service, 330

About the Editor

John D. Hedengren is a professor in the Department of Chemical and Biological Engineering at Brigham Young University, where he leads the Process Research and Intelligent Systems Modeling (PRISM) group. He received the B.S. and M.S. degrees in chemical engineering from Brigham Young University and the Ph.D. degree from the University of Texas at Austin in 2005. He developed advanced process control applications for polymer production at ExxonMobil Chemical Company and commissioned nonlinear model predictive controllers with PAS, Inc. before joining the faculty at Brigham Young University in 2011.

His teaching covers process dynamics and control, engineering computing, data science, machine learning, and dynamic optimization, with course material published online for use inside and outside the classroom. He developed the Temperature Control Lab, a portable heat transfer device now used by many universities for process control and data science instruction, and he maintains the APMonitor and GEKKO optimization software used in those courses. He serves as Chair of Control Education for the IEEE Control Systems Society and as a trustee of the CACHE Corporation. He received the AIChE David Himmelblau Award in 2014, the AIChE Computing Practice Award in 2018, and the John R. Ragazzini Award in 2024, and he was inducted into the Process Automation Hall of Fame in 2025.

Volumes in This Series

Interactive Control Engineering: Principles and Applications has 20 volumes. Nine are organized by control technology and eleven by industry or application area. Every chapter in every volume has a simulator at https://apmonitor.com/apps/.

1 **Process Dynamics, Modeling, and Identification.** How plants respond, and how to find the model.
2 **Practical PID Control.** Tuning, anti-windup and implementation that works on real loops.
3 **Feedforward, Cascade, and Ratio Control.** Choosing a structure from where the disturbance enters.
4 **Override, Split-Range, and Constraint Control.** When limits decide what the controller may do.
5 **Delays, Slow Measurements, and State Estimation.** Controlling what arrives late or cannot be measured.
6 **Multivariable and Plantwide Control.** Pairing loops, managing inventories, closing the recycle.
7 **Nonlinear Dynamics, Oscillations, and Stabilization.** Plants that will not stay put without control.
8 **Batch, Run-to-Run, and Hybrid Control.** Recipes, endpoints and learning from the last run.
9 **Model Predictive Control and Economic Optimization.** Planning moves ahead, and choosing where to operate.
10 **Control in Upstream and Midstream Oil and Gas.** From the wellbore to the terminal.
11 **Control in Petroleum Refining.** Throughput, quality and limits from crude unit to blending.
12 **Control in Petrochemicals and Polymers.** Severity, selectivity and grade from cracker to pellet.
13 **Control in Chemical Plants.** Recycle, equilibrium and energy in continuous chemical processes.
14 **Control in Semiconductor Manufacturing.** Run-to-run control where little can be measured in the chamber.
15 **Control in Mining and Mineral Processing.** Variable ore, late assays and equipment at its limit.
16 **Control in Iron, Steel, and Metals Production.** Slow furnaces, fast strip, and measurements that arrive once.
17 **Control of Energy Conversion and Storage.** Fast regulation and slow dispatch on a changing grid.
18 **Control of Nuclear and Fusion Systems.** Reactor physics, protection limits and coordinated control.
19 **Control of Biological Systems and Human Performance.** Variable subjects, sparse samples and limits that cannot be crossed.
20 **Motion Control, Robotics, and Autonomous Systems.** Tracking, navigation, contact and allocation.