INTERACTIVE CONTROL ENGINEERING

Feedforward, Cascade, and Ratio 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 3

Feedforward, Cascade, and Ratio Control

Choosing a structure from where the disturbance enters

John D. Hedengren
Editor

Advanced Process Solutions, LLC

Interactive Control Engineering: Principles and Applications
Volume 3: Feedforward, Cascade, and Ratio 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/v03-feedforward-cascade/

How to cite this volume. John D. Hedengren, Ed., *Feedforward, Cascade, and Ratio Control*, vol. 3 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 choosing a control structure from where the disturbance enters. A single feedback loop reacts only after the upset has reached its measurement. Chapters 2 to 8 show the three standard answers on process plants: a fast inner loop nested in a slow one, a measured disturbance fed forward, and a ratio held between two flows, with cross-limiting where the order of two moves is a matter of safety. Chapters 9 to 12 carry the same structures to a power converter, a landing rocket, an orbital rendezvous and a climbing tether, where the inner loop removes a difficulty the outer loop could not handle.

Volume 2 covers the tuning of each loop used here. Volume 4 takes up what happens when a loop reaches a limit.

John D. Hedengren
2026

Contents

Preface v

Notation x

1 Introduction: Feedforward, Cascade, and Ratio Structures 1

1.1 The Question This Volume Answers . . . 1
1.2 A Short History . . . 2
1.3 Notation and Conventions . . . 3
1.4 Disturbance Paths and the Structures That Meet Them . . . 4
1.5 Instruments and Final Elements in This Volume . . . 11
1.6 The Simulator Framework . . . 12
1.7 Working With AI Agents While Studying . . . 16
1.8 Chapter Roadmap . . . 17
1.9 Exercises . . . 20

2 Matching Loop Structure to Disturbance Entry in a Tank Level Simulator 22

2.1 Introduction . . . 23
2.2 Background . . . 24
2.3 Process Model . . . 25
2.4 Dynamic Analysis . . . 29
2.5 Controller Design . . . 31
2.6 Results . . . 34
2.7 Technology in Practice . . . 38
2.8 Limitations . . . 40
2.9 Discussion . . . 40
2.10 Conclusion . . . 41
2.11 Exercises . . . 43

3 Measuring the Inner Loop Speed a Fermenter Oxygen Cascade Requires 46

3.1 Introduction . . . 47
3.2 Background . . . 48
3.3 Process Model . . . 50
3.4 Dynamic Analysis . . . 56
3.5 Controller Design . . . 57
3.6 Results . . . 59
3.7 Technology in Practice . . . 66
3.8 Limitations . . . 68
3.9 Discussion . . . 69
3.10 Conclusion . . . 69
3.11 Exercises . . . 71

4 Fired Heater Simulator for Cross-Limiting, Oxygen Trim and an Optimum Held by a Constraint 75

4.1 Introduction . . . 76
4.2 Combustion Control on Fired Equipment . . . 77
4.3 Process Model . . . 77
4.4 Implemented Control System . . . 81
4.5 Results . . . 82

4.6 Technology in Practice . . . 87
4.7 Limitations . . . 89
4.8 Conclusion . . . 89
4.9 Exercises . . . 91

5 Keeping an Organic Rankine Expander Dry When Its Waste Heat Falls 94
5.1 Introduction . . . 95
5.2 From Low-Grade Heat to a Superheat Loop . . . 96
5.3 A Closed R1233zd(E) Circuit Between Two Water Loops . . . 97
5.4 Loops, a Selector and a Protection Layer in Code . . . 103
5.5 What Structure Recovers and What the Plant Withholds . . . 107
5.6 Technology in Practice . . . 117
5.7 Limitations . . . 119
5.8 Conclusion . . . 119
5.9 Exercises . . . 121

6 Municipal Water Treatment Simulator for Feedforward Control of Coagulant Dosing 125
6.1 Introduction . . . 126
6.2 Background . . . 127
6.3 Process Model . . . 128
6.4 Controller Design . . . 131
6.5 Results . . . 132
6.6 Technology in Practice . . . 137
6.7 Limitations . . . 139
6.8 Discussion . . . 139
6.9 Conclusion . . . 140
6.10 Exercises . . . 142

7 Soft Drink Carbonation Simulator Whose Foam Limit Moves With Temperature 145
7.1 Introduction . . . 146
7.2 Fixed Air, Proportioners and Fillers . . . 147
7.3 The Line as Modeled . . . 148
7.4 The Margin Against Foaming . . . 154
7.5 Implemented Controller . . . 155
7.6 Scenarios and Measured Results . . . 158
7.7 Technology in Practice . . . 163
7.8 Limitations . . . 165
7.9 Conclusion . . . 165
7.10 Exercises . . . 167

8 Timing the Quench to Hold mRNA Lipid Nanoparticle Size at Higher Throughput 171
8.1 Introduction . . . 172
8.2 Ethanol, Jets and the Arrest of Fusion . . . 173
8.3 The Skid as Modeled . . . 174
8.4 Calibration and Its Domain . . . 180
8.5 Implemented Control System . . . 180
8.6 Nine Cases and a Challenge, Measured . . . 183
8.7 Technology in Practice . . . 188
8.8 Limitations . . . 190
8.9 Conclusion . . . 191
8.10 Exercises . . . 193

9 An Inner Current Loop Against the Right-Half-Plane Zero of a Boost Converter 196
9.1 Introduction . . . 197
9.2 Background . . . 198
9.3 Converter Model . . . 199
9.4 Origin and Load Dependence of the Right-Half-Plane Zero . . . 203
9.5 Controller Design . . . 204
9.6 Results . . . 209
9.7 Technology in Practice . . . 217
9.8 Limitations . . . 219
9.9 Discussion . . . 220
9.10 Conclusion . . . 220
9.11 Exercises . . . 222

10 Commanding an Attitude Instead of a Force in a Rocket Landing Simulator 225
10.1 Introduction . . . 226
10.2 Background . . . 227
10.3 Vehicle Model . . . 228
10.4 Linearized Dynamics and Loop Bandwidths . . . 234
10.5 Controller Design . . . 236
10.6 Scenarios . . . 239
10.7 Results . . . 239
10.8 Defects in the Application Measurements . . . 246
10.9 Technology in Practice . . . 246
10.10 Limitations . . . 248
10.11 Discussion . . . 249
10.12 Conclusion . . . 249
10.13 Exercises . . . 251

11 Canceling Coriolis Coupling by Feedforward in an Orbital Rendezvous Simulator 255
11.1 Introduction . . . 256
11.2 Background . . . 257
11.3 Relative Motion Model . . . 258
11.4 Controller Design . . . 264
11.5 Scenario Definitions . . . 268
11.6 Results . . . 268
11.7 Technology in Practice . . . 277
11.8 Limitations . . . 279
11.9 Discussion . . . 280
11.10 Conclusion . . . 280
11.11 Exercises . . . 282

12 Gravity Feedforward and Input Shaping for a Speculative Space Elevator Climber 286
12.1 Introduction . . . 287
12.2 Elevator Concepts, Ribbon Dynamics and Shaped Motion . . . 288
12.3 Climber, Ribbon and Loads . . . 290
12.4 Gravity That Fades and a Mode That Moves . . . 296
12.5 Five Layers, One Scan per Second . . . 297
12.6 Cases and What They Measure . . . 300
12.7 Technology in Practice . . . 306
12.8 Limitations . . . 308
12.9 Conclusion . . . 308
12.10 Exercises . . . 310

13 Summary and Outlook **313**
13.1 What the Chapters Showed . . . 313
13.2 Choosing a Structure . . . 316
13.3 Where Practice Is Heading . . . 317
13.4 Engineering With AI Agents . . . 318
13.5 Gaps in Feedforward, Cascade and Ratio Practice . . . 321
13.6 Exercises . . . 325

Answers to Selected Exercises **326**

Glossary **331**

Index of Control Topics **334**

Index of Applications **336**

About the Editor **338**

Volumes in This Series **339**

Notation

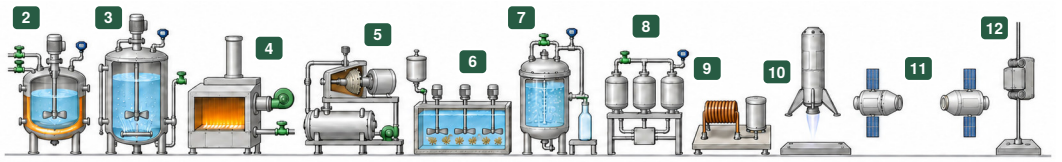
The chapters share the symbols below. A symbol used by one chapter only is defined where it first appears, and each chapter states its own units.

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: Feedforward, Cascade, and Ratio Structures



The chapters 2 tank, two upsets 3 fermenter oxygen 4 fired heater 5 organic Rankine cycle 6 water treatment 7 soft drink carbonation 8 nanoparticle skid 9 boost converter 10 rocket landing 11 orbital rendezvous 12 space elevator climber

Notice that the first eight units read their feedforward from meters through a balance, while the three vehicles take theirs from a known force. Illustration, not to scale.

IN PLAIN TERMS **Act Before the Upset Arrives**

The main idea. A simple control loop waits until a change has pushed its reading off target, and only then does it act. The methods in this book give it an earlier look. It can watch what is coming in, hold two flows in a fixed proportion, or let a fast inner loop fix trouble at the valve.

The challenge. Each method helps only against an upset that travels its own path. A correction that comes too early or is the wrong size becomes an upset of its own.

Why it matters. The same few methods run water plants, drink lines, power supplies and spacecraft, so choosing them well pays almost everywhere.

THREE WAYS TO READ THIS CHAPTER

Explore	8 min	This page, Section 1.2 with its timeline and two History boxes, the Did you know? boxes in Sections 1.1 and 1.4.3, 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 heater example worked, then Exercises 1 to 10.
Practice	5 min	This page, Section 1.4.7 with Table 1.3, Section 1.5 with Figures 1.6 and 1.9, and Section 1.8 with Figure 1.11.

1.1 The Question This Volume Answers

A single feedback loop learns about a disturbance only when the disturbance has moved the measurement. By then the plant has already paid for it. Tuning decides how quickly the loop answers after that moment. The structure of the loop decides whether the controller can answer earlier, and that choice is made before any tuning rule is applied. A loop is retuned many times in its life and restructured rarely, so a missing measurement is a lasting limit while a poor gain is a passing one [1]. This volume answers one question: where does each disturbance enter the plant, and which structure meets it there?

Three classical structures extend the single loop, and this volume measures each against the disturbance it was built for and against the ones it was not. Feedforward measures a disturbance and moves the actuator before the controlled variable responds. Ratio control holds one flow in proportion to another and lets a slow feedback loop correct the proportion. Cascade control closes a fast inner loop around the actuator, so that upsets entering there are removed before they reach the slow process. Cross-limiting adds a pair of selectors to a ratio, so that the error that is dangerous cannot occur while two flows of different speed move together.

DID YOU KNOW? Twice the Refinery for the Same Money

Cascade, selector and midranging structures appeared in the power and process industries, where much of PID control was also developed [2]. Between 1930 and 1940, automatic controls doubled the capacity that United States oil refineries could achieve for the same investment [3].

Eleven chapters test these structures on eleven plants, each with an app that runs in a browser. Chapter 2 opens with a tank that has one disturbance at its valve and one at its outlet, so that each structure can be seen to reject one and ignore the other. Chapters 3 to 8 are process plants: a fermenter, a fired heater, a waste-heat power cycle, a water treatment plant, a soft drink line and a nanoparticle formulation skid. Chapters 9 to 12 carry the same ideas to a power converter, a landing rocket, an orbital rendezvous and a space elevator climber, where the feedforward often comes from a known force and not from a meter.

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

- Trace each disturbance to the point where it enters, and say which measurement could see it first.
- Design a static feedforward from a balance and a lead–lag feedforward from two step tests, and predict what a wrong lag costs.
- Build a ratio with a feedback trim, and say why the trim multiplies the ratio and is not added to it.
- Decide whether a cascade pays, from the speed of its inner loop and the place where the disturbance enters.
- Arrange two flows of different speed so that the hazardous error cannot occur, and state what that arrangement costs.

This chapter develops each skill on one small example and names the chapter that uses it. PID tuning is not repeated here. Volume 2 develops it, and every loop in this chapter is a PI tuned from its model by the rules of that volume. Standard texts give fuller treatments of the structures [4, 5].

1.2 A Short History

The milestones of Figure 1.1 follow the three structures through the kinds of plant in this volume. Ratio control rests on proportions that physics sets. A paper of 1803 reported that water takes up twice or three times as much gas under two or three atmospheres as under the common pressure [6], and Chapter 7 doses carbon dioxide on that law in proportion to its product flow. A mechanical linkage that moves the fuel valve and the air damper from one actuator still sets the ratio of many simple burners, and adjusting it is a task for a skilled mechanic. A flow loop on one line and a flow transmitter on the other is the more automated approach [7].

First results on cascade control were published in the mid-1950s [17], among them the quantitative analysis of 1956 [8]. In a pneumatic boiler feedwater loop the output of the level controller is the setpoint of the flow controller, so a change of feedwater pressure is countered by the flow controller without the level controller having to react [7]. Computers arrived as supervisors over such loops.

HISTORY 1959: A Computer That Only Moved Setpoints

The first process control computers were supervisory. They adjusted the setpoints of pneumatic or electric controllers, so the plant kept working when the computer was off or busy with other tasks. The first computer built for process use went fully online in March 1959 on a refinery polymerization unit. Every hour it adjusted 19 parameters, and it raised the conversion of propylene to gasoline from 80 to 91 percent, paying for itself in less than three years. A process stream analyzer was the key instrument, because measuring product quality and closing the mass balance was the ultimate feedback if it could be applied quickly enough [3]. The slow trim of Section 1.4.4, which corrects a ratio from a late quality reading, is the same arrangement in a single loop.

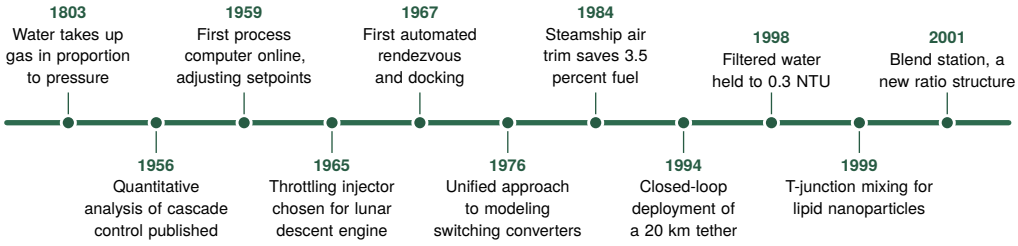


Figure 1.1 Milestones of feedforward, ratio and cascade control in the plants of this volume. The dates are from [6] (1803), [8] (1956), [3] (1959), [9] (1965), [10] (1967), [11] (1976), [12] (1984), [13] (1994), [14] (1998), [15] (1999) and [16] (2001). Notice that the vehicle and converter milestones sit among the process ones: the same structures served both. Events are evenly spaced, not to a time scale.

The vehicles of this volume take their feedforward from physics. A throttling concept for the lunar descent engine was selected in January 1965 [9], and two uncrewed spacecraft made the first automated rendezvous and docking in October 1967 [10]. A flight of 1994 deployed a tethered payload along the local vertical under a closed-loop law [13]. A conference paper of 1976 set out a unified approach to modeling the power stages of switching converters [11], and Chapter 9 uses its averaged form. In the process plants, a steamship fleet reported fuel savings of 3.5 percent from automatic combustion air trim in 1984 [12]. A rule of 1998 held the filtered water of large surface-water systems to 0.3 NTU in 95 percent of each month's readings [14]. T-junction mixing was introduced for lipid-based nanoparticles in 1999 [15], and a new ratio structure, the blend station, appeared in 2001 [16].

> ⧖ HISTORY **1963: Feedforward Beside Feedback**
>
> Feedback is reactive: an error must exist before it acts. When a disturbance can be measured before it enters, a controller can counteract it first, and this is feedforward. Because it tries to match two signals, feedforward needs a good process model, or its correction comes with the wrong size or at the wrong time. It runs open loop, acts on plans, is sensitive to model error and cannot destabilize the loop, while feedback is robust to model error and can. The value of systems with two degrees of freedom, which combine the two, was emphasized in 1963, and combining them is often the better choice [2]. The outer columns of Table 1.1 show a correction at the wrong time.

1.3 Notation and Conventions

The chapters share the symbols of the notation page. The setpoint is SP, the process variable is PV and the manipulated variable is MV. A tag such as TIC, FIC or FT names a temperature controller, a flow controller or a flow transmitter, and a subscript m marks a measured value. The controller settings are K_c and τ_I, and a process model has a gain K_p, a time constant τ_p and a dead time θ_p. A disturbance d reaches the PV through its own model, with gain K_d, time constant τ_d and dead time θ_d. In a cascade the outer or primary loop sets the setpoint of the inner or secondary loop, and τ_{in} is the closed-loop time constant of the inner loop. In a ratio, the wild flow is the one that is measured and not manipulated, and R is the ratio applied to it.

Each chapter keeps the time unit of its plant, from milliseconds in the converter to minutes and hours in the water plant. The running example of this chapter uses minutes, except Section 1.4.6, which uses seconds.

Four symbols need care across chapters. This chapter writes K_d for a disturbance gain, as Chapter 2 does. Chapter 3 uses K_{ff} for a confidence factor on its feedforward, where this chapter uses K_{ff} for

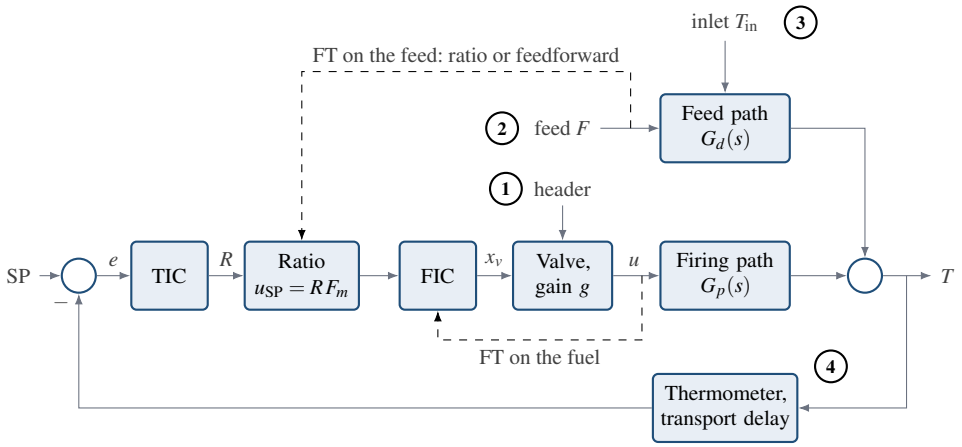


Figure 1.2 The Running Example With Every Structure of This Section. The TIC Trims the Ratio, the Ratio Station Sets the Firing Setpoint From the Measured Feed, and the FIC Moves the Fuel Valve. Dashed Lines Are Measurements Used Ahead of the PV. Circles Mark the Entry Points: (1) Header Pressure at the Valve, (2) Feed Flow, (3) Inlet Temperature, (4) the Measurement.

the feedforward gain itself. Chapter 4 writes λ for the burner air ratio and δ for the margin of its selectors, and Section 1.4.6 follows it. The shared framework uses a for the controller action and h for the integration step, so neither letter is used here for anything else.

1.4 Disturbance Paths and the Structures That Meet Them

1.4.1 The Running Example

The running example is a gas-fired heater that warms a stream of water. The MV is the firing demand u, in percent of the burner's maximum duty, and the PV is the outlet temperature T in °C. At design the feed is $F_0 = 10\,\mathrm{kg/s}$ at $T_{in} = 20\,^{\circ}\mathrm{C}$, and a firing of 40 % gives 60 °C. The steady energy balance, scaled to these numbers, is Equation (1.1).

$$T = T_{in} + u\,\frac{F_0}{F} \tag{1.1}$$

where F is the feed flow. At design one percent of firing is worth 1 °C, and one more kg/s of feed is worth −4 °C.

Two step tests give the dynamics, as shown in Equation (1.2).

$$G_p(s) = \frac{K_p\,\mathrm{e}^{-\theta_p s}}{\tau_p s + 1}, \qquad G_d(s) = \frac{K_d\,\mathrm{e}^{-\theta_d s}}{\tau_d s + 1} \tag{1.2}$$

with $K_p = 1\,^{\circ}\mathrm{C}/\%$, $\tau_p = 4\,\mathrm{min}$ for the firing path and $K_d = -4\,^{\circ}\mathrm{C}$ per kg/s, $\tau_d = 2\,\mathrm{min}$ for the feed path. Both paths share the dead time $\theta_p = \theta_d = 1\,\mathrm{min}$, the transport delay from the heater outlet to the thermometer. The firing path is slower because heat must pass the firebox and the tube wall, while a change of feed acts on the water at once. The dynamic runs below are linear in deviation variables about the design point. The temperature controller is a PI tuned by the SIMC rule with $\tau_c = \theta_p$, which gives $K_c = 4/(1 \times 2) = 2\,\%/^{\circ}\mathrm{C}$ and $\tau_I = 4\,\mathrm{min}$, and it executes every 0.1 min in the velocity form of Section 1.6.2.

Figure 1.2 shows the whole structure the section builds up. Each subsection adds one part of it. The numbered circles mark the four places where a disturbance can enter.

1.4.2 Where a Disturbance Enters

With a controller $C(s)$ in a single loop, a disturbance D that reaches the PV through G_d leaves the deviation of Equation (1.3).

$$Y(s) = \frac{G_d(s)}{1 + C(s)\,G_p(s)}\,D(s) \tag{1.3}$$

The numerator is the path from the entry point to the measurement. The denominator is the feedback, and it acts only on what has already arrived. Nothing in the controller can make G_d smaller. A structure helps only by measuring something that lies earlier on the disturbance's path than the PV does. Figure 1.2 gives four entry points, and each has a different remedy.

- *At the actuator* (1). A fall in header pressure lowers the fuel that a given valve lift passes. A flow meter downstream of the valve sees it within seconds, so a cascade with an inner flow loop removes it.
- *At the process, measured* (2). A change of feed flow enters the tubes. A flow meter on the feed sees it before the temperature does, so feedforward or a ratio can act on it.
- *At the process, unmeasured* (3). A change of inlet temperature has no transmitter here. Only feedback on the PV, or a slow trim, can correct it.
- *At the measurement* (4). A slow, biased or noisy thermometer corrupts the information every structure relies on. No rearrangement of controllers removes it. It needs a better instrument, a validation check or an estimator, which Volume 5 treats.

The same disturbance can change category with the instruments installed. A feed flow without a meter is entry point 3, and an inlet temperature with a transmitter becomes entry point 2.

Chapter 2 is built on this list. Its tank has a supply pressure at the valve and an outlet demand at the vessel, and the chapter shows that each structure rejects the disturbance on its own path and leaves the other nearly untouched. Chapter 3 sorts three lags of nearly equal size the same way: a cascade reaches the agitator drive, a feedforward reaches the feed, and nothing reaches the oxygen probe. The roadmap of Section 1.4.7 lists the paths of every chapter.

1.4.3 Static and Lead–Lag Feedforward

A feedforward adds $G_{ff}(s)$ times the measured disturbance to the MV. The PV is then $(G_d + G_p G_{ff})\,D$, which is zero for the ideal compensator of Equation (1.4).

$$G_{ff}(s) = -\frac{G_d(s)}{G_p(s)} = -\frac{K_d}{K_p}\,\frac{\tau_p s + 1}{\tau_d s + 1}\,\mathrm{e}^{-(\theta_d - \theta_p)s} \tag{1.4}$$

It has three parts. The gain $-K_d/K_p$ is the static feedforward, which holds the steady state. The ratio of lags is a lead–lag, which matches the timing. The delay term can be built only when the disturbance path is the slower one to reach the PV, $\theta_d \geq \theta_p$. Otherwise the ideal compensator would have to act before the disturbance is measured, and the design must settle for an approximation, for which tuning rules exist [18, 19].

For the heater the delays are equal and the compensator is $4\,(4s+1)/(2s+1)$ % per kg/s. The static part follows from Equation (1.1) without any test: holding T needs $u = (T - T_{\mathrm{in}})\,F/F_0$, whose slope at design is 4 % per kg/s. A static feedforward from a balance needs only the gains, which are known better than the dynamics.

Static feedforward alone leaves an error while the two paths respond at different speeds. For a step D and the exact static gain, the error integrates to Equation (1.5).

$$\int_0^\infty e\,\mathrm{d}t = K_d\,D\left[(\theta_p + \tau_p) - (\theta_d + \tau_d)\right] \tag{1.5}$$

The bracket is the difference between the mean response times of the two paths. When the firing path is the slower one, as here, the feedforward is late and the error keeps the sign of the uncompensated

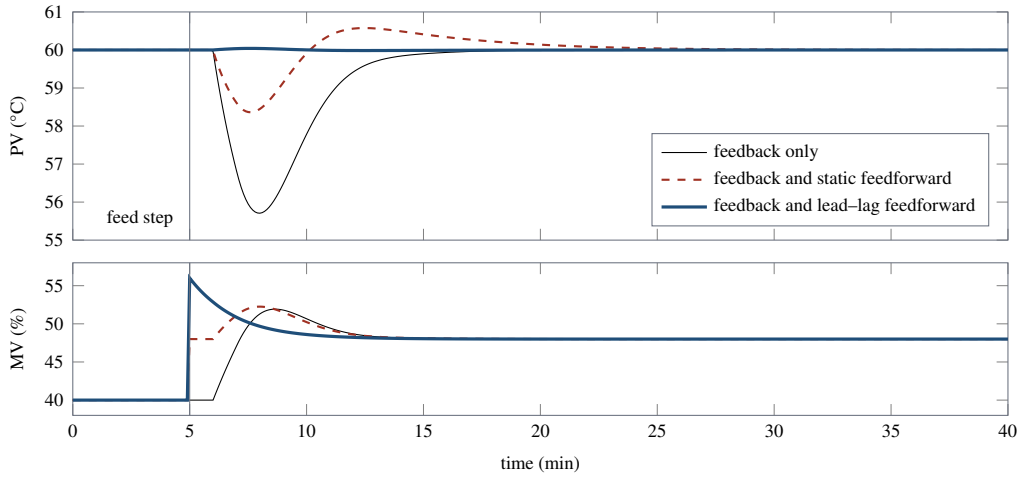


Figure 1.3 A Feed Step From 10 to 12 kg/s at 5 min on the Heater. Thin Solid: PI Feedback Only. Dashed: Feedback With the Static Feedforward of 4 % per kg/s. Thick Solid: Feedback With the Lead–Lag $4(4s+1)/(2s+1)$.

Table 1.1 Feedforward on the Feed Step of Figure 1.3. The Lead–Lag Is Designed for $\tau_d = 2$ min. The Two Outer Columns Run It on Plants Whose Feed Path Is Faster or Slower. Peak Error in °C, IAE in °C min Over 60 min.

	Plant $\tau_d = 1$ min		Plant $\tau_d = 2$ min		Plant $\tau_d = 3$ min	
Structure	Peak	IAE	Peak	IAE	Peak	IAE
Feedback only	5.89	16.8	4.29	16.0	3.39	16.0
Feedback and static feedforward	3.52	14.7	1.64	8.2	0.63	3.5
Feedback and lead–lag feedforward	1.92	7.0	0.04	0.2	1.06	5.0

disturbance. A step of 2 kg/s leaves −16 °C min, with a peak of 2.0 °C against 8 °C with no control. When the disturbance path is the slower one, the feedforward is early and the error takes the opposite sign. With $\tau_d = 8$ min the same static feedforward leaves 32 °C min of the opposite sign. An early feedforward is a disturbance of its own.

Figure 1.3 adds feedback to each feedforward, and Table 1.1 gives the measures. Feedback alone peaks at 4.29 °C. Its error integrates to 16 °C min, which is the firing change of 8 % times τ_I/K_c, the signature of any PI controller on a load. The static feedforward halves the IAE and cuts the peak to 1.64 °C. The lead–lag removes the error except for 0.04 °C left by the control period. It does so by moving the firing 16 % at once, twice its final change, because the lead of 4 min over a lag of 2 min doubles the initial response. That overshoot of the MV is what the slow firing path needs.

> **+ DID YOU KNOW? A Faster Upset Peaks Higher but Costs Nearly the Same**
>
> From the simulator: under feedback alone, the feed step of Table 1.1 peaks at 5.89 °C when its path lag is 1 min and at 3.39 °C when it is 3 min, yet the integrated error is 16.8 against 16.0. A PI pays for a load by its own settings, whatever the load's speed.

The outer columns of Table 1.1 show why a lead–lag is tuned with care. When the feed path is faster than the model, the lead–lag is still the best of the three. When it is slower, the lead–lag acts early and does worse than the static gain, which happens to match the slower path more closely. A lead–lag whose gain is 25 % low, three in place of four, leaves a quarter of the disturbance to

feedback, with a peak of 1.04 °C and an IAE of 4.0 °C min. Feedback is therefore kept beside every feedforward. It removes the residue of a wrong gain, and the feedforward removes the bulk of a measured disturbance that feedback would see late [20].

Every chapter of this volume uses a feedforward or a ratio, and most use a static one built from a balance or a known force. In Chapter 2 the outlet demand and the inlet flow enter the same integrator, so the ideal compensator is a pure gain. Chapter 4 inverts the heat balance of its heater, Chapter 9 computes a duty and a current from its power balance, and Chapter 10 and Chapter 12 hand the weight of the vehicle to the loop as a bias. Chapter 11 cancels the coupling terms of its equations of motion and measures how much a wrong orbital rate costs at two bandwidths. Timing decides two chapters. Chapter 3 must lag its feedforward to match the time the feed takes to reach the cells through a glucose pool, or the oxygen overshoots upward. Chapter 5 computes the lost heat from the source meters and lowers the pump's flow setpoint within seconds. These two chapters and Chapter 4 enter the feedforward as the output bias of the PID, so that switching it on or off moves the integral by the same amount and the output does not bump.

> ✦ DID YOU KNOW? **Feedforward in Markets, Firms and Muscles**
>
> A standard text compares feedforward with a planned economy and with running a company on strategic plans, and notes that it has been suggested as an essential, trained part of how humans control their motion [2]. The runner of Volume 19, Chapter 6 gets a measured feedforward of the incline, because the angle is known the instant it changes.

1.4.4 Ratio Control With a Feedback Trim

A ratio station sets the MV in proportion to a measured wild flow. For the heater the MV is the firing setpoint and the wild flow is the feed, as shown in Equation (1.6).

$$u_{\mathrm{SP}} = R\,F_m, \qquad R = R_0 + \Delta R \tag{1.6}$$

where R_0 is the design ratio and ΔR is the output of a slow feedback controller on the PV. Equation (1.1) gives $R_0 = (T - T_{\mathrm{in}})/F_0 = 4\,\%$ per kg/s. Linearized at the design point, the ratio is the static feedforward of Section 1.4.3. Ratio control is the oldest form of feedforward, long used on blending and combustion services [16, 21]. For the heater the balance makes the firing proportional to the feed at a fixed temperature rise. A feed of 15 kg/s needs 60 % of firing, and $R_0 F$ delivers it. The ratio and the static feedforward therefore agree at every feed, and they differ only in how a correction combines with them.

The trim exists to correct what the ratio does not know. Suppose the inlet water cools from 20 to 10 °C, with no transmitter to report it. The heater now needs $R = 5\,\%$ per kg/s. At the design feed the TIC finds the missing 10 % of firing either way. A trim that multiplies, as in Equation (1.6), settles at $\Delta R = 1$. A trim that adds a bias b to $R_0 F_m$ settles at $b = 10\,\%$. The two agree until the feed moves. At 15 kg/s the multiplied ratio gives the 75 % the heater needs, while the added bias gives 70 %, and the outlet runs 3.3 °C cold until the slow trim finds the difference. At 5 kg/s the added bias overfires by 5 % and the outlet runs 10 °C hot. An error in the ratio scales with the wild flow, so the correction should scale with it too.

The trim is slow on purpose. A trim as fast as the main loop would also react to every transient the ratio leaves while the wild flow changes, so it is tuned to follow composition, heating value or a fouled exchanger over many residence times. A ratio can also be applied to the setpoint of the wild flow in place of its measurement. The setpoint basis moves both flows at once, and the measurement basis follows the flow that actually arrived when a valve cannot keep up.

Five chapters build on this subsection. Chapter 6 computes the coagulant dose from the raw-water turbidity and lets the settled turbidity trim it through a multiplier. It argues the case made above, that a correction found at one raw turbidity remains right at another. Chapter 7 blends syrup into water at a ratio trimmed on a refractometer, doses carbon dioxide in proportion to the product flow, and

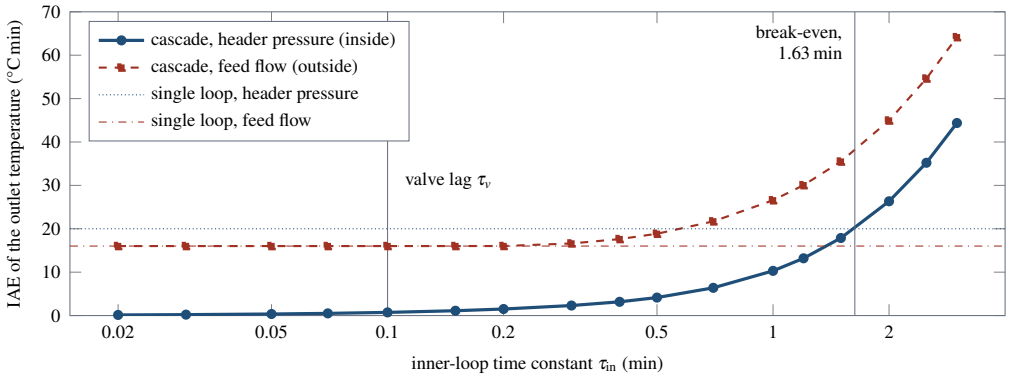


Figure 1.4 What the Speed of the Inner Loop Buys. The Heater's TIC Is Held at $K_c = 2$, $\tau_I = 4$ min. Thick Solid With Circles: Cascade Against a 36 % Fall in Header Pressure. Dashed With Squares: Cascade Against a Feed Step of 2 kg/s. Dotted and Dash-Dotted Lines: the Single Loop Against the Same Two Upsets.

compares the measured-water and setpoint bases when a water valve falls behind. Chapter 8 solves four feed flows that hold a total flow, two ratios and a buffer strength at once. Chapter 4 sets the air in ratio to the fuel and trims the ratio on stack oxygen, and Chapter 2 expresses its feedforward gain as a ratio of flow setpoint to demand when the cascade is selected.

1.4.5 Cascade Control and the Inner-Loop Speed Rule

The fuel valve of the heater passes a flow $u = g\,x_v$, where x_v is the lift and g grows with the square root of the pressure drop. When the header pressure falls by 36 %, g falls from 1 to 0.8 and a lift of 40 % passes 8 % less fuel. That upset enters at point (1) of Figure 1.2, upstream of the whole firing path. With the TIC on the valve, the temperature falls through G_p before the controller can respond. A cascade places a flow controller between them. The TIC asks for a fuel flow, and the FIC moves the valve until the meter reads it. The header pressure is then removed inside the inner loop, before it reaches the firebox.

The inner loop has dynamics of its own. Here the valve and the fuel line respond as a lag of $\tau_v = 0.1$ min. A PI on that lag with $\tau_I = \tau_v$, or a proportional controller alone, closes the inner loop with the time constant of Equation (1.7).

$$\tau_{in} = \frac{\tau_v}{K_{c,in}\,g} \quad \text{(PI)}, \qquad \tau_{in} = \frac{\tau_v}{1 + K_{c,in}\,g} \quad \text{(P only)} \tag{1.7}$$

The outer loop sees this closed inner loop in place of the valve. If it is faster than the valve was, the outer loop gains. If it is slower, the outer loop has a slower actuator than before.

Figure 1.4 measures the trade on the heater. The TIC keeps its tuning throughout, and only the inner gain changes. With a single loop the header pressure costs a peak of 3.09 °C and an IAE of 20.0 °C min. With an inner loop of 0.05 min the peak is 0.11 °C and the IAE 0.36 °C min, 56 times smaller. The feed step, which enters outside the inner loop, costs the same 16.0 °C min with or without the cascade, because no flow meter on the fuel sees it.

As the inner loop slows, the two curves separate the two parts of the speed rule. The rule, stated since the first quantitative analysis of cascade control, asks for an inner loop substantially faster than the outer one [8]. Against the header pressure the cascade keeps 93 % of its attainable benefit at $\tau_{in} = 0.2$ min, 80 % at 0.5 min and none at 1.63 min. Against the feed step it costs nothing until the inner loop is as slow as the valve lag it encloses, about 0.1 min. It then costs 18 % at 0.5 min and nearly three times the single-loop IAE at 2 min. The usual statement of the rule is three to five

Table 1.2 The Three Arrangements of Figure 1.5. The Fuel Time Is How Long the Fuel Takes to Cover 63 % of the Step.

	Increase, 40 to 60 %		Decrease, 60 to 40 %		Time short
Arrangement	Lowest λ	Fuel time (s)	Lowest λ	Fuel time (s)	of air (s)
Fuel leads	0.880	2.0	1.151	2.0	10.8
Air leads	1.150	12.2	1.095	12.2	0
Cross-limited, $\delta = 0$	1.150	12.2	1.152	2.0	0
Cross-limited, $\delta = 2\,\%$	1.133	10.9	1.151	2.0	0

times faster [4, 22]. The outer loop here has a closed-loop time constant of $\tau_c + \theta_p = 2\,\text{min}$, so that statement gives 0.4 to 0.67 min. It still collects 71 to 85 % of the benefit, and it already costs 10 to 32 % on the feed step outside the inner loop. A factor of ten costs nothing measurable here. A sharper statement for this heater is that the inner loop should be no slower than the actuator lag it replaces, since every disturbance that the inner loop cannot see then pays for its delay.

A cascade has two further uses that the curves do not show. The inner loop linearizes the actuator. The TIC asks for a flow, and the square root of the valve law, a sticking stem or a worn pump no longer reaches the outer loop. The inner loop is also where a limit can be enforced, since a flow or a current that is controlled can be capped. Tuning both loops from their models is treated in [23]. When the secondary measurement responds to the MV but does not lie on the path to the primary, the parallel arrangement applies [24].

Seven chapters measure this subsection. Chapter 2 finds a stability limit for an inner loop with no plant dynamics, set by one sample of delay, which the supply pressure carries in and out of range on every cycle at some gains. Chapter 3 sweeps the inner speed of an agitator loop, finds that a detuned cascade is three times worse than no cascade, and shows that the inner integral time moves the break-even point by a factor of 1.8. Chapter 5 wraps a flow loop around its feed pump. Chapter 9 closes an inner current loop, which enforces a current limit and lets the voltage loop run harder against a right-half-plane zero. Chapter 10 commands an attitude from a position loop and sweeps the separation of their bandwidths, and Chapter 4 and Chapter 7 put a flow loop under every ratio.

1.4.6 Cross-Limiting of Fuel and Air

A burner needs air in proportion to fuel, and the two errors are not equal. Excess air carries heat up the stack. Too little air leaves unburned fuel in a hot enclosure, which is a hazard. The heater burns gas at a design air ratio $\lambda_0 = 1.15$, where $\lambda = F_a/F_f$ is the air supplied over the air the fuel needs. The fuel flow loop closes in 2 s and the forced-draft fan in 10 s. Times in this subsection are in seconds.

Three arrangements are compared on a firing step from 40 to 60 %, the change the ratio station of Section 1.4.4 asks for when the feed rises from 10 to 15 kg/s. In *fuel leads* the fuel follows the demand and the air follows the measured fuel. In *air leads* the air follows the demand and the fuel follows the measured air. Cross-limiting combines the two with a low selector on the fuel and a high selector on the air, as shown in Equation (1.8).

$$F_f^{\text{SP}} = \min\Big[u,\ \frac{(1+\delta)\,F_{a,m}}{\lambda_0}\Big], \qquad F_a^{\text{SP}} = \max\big[\lambda_0\,u,\ (1-\delta)\,\lambda_0 F_{f,m}\big] \tag{1.8}$$

where δ is a small margin by which the fuel may run ahead of the air. The fuel can rise no faster than the measured air can burn it, and the air can fall no faster than the measured fuel allows. On an increase the air leads, and on a decrease the fuel leads [25].

Figure 1.5 and Table 1.2 give the result. With fuel leading, the fuel reaches its new flow in seconds and the fan follows five times more slowly. The burner runs short of air for 10.8 s, down to $\lambda = 0.880$. Air leading never runs short, but the fuel now waits for the fan in both directions, and on the decrease

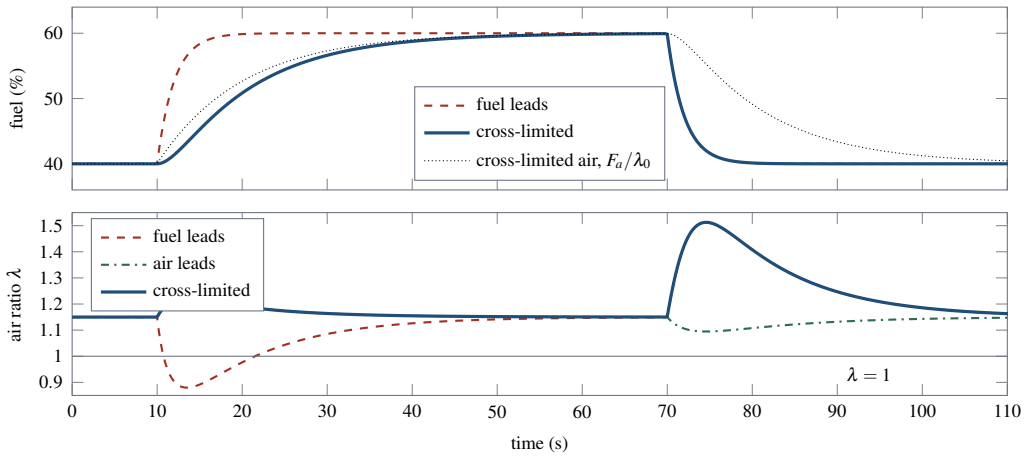


Figure 1.5 Firing Demand Stepped From 40 to 60 % at 10 s and Back at 70 s. Top: Fuel Flow, and the Cross-Limited Air in Units of the Fuel It Can Burn at λ_0. Bottom: Air Ratio. Dashed: Fuel Leads. Dash-Dotted: Air Leads. Thick Solid: Cross-Limited. Below the Thin Line at $\lambda = 1$ the Burner Is Short of Air.

the fuel stays high for 12 s after the demand has fallen. Cross-limiting takes the safe half of each. It is as fast as fuel leads on the way down and as safe as air leads on the way up. Its cost is the heat delay on an increase, the sum of the two loop lags, which the TIC sees as a slower actuator. A margin of $\delta = 2\,\%$ recovers 1.3 s of that delay and lets λ dip to 1.133. Every scheme runs with extra air for some seconds in at least one direction, which is the price of safety and is paid in fuel.

Chapter 4 builds this arrangement on a refinery heater with a fan six times slower than its fuel valve. It then trims the ratio on stack oxygen and adds a carbon monoxide override, a selector that Volume 4 develops. The asymmetry argument applies wherever one direction of a ratio error is a hazard, as with reactant feeds to an exothermic reactor.

1.4.7 Which Structure for Which Path

The section reduces to one rule. A structure helps against a disturbance only if it measures the disturbance, or a variable on its path, earlier than the PV does. Feedforward and ratio act on a measured disturbance at the process. A cascade acts on any upset between the controller output and the inner measurement. A trim corrects what was not measured, slowly. Nothing here corrects the measurement. A structure applied to the wrong path is not merely useless. A feedforward from a variable that the MV itself moves can oscillate, as in Chapter 2, and a slow inner loop delays every disturbance outside it.

Table 1.3 applies the rule to every chapter of the volume. It lists the main disturbances each chapter applies, where each enters, and the structure the chapter tests against it. The final chapter returns to this table with the measured result of each pairing.

Three patterns run through the table. The process plants of Chapters 3 to 8 combine a measured disturbance, met by feedforward or a ratio, with an unmeasured one, met by a slow trim. The vehicles of Chapters 10 to 12 take their feedforward from physics, so the model and not a meter is what can be wrong. Several chapters keep one disturbance that no structure reaches, such as the probe lag of Chapter 3 or the condenser capacity of Chapter 5, and those chapters state the limit flatly. A broader view of choosing controlled and secondary variables across a whole plant is given in [26], and Volume 6 develops it.

Table 1.3 Disturbance Paths in This Volume and the Structure Each Chapter Tests Against Them

Ch.	Disturbance	Where it enters	Structure tested
2	Outlet demand	Vessel, measured	Outlet feedforward (inlet-flow feedforward fails)
	Supply pressure	Valve	Level-to-flow cascade
3	Feed step	Culture, measured	Carbon-balance feedforward with a lag
	Gassing on the shaft	Actuator	Speed cascade
	Probe lag	Measurement	None
4	Process flow	Process, measured	Heat-balance feedforward
	Load steps	Fuel and air pair	Cross-limiting
	Heating value	Burner, unmeasured	Oxygen trim of the ratio
5	Source heat	Evaporator, measured	Measured-heat feedforward
	Pump wear, density	Actuator	Inner flow loop
	Warm cooling water	Condenser capacity	None
6	Raw-water storm	Intake, measured	Computed dose
	Organic carbon	Intake, laboratory only	Multiplying ratio trim
7	Filler rate	Throughput, measured	Throughput feedforward and ratio blend
	Syrup strength	Blend, unmeasured	Brix trim of the ratio
	Water header dip	Water valve	Ratio on the measured flow
8	Throughput change	Four feeds	Four-feed ratio block
	Size drift	Late laboratory result	Slow trim with a soft sensor
9	Load and supply voltage	Power stage, measured	Feedforward on both loops
	Overload	Inductor current	Inner current loop and its limit
10	Weight under tilt	Known force	Gravity feedforward
	Crosswind	Side force	Position-to-attitude cascade, no integral
11	Gravity gradient, Coriolis terms	Known coupling between axes	Model feedforward that cancels the coupling
12	Effective gravity along the route	Known force that changes sign	Gravity feedforward
	Ribbon mode	The climber's own moves	Input shaping

1.5 Instruments and Final Elements in This Volume

Every structure of this volume adds a measurement or a loop, so it can be no better than the instrument it adds. Figure 1.6 lists what the controllers of each chapter read and what they move. The flow meter and the control valve or flow loop lead, with six chapters each, because a ratio holds only what a meter reports and a flow loop is the usual inner loop. Analyzers and quality probes follow with five chapters, nearly always as the input of a slow trim.

A ratio station, or signal multiplying relay, multiplies the signal of the wild flow transmitter by a constant and sends the product to the captive flow controller as its setpoint [7]. The ratio is then only as good as the two meters. A Coriolis meter senses the inertial force of the fluid in its vibrating tubes, so it is a true mass flow instrument with no straight-pipe requirement. External vibration and gas bubbles trapped in bent tubes cause errors [7]. Figure 1.7 shows its principle, and the technology section of Chapter 2 places such a meter under its inner loop.

In a cascade the secondary loop must respond faster than the primary, and the scheme cannot work if that order is reversed [7]. The lowest loop is often in the valve itself. It is customary to close inner loops on valve positioners to reduce friction and nonlinearity at the lowest level of the hierarchy [2], so the fuel loop of Figure 1.2 on such a valve is three loops deep (Figure 1.8). A speed drive with a tachometer is likewise the secondary controller of a cascade, taking its setpoint from a process controller [7], as the agitator of Chapter 3 does.

A supervisory computer sends setpoints to regulatory controllers that run in automatic with a remote setpoint, which forms a two-layer control system [7]. Loops in such systems are often nested or cascaded, with the setpoint of one loop based on the process variable of another, and cycle times

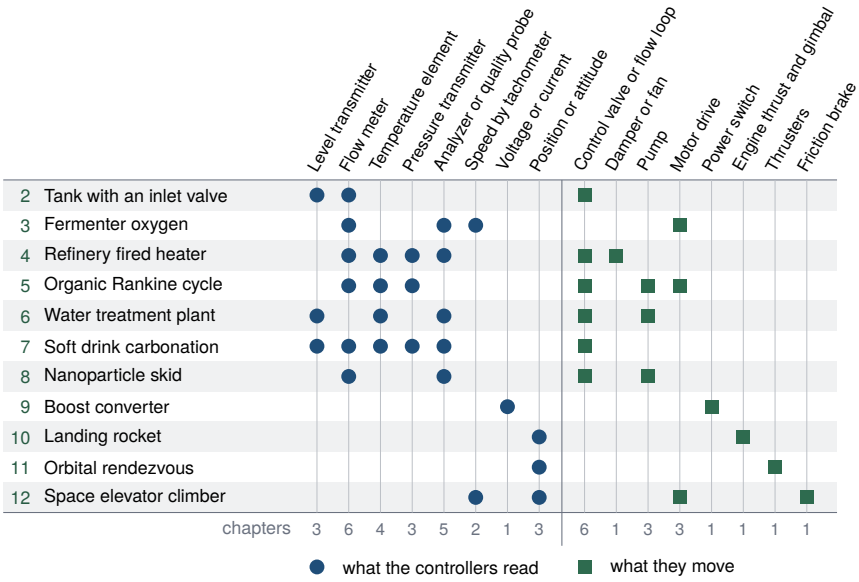


Figure 1.6 What the controllers of each chapter read and move: a round mark where a controller reads the instrument and a square mark where it moves the element, as the chapter's variables table and its technology section name them. A value computed from other readings, such as a superheat, is marked under the instruments it is computed from. Notice that the flow meter and the valve or flow loop lead the counts, with six chapters each.

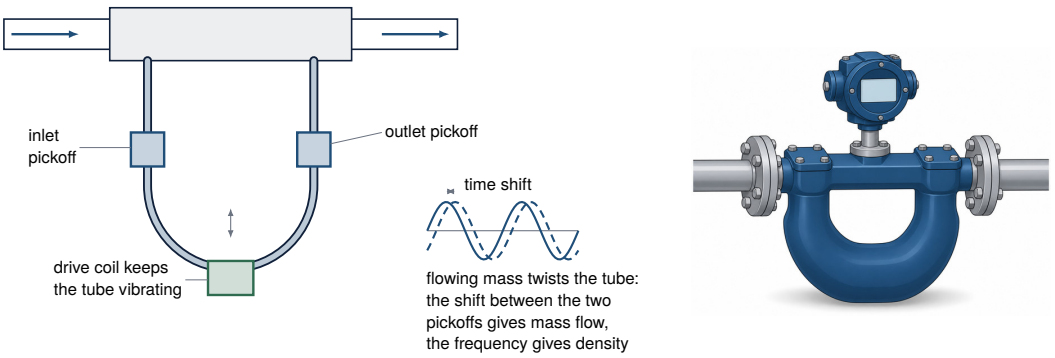


Figure 1.7 Principle of a Coriolis mass flow meter. Notice that it reads mass flow directly, which is what a ratio of two liquid streams needs, and that the reading depends on a steady vibration of the tubes.

run from milliseconds to minutes [27]. A converter splits its cascade across hardware. A gate array runs the fast controller, and an outer loop on a processor core supplies its reference at a lower rate [28]. Figure 1.9 places the loops of this volume on their platforms. The apps run from the 0.2 ms period of Chapter 9 to the 0.2 min period of Chapter 6, and the supervisor of Chapter 8 acts every 30 s.

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

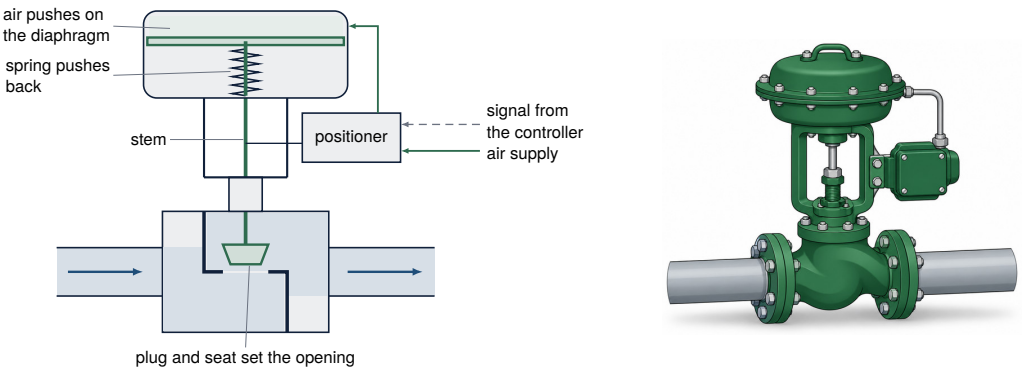


Figure 1.8 A globe control valve with its positioner. Notice that the positioner compares the controller's signal with the stem position and moves the stem itself, so it is the innermost loop of every valve cascade in this volume.

Layer	What it does in this volume	Where it runs	How often
Supervisor and slow trims	a trim on a late analyzer, a supervisor that picks a mixing loop, a predictive controller over a cascade	**Supervisory computer** that writes setpoints to the loops below	every 30 s in the formulation skid; trims over many residence times
Outer loop with feedforward, ratio	the primary PID with the feedforward as its bias, a ratio station, the cross-limiting selectors	**Distributed control system** or PLC function blocks; a **flight computer** on a vehicle	0.2 s to 0.2 min in the plants; 10 ms to 1 s in the vehicles
Inner loop	a flow loop under a ratio, a drive's own speed loop, a converter's current loop	**Flow controller**, transmitter or positioner; a **speed drive**; a **gate array**	0.25 s for a flow loop; 0.2 ms in the converter
Sensors and final elements	meters, analyzers and probes; valves, pumps, drives, a power switch, an engine, thrusters, a brake	wires or a fieldbus; a **positioner** closes the lowest loop on the valve stem	device lags from 0.2 ms to 20 s in the models

Figure 1.9 The control layers of this volume, the platform that hosts each and how often each acts. Platforms are from [7], [27], [2] and [28], and the app periods from the chapters' technology sections. Notice that each layer writes the setpoint of the one below it, so a cascade runs through the whole stack.

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 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.

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.

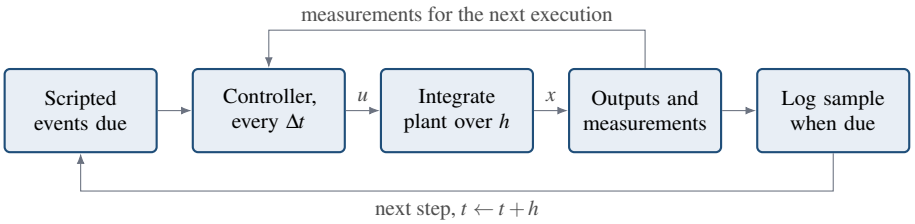


Figure 1.10 Work of the Engine in One Integration Step

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 (1.9) to (1.12) for execution k and control period Δt.

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k, \qquad D_k = -a\,K_c\,\tau_D\,\frac{\mathrm{PV}_k^f - \mathrm{PV}_{k-1}^f}{\Delta t} \tag{1.9}$$

$$S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \tag{1.10}$$

$$v_k = u_b + S_k + I_{k-1}, \qquad u_k = \mathrm{sat}(v_k) \tag{1.11}$$

$$I_k = I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{T_t}\right]\Delta t \tag{1.12}$$

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 (1.12) drives I toward the value that reproduces the output applied, which is back-calculation [29]. With no integral action, Equation (1.12) 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 $I_k = u_{\mathrm{man}} - u_b - S_k$, 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 (1.13).

$$x_{k+1} = A\,x_k + B\,u_k + E\,w_k + c, \qquad y_k = C\,x_k \tag{1.13}$$

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 (1.14) over a horizon of N steps.

$$J = \sum_{k=1}^{N} (y_k - r_k)^{\top} Q\,(y_k - r_k) + \sum_{j} \Delta u_j^{\top} R\,\Delta u_j + \sum_{i} \rho_i\, s_i^2 \tag{1.14}$$

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 [30]. 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 [31]. 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 [32].

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.

	Feedforward from a meter and a balance	Feedforward from a known force or model
Each structure on its own path	(2) feedforward from the wrong flow oscillates	
Feedforward gain and timing	(5) measured heat protects the expander	(11) coupling canceled exactly (12) a weight that fades and reverses
Ratio with a slow trim	(6) dose ahead of a slow basin (8) four ratios held (7) blend trimmed by an analyzer	
Inner-loop speed and limits	(3) a detuned cascade does worse (9) current loop and its limit	(10) position commands an attitude
Cross-limiting and override	(4) fuel and air of different speeds	

Figure 1.11 Method map of the volume. Each chapter is placed by the source of its feedforward, from Table 1.3, and by the structural lesson it centers on, shortened from Table 1.5. The placement is the editor's reading of the two tables. Notice that the three vehicles take their feedforward from physics, while the plants and the converter read it from meters.

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.
- *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 13.

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 where their feedforward comes from and by the structural lesson each centers on.

Table 1.5 Roadmap of the Volume

Ch.	Plant	The structural lesson	Level	Time
2	Tank on level with an inlet valve	Each structure rejects the disturbance on its own path. A feedforward from the wrong flow oscillates.	introductory	2.5 h
3	Fed-batch fermenter, dissolved oxygen	The inner-loop speed rule measured, and a feedforward that must wait for the culture.	intermediate	3 h
4	Refinery fired heater	Cross-limiting of fuel and air, oxygen trim of the ratio and a heat-balance feedforward.	intermediate	2 h
5	Waste-heat organic Rankine cycle	A flow cascade that needs a feedforward to protect the expander, and a limit no loop removes.	advanced	2.5 h
6	Municipal water treatment plant	A computed dose ahead of three hours of process, with a slow multiplying trim.	introductory	2 h
7	Soft drink carbonation line	Ratio blending with an analyzer trim, and a quality that is a pressure margin.	intermediate	2 h
8	Nanoparticle formulation skid	Four ratios held at once, and a slow trim on a result a minute old.	advanced	2.5 h
9	DC–DC boost converter	An inner current loop against a right-half-plane zero, and a current limit.	advanced	3 h
10	Rocket on a gimballed engine	A cascade whose outer loop commands an angle. Bandwidth separation and gravity feedforward.	advanced	3 h
11	Chaser approaching a target in orbit	A model feedforward that cancels coupling, and what a wrong orbital rate costs.	intermediate	3 h
12	Space elevator climber	Feedforward of a weight that fades and reverses, and shaping against a ribbon mode.	advanced	2.5 h

Chapter 2 is the first chapter to read, because it isolates each structure on one plant with two disturbances. A first course can continue with Chapters 6, 4 and 3, which add a dose computed ahead of a long process, cross-limiting and the inner-loop speed rule in that order. Chapters 5 to 8 combine several structures on one plant and can be read in any order once Sections 1.4.3 to 1.4.5 are familiar. Chapters 9 to 12 need the same sections and a little rigid-body or circuit dynamics. They form a second path for readers from electrical or aerospace engineering, best started with Chapter 11, whose feedforward is exact.

KEY POINTS

- A single loop learns of a disturbance only when the measurement has moved. A structure helps only if it measures the disturbance, or something on its path, earlier than the PV does.
- A static feedforward from a balance needs only the gains. A lead–lag also needs the timing, and an early feedforward is a disturbance of its own.
- A ratio trim should multiply the ratio, because an error in the ratio scales with the wild flow.
- A cascade pays when its inner loop is no slower than the actuator lag it replaces, and it cannot touch a disturbance outside that loop.

Where else this applies. Selectors and overrides in Volume 4, dead time in Volume 5, plant-wide structure in Volume 6.

References

[1] F. G. Shinskey, "Process control: As taught vs as practiced," *Industrial & Engineering Chemistry Research*, vol. 41, no. 16, pp. 3745–3750, 2002.

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

[3] M. Pitt, "Taking a look back at control: Part 2," *The Chemical Engineer*, no. 989, Nov. 2023, thechemicalengineer.com/features/taking-a-look-back-at-control-part-2.

[4] 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.

[5] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[6] W. Henry, "Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures," *Philosophical Transactions of the Royal Society of London*, vol. 93, pp. 29–274, 1803.

[7] 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.

[8] R. G. Franks and C. W. Worley, "Quantitative analysis of cascade control," *Industrial & Engineering Chemistry*, vol. 48, no. 6, pp. 1074–1079, 1956.

[9] W. R. Hammock, Jr., E. C. Currie, and A. E. Fisher, "Apollo experience report: Descent propulsion system," NASA Manned Spacecraft Center, Houston, TX, NASA Technical Note TN D-7143, 1973, nASA Technical Reports Server 19730011150.

[10] J. L. Goodman, "History of space shuttle rendezvous," NASA Johnson Space Center, Mission Operations Directorate, Flight Dynamics Division, Houston, TX, Report JSC-63400, Revision 3, Oct. 2011, nASA Technical Reports Server 20110023479.

[11] R. D. Middlebrook and S. Cuk, "A general unified approach to modelling switching-converter power stages," in *1976 IEEE Power Electronics Specialists Conference*, 1976, pp. 18–34.

[12] J. J. Sweeney, "Boiler Flue Gas Analysis and Automatic Combustion Air Trim Control by Measurement of Carbon Monoxide, Unburned Hydrocarbons, and Opacity," *Marine Technology and SNAME News*, vol. 21, no. 01, pp. 74–79, 1984.

[13] "Tethers in space handbook," Smithsonian Astrophysical Observatory for NASA Marshall Space Flight Center, NASA Contractor Report NASA/CR-97-206807, 1997, nTRS 19980018321.

[14] U.S. Environmental Protection Agency, "National primary drinking water regulations: Interim enhanced surface water treatment; final rule," *Federal Register*, vol. 63, no. 241, pp. 69 478–69 521, 1998, december 16, 1998.

[15] J. Hong Hoang, M. Ott, E. Samaridou, M. Beck-Broichsitter, and J. Simon, "Industrial perspective on the manufacturing of lipid nanoparticles for nucleic acid delivery," *Pharmaceutics*, vol. 18, no. 4, p. 489, 2026.

[16] T. Hägglund, "The blend station—a new ratio control structure," *Control Engineering Practice*, vol. 9, no. 11, pp. 1215–1220, 2001.

[17] D. Gromov, "Analysis of hierarchical structures for hybrid control systems," Ph.D. dissertation, Technische Universität Berlin, Berlin, 2010.

[18] J. L. Guzmán and T. Hägglund, "Simple tuning rules for feedforward compensators," *Journal of Process Control*, vol. 21, no. 1, pp. 92–102, 2011.

[19] M. Hast and T. Hägglund, "Low-order feedforward controllers: Optimal performance and practical considerations," *Journal of Process Control*, vol. 24, no. 9, pp. 1462–1471, 2014.

[20] E. J. Adam and J. L. Marchetti, "Designing and tuning robust feedforward controllers," *Computers & Chemical Engineering*, vol. 28, no. 9, pp. 1899–1911, 2004.

[21] F. G. Shinskey and B. G. Lipták, "Ratio control," in *Process Control*. Elsevier, 1995, pp. 102–105.

[22] P. R. Krishnaswamy, G. P. Rangaiah, R. K. Jha, and P. B. Deshpande, "When to use cascade control," *Industrial & Engineering Chemistry Research*, vol. 29, no. 10, pp. 2163–2166, 1990.

[23] Y. Lee, S. Park, and M. Lee, "PID controller tuning to obtain desired closed loop responses for cascade control systems," *Industrial & Engineering Chemistry Research*, vol. 37, no. 5, pp. 1859–1865, 1998.

[24] W. L. Luyben, "Parallel cascade control," *Industrial & Engineering Chemistry Fundamentals*, vol. 12, no. 4, pp. 463–467, 1973.

[25] D. Lindsley, *Power Plant Control and Instrumentation: The Control of Boilers and HRSG Systems*. Institution of Engineering and Technology, 2000.

[26] S. Skogestad, "Control structure design for complete chemical plants," *Computers & Chemical Engineering*, vol. 28, no. 1–2, pp. 219–234, 2004.

[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] P. Karamanakos, E. Liegmann, T. Geyer, and R. Kennel, "Model predictive control of power electronic systems: Methods, results, and challenges," *IEEE Open Journal of Industry Applications*, vol. 1, pp. 95–114, 2020.

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

[30] 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.

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

[32] 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 Concept check · 2 min

A steam-heated exchanger holds its outlet temperature with a steam valve. The steam header pressure swings by 20 % several times an hour, and a flow meter is installed on the steam line downstream of the valve. Which structure removes the header swing before it reaches the outlet temperature?

(a) Static feedforward from the measured process flow to the steam valve

(b) A cascade in which the temperature controller sets a steam flow controller

(c) A ratio of steam flow to process flow with no feedback trim

(d) A faster integral time on the temperature controller alone

1.2 Theory and calculation · 5 min

A process has the manipulated path $G_p(s) = 2\,\mathrm{e}^{-3s}/(6s+1)$ and a measured disturbance path $G_d(s) = 0.5\,\mathrm{e}^{-5s}/(3s+1)$, with time in minutes. Find the ideal feedforward compensator, say whether it can be built, and give the ratio of its initial response to its final response for a step disturbance.

1.3 Theory and calculation · 4 min

A static feedforward with the exact gain is applied without feedback. The manipulated path has $\theta_p + \tau_p = 5\,\mathrm{min}$, the disturbance path has $K_d = 1.5$ PV units per unit of D and $\theta_d + \tau_d = 3\,\mathrm{min}$. A disturbance step of $D = 2$ occurs. Find the integral of the error, and say whether the feedforward is early or late.

1.4 Concept check · 2 min

In a fermenter a feed step raises the oxygen demand of the culture only after a glucose pool has filled, about a minute later. A feedforward computes the agitator speed the new feed will need and applies it at once. What does it do to the dissolved oxygen?

(a) It holds the dissolved oxygen exactly, because the computed speed is right

(b) It lets the dissolved oxygen dip, as it would with no feedforward at all

(c) It drives the dissolved oxygen above the setpoint before the demand arrives

(d) It makes no difference, because the feedback trim removes the term

1.5 Theory and calculation · 6 min

(a) A heater fires $u = c\,(T_{\mathrm{SP}} - T_{\mathrm{in}})\,F$ with $c = 0.2\,\%$ per °C per kg/s, $T_{\mathrm{SP}} = 80\,^{\circ}\mathrm{C}$ and $T_{\mathrm{in}} = 20\,^{\circ}\mathrm{C}$. A ratio station sets $u = R\,F_m$. Find the design ratio R and the ratio needed after the inlet warms to 30 °C. *Checkpoint:* $R = 12$, then $R = 10$ % per kg/s

(b) With the inlet at 30 °C, a trim that adds a bias to $12\,F_m$ settles at the design feed of 5 kg/s. The feed then rises to 8 kg/s before the trim moves. Find the firing error and the outlet temperature error. *Checkpoint:* 6 % overfired, outlet 3.75 °C hot

1.6 Theory and calculation · 4 min

An agitator drive has a lag of 12 s. A proportional speed controller gives an inner loop gain $K_{c,\text{in}}K_p = 5$. Find the closed-loop time constant of the inner loop, and its ratio to an outer process whose fastest element has a time constant of 8 s.

1.7 Concept check · 2 min

A tank level is controlled by a cascade in which the level controller sets an inlet flow controller. The outlet is drawn by a metered pump whose demand steps by 50 %. What does the cascade do for this disturbance compared with a single level loop on the valve?

(a) Almost nothing, because the demand enters the tank and not the inlet line

(b) It removes most of the error, because the flow loop is fast

(c) It removes all of the error once the integral of the inner loop settles

(d) It doubles the error, because the inner loop always adds a lag

1.8 Theory and calculation · 5 min

A cross-limited burner uses the selectors of Equation (1.8) with $\lambda_0 = 1.2$ and $\delta = 0$.

(a) The demand has just risen to $u = 60$ while the measured fuel is 45 and the measured air is 60, in the units of the equation. Find both setpoints. *Checkpoint:* fuel 50, air 72

(b) Later the demand falls to $u = 40$ while the measured fuel is 55 and the measured air is 70. Find both setpoints. *Checkpoint:* fuel 40, air 66

1.9 Concept check · 2 min

A level loop adds a static feedforward from the inlet flow transmitter, which is downstream of its own control valve, with the gain that would be right for a disturbance at the outlet. What does this feedforward do?

(a) It rejects the outlet demand as a feedforward from the outlet would

(b) It rejects the supply pressure exactly, as a cascade does

(c) It has no effect, because the inlet flow is not a disturbance

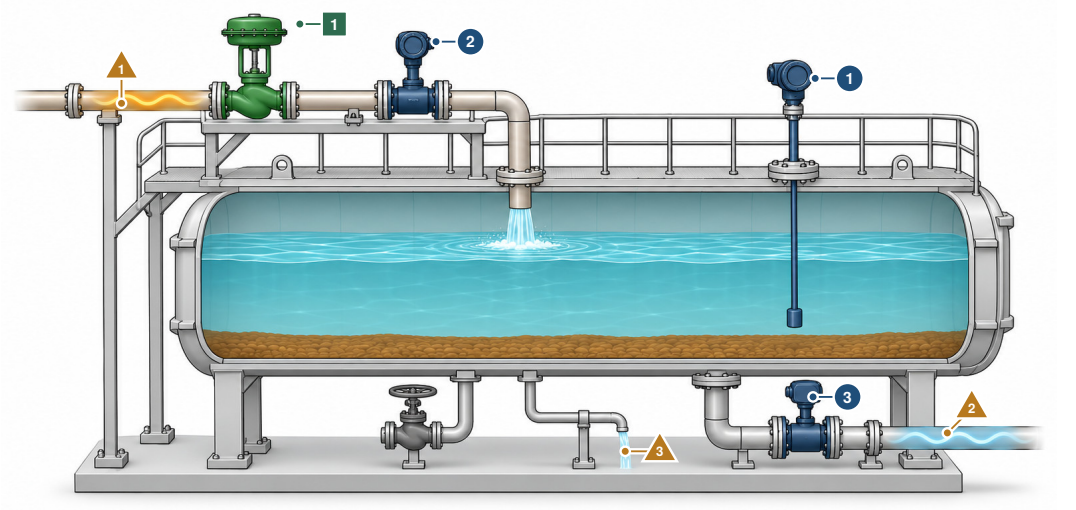
(d) It feeds the valve's own effect back to the valve, which can oscillate

1.10 Theory and calculation · 4 min

The heater of Section 1.4.3 uses the lead–lag $4\,(4s+1)/(2s+1)$ % per kg/s, with time in minutes. The feed steps by 2 kg/s. Find the change of firing it asks for at once, after 2 min, and at steady state.

CHAPTER 2

Matching Loop Structure to Disturbance Entry in a Tank Level Simulator



The controller moves 1 the inlet valve on the filling pipe

The controller reads 1 the level in the tank 2 the flow arriving through the valve 3 the flow drawn out at the outlet

What disturbs it 1 the supply pressure ahead of the valve, swinging up and down

2 the outlet demand, stepping up and back 3 a small leak that grows with the level

Notice that the supply upset enters at the valve and the demand upset at the outlet, two different points on one tank, and that the model's tank is a vertical one of uniform area, as in Figure 2.1. Illustration, not to scale.

IN PLAIN TERMS **Fix the Upset Where It Comes In**

The main idea. A controller, a small computer, holds a tank's level by opening and closing the valve that fills it. Two upsets reach the tank: the push of the water at the valve rises and falls, and the draw at the bottom jumps.

The challenge. A second controller that watches the inflow cuts the first upset within seconds but misses the second. Watching the outflow fixes the second, while watching the wrong flow makes the valve jump back and forth.

Why it matters. What the controllers read decides what they can fix, and adjusting them cannot make up for a wrong choice.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	12 min	the first page, the fact on the steam drum (Section 2.1), the history of cascade control (Section 2.2), the try-it segments of scenarios 6 and 4 (Sections 2.4 and 2.5), the fact on the inner gain (Section 2.6) 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 2.6), Technology in Practice (Section 2.7), Limitations (Section 2.8) and the open challenge

The control question A tank held on level by an inlet valve meets two upsets that enter at different points, a supply pressure swing at the valve and a step in the outlet demand. Which control structure rejects which upset, and what does a structure chosen for the wrong one cost?

Learning objectives

- Derive the valve gain, the linearized level model and the ideal outlet feedforward gain from the mass balance, and explain why that compensator is a pure gain.
- Compute the roots and the Jury limit of the inner flow loop, and state how the limit moves with the supply pressure.
- Explain why feedforward from the inlet flow transmitter oscillates and halves the authority of the level controller.
- Measure each structure against each upset separately, and read the level error together with the valve travel.
- Judge a claim about a structure by where the upset enters, not by one aggregate error.

Prerequisites Sections 1.4.2, 1.4.3 and 1.4.5 (entry points, static feedforward, inner-loop speed), 1.6.2 (velocity form) and 1.6.1 (clocks and noise).

The app https://apmonitor.com/apps/cascade/

See also Volume 1, Chapter 2 for a gravity-drained tank whose gain and time constant move with the level, and Volume 1, Chapter 3 for step tests on two tanks whose order changes with a splitter valve. Volume 2, Chapter 2 for the velocity-form PID with back-calculation, and Volume 2, Chapter 5 for the integrating level that this tank becomes with the leak closed.

2.1 Introduction

Tuning a controller changes how quickly a loop rejects a disturbance. Restructuring the loop changes whether it can reject that disturbance at all. The distinction is routine in operating plants, where the fastest route to a steadier column or a steadier furnace is often a secondary measurement and a second controller rather than a better gain, and it is the distinction that control structure design addresses before any tuning rule is applied [1, 2]. A single loop measures its controlled variable, so every disturbance reaches it through the process dynamics that separate the disturbance entry point from the measurement. When those dynamics are slow, the loop responds late. Cascade and feedforward are the two classical remedies, and they are not interchangeable. Cascade inserts a faster measurement between the actuator and the slow process, so upsets that enter upstream of that measurement are caught inside the inner loop [3, 4]. Feedforward measures the disturbance itself and moves the actuator before the controlled variable has responded [5, 6].

Textbook treatments of both structures present a worked transfer function and a simulation in which the improvement is demonstrated [7, 8]. The negative half of the result is harder to convey, that a structure chosen for the wrong disturbance gives almost no benefit and can perform worse than the single loop it replaced. A student who has applied cascade to a supply-pressure disturbance may apply cascade again to a load disturbance, where the inner loop is outside the disturbance path and contributes nothing. An experiment that applies both disturbances together hides this point, because the aggregate error improves for reasons the experiment cannot separate.

Tank level control is well suited to the demonstration. The vessel supplies a capacity large enough that the controlled variable is slow, the inlet valve supplies a manipulated path with a square-root pressure dependence that makes a supply disturbance multiplicative rather than additive, and the outlet draw supplies a load disturbance that enters the same mass balance as the manipulated flow. The model rests on a liquid mass balance, the conservation law that produces first-order level dynamics on a vessel with a level-dependent outflow [7, 8], and the separation between the fast valve and the slow inventory is the time-scale separation that cascade control requires [9]. Two flow transmitters on the two sides of the vessel then make all four structures physically available on one apparatus, so the plant is unchanged when the structure changes. The experiments on this plant cover process order,

disturbance entry location, cascade structure, static feedforward, integral anti-windup and sampled measurement. They apply the sorting of disturbances by entry point that Section 1.4.2 sets out, and this tank is the example that section names for it.

> ✦ DID YOU KNOW? **The Same Two Structures in a Steam Drum**
>
> A steam drum is held by three-element control, which reads the feedwater flow, the steam flow and the level. A study of practice at solar power plants calls it the common industry-standard scheme for drum level [10]. Volume 18, Chapter 4 builds it on a drum as a cascade plus a feedforward, the pairing this chapter measures on a tank.

This chapter places four control structures on one tank in a single browser app and applies the two disturbances separately, so that the pairing between a structure and the disturbance it rejects is measured rather than asserted. The structures themselves are standard. The level loop, both feedforward paths and the level-to-flow cascade share one plant, one actuator and one sample period, so a comparison between them is free of every confounding difference except the structure. The separated disturbance protocol, and the reporting that goes with it, is the method of the chapter. Section 2.2 traces the two structures through the literature. Section 2.3 gives the model in enough detail to rebuild the simulator from this chapter alone. Section 2.4 linearizes it, derives the ideal feedforward compensator and compares the two loops at the frequency of the pressure swing. Section 2.5 states the control algorithm, the four structures, the stability limit of the inner loop and the algebraic feedback that the inlet feedforward closes. Section 2.6 reports a step test, the combined and separated scenarios, an inner-loop gain sweep and the scenarios of the app. Section 2.7 places the loop on real devices and platforms. Sections 2.8 to 2.10 give the limitations, a discussion and the conclusion, and the exercises follow in Section 2.11.

2.2 Background

How cascade control came from pneumatic practice to its first quantitative treatment [3] is told in the history box below, and the design rule still in use is that the secondary loop must be substantially faster than the primary one for the arrangement to give a benefit. Later work sharpened the rule into an explicit comparison against single-loop control across ratios of primary to secondary dynamics, and identified the region in which cascade gives no worthwhile benefit at all [4]. Chapter 3 measures that region on a fermenter, where a detuned inner loop performs worse than no cascade. Variants followed for cases where the secondary measurement responds to the manipulated variable but not to the primary disturbance, which requires the parallel rather than the series arrangement [11], and analytic tuning rules were derived for both controllers of a series cascade from the two identified models [12].

> ⧖ HISTORY **First Results on Cascade Control, 1956**
>
> In the pneumatic form of a boiler feedwater system, the level controller's air signal of 3 to 15 psi is piped to a flow controller as its setpoint, and only the flow controller moves the valve. A swing in feedwater pressure is then countered by the flow controller before the drum level has to react [13]. Much of the development of PID control took place in the power and process industries, where structures such as cascade, selector and midrange control also appeared [9]. The first published results on cascade control date from the mid-1950s, in a paper of 1956 by Franks and Worley. A later view counts cascade as a precursor of hierarchical control, because the outer loop is designed as if the closed inner loop were a fairly simple system [14].

Feedforward control developed along a separate route. Ratio control on blending and combustion services is the oldest form, where a measured wild stream sets the demand for a manipulated stream through a fixed factor [6]. The general form replaces the fixed factor by the compensator of Section 1.4.3, which in the usual case is a lead-lag with a dead time and cannot be realized exactly.

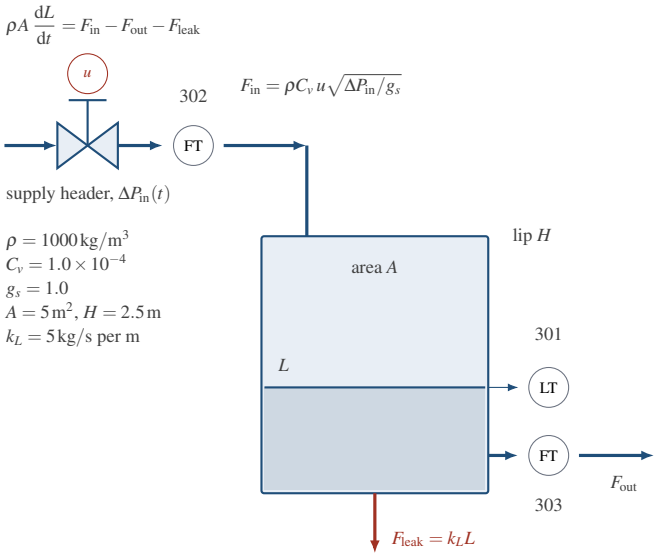


Figure 2.1 Instrumented tank with an inlet control valve. The valve lift u is the manipulated variable in all four structures, the outlet demand F_{out} and the supply pressure drop ΔP_{in} are the two upsets, and LT-301, FT-302 and FT-303 are the three transmitters the controllers are allowed to read.

The design literature is largely about what to do when that inverse is improper or unstable, how to tune the lead and the lag when the two paths carry different delays, and how much robustness the compensator gives up when the plant moves [5, 15, 16]. Advanced PID practice treats both structures as the standard extensions of the single loop [17], and process control texts introduce them together for the same reason, because the two answer complementary questions [7].

The industrial evidence is that structural choices persist and tuning choices do not. A loop is retuned many times over its life and rewired rarely, so a missing secondary measurement is a permanent limitation while a poor gain is a temporary one, as Section 1.1 also argues. Surveys of industrial practice report that the proportional-integral-derivative algorithm remains the dominant regulatory controller and that cascade and feedforward are its most common structural extensions [18]. Practitioner assessments of what plants actually run report the gap between the structures taught and the structures installed, and place the regulatory layer, not the advanced layer, at the center of delivered benefit [2]. Plantwide design methodology reaches the same conclusion from the other direction, treating the selection of controlled and secondary variables as the decision that bounds everything downstream [1].

2.3 Process Model

The simulated plant is a vertical tank of uniform cross-section fed through a control valve from a supply header and drawn down by a metered outlet flow, with a small level-proportional leak that makes the vessel self-regulating. The arrangement and the symbols used throughout are shown in Figure 2.1.

2.3.1 Variable Classification

The classification of every signal the application exposes is given in Table 2.1. One differential state carries the whole plant. The inlet flow is not a state, because the valve law is algebraic, but it becomes a secondary controlled variable when the cascade is selected, and that change of role is the only effect the structure selection has on the process side of the simulation.

Table 2.1 Variable Classification

Symbol	Role	Unit	Range
L	state, primary CV	m	0 to H
L^{SP}	primary setpoint	m	0–2.5
u	MV, valve lift	%	0–100
F_{in}	secondary CV, cascade	kg/s	0–46.9
$F_{\mathrm{in}}^{\mathrm{SP}}$	secondary setpoint	kg/s	0–40
F_{out}	DV, load	kg/s	0–20
$\overline{\Delta P}_{\mathrm{in}}$	DV, mean drop	bar	2–22
a	DV, swing amplitude	bar	0–10
k_L	DV, leak coefficient	kg/s/m	0–10
F_{leak}	internal flow	kg/s	$k_L L$

2.3.2 Mass Balance, Valve Law and Leak

The single state equation is the liquid mass balance on the tank, written for constant density and uniform cross-section, as shown in Equation (2.1).

$$\rho A \frac{\mathrm{d}L}{\mathrm{d}t} = F_{\mathrm{in}} - F_{\mathrm{out}} - F_{\mathrm{leak}} - F_{\mathrm{spill}}, \qquad 0 \le L \le H \tag{2.1}$$

where ρ is the liquid density, A is the tank cross-sectional area, L is the level, F_{in} is the inlet mass flow through the control valve, F_{out} is the metered outlet demand and F_{leak} is the leak. Each Euler substep floors the level at zero and caps it at the lip height H. Any excess inventory leaves as spill: if an unconstrained step would give $L^{*} > H$, the model records $F_{\mathrm{spill}} = \rho A (L^{*} - H)/h$ over substep h, accumulates the spilled mass, and sets $L = H$. All reported operating comparisons remain below the lip, so spill is zero in those runs.

The valve delivers the flow of Equation (2.2), which is the standard incompressible sizing relation written with lift in percent.

$$F_{\mathrm{in}} = \rho\, C_v\, u \sqrt{\Delta P_{\mathrm{in}} / g_s} \tag{2.2}$$

where C_v is the valve coefficient, u is the lift in percent of travel, ΔP_{in} is the pressure drop across the valve in bar and g_s is the specific gravity of the liquid. The constant C_v carries units of m^3/s per percent of lift per $\sqrt{\mathrm{bar}}$, so the product is a mass flow in kg/s. Flow is linear in the lift, which means the valve has no installed characteristic of its own, and it is proportional to the square root of the drop, which is the only nonlinearity in the plant. The inlet flow gain of Equation (2.3) is the quantity every structure in this chapter depends on.

$$K_v(\Delta P_{\mathrm{in}}) = \frac{\partial F_{\mathrm{in}}}{\partial u} = \rho\, C_v \sqrt{\Delta P_{\mathrm{in}} / g_s} \tag{2.3}$$

At the design drop of 12 bar the gain is 0.3464 kg/s per percent, at 2 bar it is 0.1414 kg/s per percent and at 22 bar it is 0.4690 kg/s per percent, a factor of 3.32 across the range the supply sinusoid sweeps.

The leak is taken proportional to level, as shown in Equation (2.4).

$$F_{\mathrm{leak}} = k_L\, L \tag{2.4}$$

where k_L is the leak coefficient. This term is what makes the tank self-regulating instead of integrating. Holding the valve still and waiting gives the steady level of Equation (2.5), and holding the level still

requires the valve position of Equation (2.6).

$$L_\infty = \frac{K_v(\Delta P_{\text{in}})\,u - F_{\text{out}}}{k_L} \tag{2.5}$$

$$u_\infty = \frac{F_{\text{out}} + k_L L}{K_v(\Delta P_{\text{in}})} \tag{2.6}$$

The right-hand side of Equation (2.6) is what the positional control law carries as its bias. At the design point of $L = 1.0\,\text{m}$, $F_{\text{out}} = 2.0\,\text{kg/s}$ and $\Delta P_{\text{in}} = 12\,\text{bar}$ it evaluates to 20.21 %, passing 7.0 kg/s of which 2.0 kg/s leaves through the outlet and 5.0 kg/s through the leak.

2.3.3 Disturbance Inputs

The outlet demand is a metered draw set directly by the operator, so it enters Equation (2.1) additively and is completely independent of the level and of the valve. The disturbance protocol steps it from 2 kg/s to 12 kg/s and back, in steps of exactly $\pm 10\,\text{kg/s}$.

The supply pressure drop is the second upset and it enters multiplicatively through Equation (2.2), as shown in Equation (2.7).

$$\Delta P_{\text{in}}(t) = \max\bigl(0,\ \overline{\Delta P}_{\text{in}} + a\,\sin(t/10)\bigr) \tag{2.7}$$

where $\overline{\Delta P}_{\text{in}}$ is the operator-set mean, a is the amplitude and t is the simulated time in seconds. The default $\overline{\Delta P}_{\text{in}} = 12\,\text{bar}$ and $a = 10\,\text{bar}$ sweep the drop between 2 bar and 22 bar with a period of $2\pi \times 10 = 62.8\,\text{sec}$, and setting the amplitude to zero removes the swing for step tests. The sinusoid runs on the simulated clock and its phase resets with the tank, so a run is repeatable at any speed setting.

The square root has two consequences that precede any controller. The swing is not symmetric in flow, because at a fixed lift of 20.21 % the inlet flow ranges from 2.86 kg/s to 9.48 kg/s rather than symmetrically about 7.0 kg/s. The mean is also displaced, because the average of $\sqrt{\Delta P_{\text{in}}}$ over one cycle is 3.276 against $\sqrt{12} = 3.464$, so a zero-mean pressure swing lowers the mean delivered flow by 5.4 % and the steady lift needed to hold 1.0 m rises from 20.21 % to 21.37 %. Integral action must supply that 1.16 % of lift even though the disturbance itself has zero mean.

2.3.4 Instrumentation and Numerics

Three transmitters are simulated and the controllers are permitted to read nothing else. Each adds an independent zero-mean Gaussian deviate drawn by the Box-Muller transform and each reading is floored at zero, as shown in Equation (2.8).

$$\begin{aligned} L^m &= \max(0,\ L + \sigma_L\,\varepsilon_1) \\ F_{\text{in}}^m &= \max(0,\ F_{\text{in}} + \sigma_F\,\varepsilon_2) \\ F_{\text{out}}^m &= \max(0,\ F_{\text{out}} + \sigma_F\,\varepsilon_3) \end{aligned} \tag{2.8}$$

where $\sigma_L = 0.004\,\text{m}$ is the level transmitter noise, $\sigma_F = 0.05\,\text{kg/s}$ is the noise on both flow transmitters and ε_i are independent standard normal deviates. There is no transmitter lag, no span error, no bias and no filtering. The app draws its noise from the browser's unseeded generator, as the standalone apps of Section 1.6.1 do, so repeated visits differ. The reference experiments below inject three declared random streams into the shipped model for reproducibility.

In manual mode the operator sets the valve lift directly, and the app reports the flow that lift passes at the current mean pressure and the steady level it produces from Equation (2.5). Stepping the lift with the pressure swing at zero amplitude is the step test of Section 2.6.2, from which the tuning is identified.

The plant is advanced by explicit Euler in substeps of at most 0.05 sec, and both controllers execute once per 1.0 sec of simulated time, which is twenty substeps. Simulated time advances at

Table 2.2 Simulation Parameters and Default Tuning

Symbol	Meaning	Value	Group
ρ	liquid density	1000 kg/m^3	fixed
C_v	valve coefficient	1.0×10^{-4}	model
A	tank area	5.0 m^2	model
g_s	specific gravity	1.0	model
H	lip height	2.5 m	model
σ_L	LT-301 noise	0.004 m	instrument
σ_F	FT-302, FT-303 noise	0.05 kg/s	instrument
k_L	leak coefficient	5.0 kg/s per m	disturbance
F_{out}	outlet demand	2.0 kg/s	disturbance
$\overline{\Delta P}_{\text{in}}$	mean supply drop	12 bar	disturbance
a	pressure swing amplitude	10 bar	disturbance
Δt	controller sample period	1.0 sec	fixed
h	integration substep	0.05 sec	fixed
K_{c1}	level gain, direct to valve	20 %/m	tuning
K_{I1}	level reset, direct to valve	0.4	tuning
K_{c1}	level gain, cascade	6.25	tuning
K_{I1}	level reset, cascade	0.125	tuning
K_{D1}	level derivative gain	0	tuning
u_{bias}	level loop bias, direct	20.2 %	tuning
u_{bias}	level loop bias, cascade	7.00 kg/s	tuning
K_{ff}	static feedforward gain	2.887	tuning
K_{c2}	flow loop gain	1.0	tuning
K_{I2}	flow loop reset	1.0	tuning
u_{bias2}	flow loop bias	20.2 %	tuning

$1\times$ to $10\times$ wall time, and as in the framework apps the speed changes the pace and not the result. Within one control execution the order is fixed, the three transmitters are read first, then the selected structure computes a lift, then the lift is applied, then the sample is recorded. Because the transmitters are read before the new lift is applied, the measured inlet flow at sample k reflects the lift commanded at sample $k-1$, which puts one sample of delay in the inner cascade loop and matters for its stability limit.

2.3.5 Parameters

Every parameter of the simulation is listed in Table 2.2 with its default value. No parameter is hidden, and only the liquid density, the controller sample period and the integration substep are fixed in code.

2.3.6 Assumptions

The plant is deliberately simple so that the structural comparison is not contaminated by effects that would need their own explanation. Each assumption below is stated with the condition under which it stops being reasonable.

The tank contents are perfectly mixed and the cross-section is uniform, so one lumped level describes the inventory. This fails for a horizontal drum or a dished head, where the area varies with level and the time constant varies with it.

The valve is algebraic. There is no stroking time, no stiction, no deadband and no positioner dynamics, so a commanded lift appears as a flow within the same sample. The inner cascade loop is the most sensitive to this assumption. A real valve with a stroking time of two or three seconds would put a lag inside the inner loop comparable with the 1 sec sample period, which lowers the inner gain that can be used, and with it the pressure rejection the cascade delivers.

The supply pressure drop is exogenous. Opening the valve does not pull the header down, which fails when the tank inlet takes a large fraction of header capacity and the disturbance becomes partly a consequence of the manipulated flow.

The outlet is a perfectly metered draw independent of level, which is the behavior of a downstream pump on flow control. A gravity outlet would make F_{out} a function of L and would add self-regulation the level loop does not currently have.

The leak is linear in level rather than proportional to $\sqrt{L}$ as an orifice would be. Linearity keeps the plant exactly first order at every operating level, which is convenient for the structural argument but means the gain and the time constant do not move with the level as they do on a gravity-drained vessel. Volume 1, Chapter 2 treats such a vessel, whose gain and time constant both follow the square root of the level.

The transmitters are instantaneous and their noise is white. Real differential-pressure level transmitters carry several seconds of lag and real flow measurements are noisier at the bottom of their span, both of which limit how hard the inner loop and the feedforward gain can be pushed.

There is no transport delay anywhere in the plant. This is the assumption that most favors feedforward, because the ideal compensator for this plant is a constant and remains realizable. A delay in either the manipulated path or the disturbance path would make the ideal compensator improper and would force the approximations that occupy the feedforward design literature [15].

Explicit Euler at 0.05 sec is accurate here because the substep is four orders of magnitude below the 1000 sec level time constant and the only other dynamics in the plant are algebraic.

2.4 Dynamic Analysis

Three models must be kept apart in what follows. The simulation model is Equations (2.1) to (2.7) and is what the application integrates. The control model is the linearization below and is what the tuning and the structural argument are built on. The controller's internal model is nothing at all, because both controllers are fixed-gain proportional-integral algorithms that carry no plant description.

2.4.1 Linearization About the Design Point

Perturbing Equations (2.1) to (2.4) about $L_0 = 1.0\,\text{m}$, $u_0 = 20.21\,\%$, $\Delta P_0 = 12\,\text{bar}$ and $F_{\text{out},0} = 2.0\,\text{kg/s}$ gives the first-order form of Equation (2.9).

$$\tau \frac{\mathrm{d}L'}{\mathrm{d}t} + L' = K_p\, u' + K_d\, F'_{\text{out}} + K_q\, \Delta P'_{\text{in}} \tag{2.9}$$

where the primes denote deviations from the design point and Equation (2.10) defines the four constants.

$$\begin{aligned} \tau &= \rho A / k_L, \quad & K_p &= K_v / k_L, \\ K_d &= -1/k_L, \quad & K_q &= F_{\text{in},0} / (2\,\Delta P_0\, k_L) \end{aligned} \tag{2.10}$$

At the design point these evaluate to $\tau = 1000\,\text{sec}$, $K_p = 0.0693\,\text{m}$ per percent, $K_d = -0.200\,\text{m}$ per kg/s and $K_q = 0.0583\,\text{m/bar}$. The three transfer functions from Equation (2.9) share the same single pole, which is the fact the rest of this section exploits.

2.4.2 Ideal Feedforward Compensator

The ideal feedforward compensator for a measured disturbance is the negative ratio of the disturbance transfer function to the manipulated transfer function, whose general form with a lead-lag and a delay is Equation (1.4). For the outlet demand that ratio is given by Equation (2.11).

$$G_{\text{ff}}(s) = -\frac{G_d(s)}{G_p(s)} = -\frac{K_d/(\tau s + 1)}{K_p/(\tau s + 1)} = \frac{1}{K_v(\Delta P_{\text{in}})} \tag{2.11}$$

The pole cancels exactly, so the compensator is a pure gain with no lead, no lag and no delay. This is not a modeling approximation. It is a structural property of the plant, because the outlet demand and

the manipulated inlet flow enter the same integrator at the same point, so the dynamics between the disturbance and the level are identical to the dynamics between the valve and the level, and inverting the ratio leaves the algebraic valve gain behind. At 12 bar Equation (2.11) evaluates to 2.8868 % per kg/s, which is the default value of 2.887. At 4 bar the exact value is 5.000 and at 22 bar it is 2.132, and the application reports the exact value for the current mean pressure.

🕮 **Pause.** Try Exercise 2.2 before reading on.

The same construction applied to the pressure upset gives $K_q/K_p = u_0/(2\Delta P_0) = 0.842\,\%/\text{bar}$, also a constant. A pressure feedforward is realizable in principle, but it would be exact only to first order because the true requirement is $u \propto 1/\sqrt{\Delta P_{\text{in}}}$, and it would need a pressure transmitter and an explicit model of the valve. The application does not provide that path. The cascade reaches the same upset without any valve model at all, which is the practical reason cascade is preferred for supply-side disturbances.

2.4.3 Sensitivity at the Disturbance Frequency

The pressure swing has a single frequency, $\omega_d = 0.1\,\text{rad/s}$, and the two loops meet it with very different authority. The level loop sees the process through the 1000 sec lag, so at ω_d the plant contributes $|G_p(j\omega_d)| = K_p/\sqrt{1+(\omega_d\tau)^2} = 6.93\times10^{-4}$ m per percent, a factor of 100 below its static gain. The velocity-form proportional-integral controller contributes $|C_1| = 20.6\,\%/\text{m}$ at the same frequency, so with the one-sample measurement delay the loop gain is $|L_1(j\omega_d)| = 0.0143$ and the sensitivity is $|S_1(j\omega_d)| = 1.005$. The level loop does not attenuate the pressure swing. It amplifies it by half a percent.

The flow loop sees an algebraic plant. Its loop transfer function is $C_2(z)K_v z^{-1}$, with the sample delay identified above, so at ω_d the loop gain is $|L_2| = 3.50$ and the sensitivity is $|S_2| = 0.271$. The inner loop removes 72.9 % of the flow upset at the frequency at which the outer loop removes none of it. That single pair of numbers is the whole argument for cascade on this plant, and it follows from the position of the measurement rather than from any tuning constant.

> **TRY IT IN THE APP 6 · Cascade, Pressure Swing**
>
> Press PID auto, choose Cascade with its outlet trim off, turn the sin(t/10) swing on and leave the demand alone for at least 200 s. Watch the valve lift: it sweeps widely every cycle while the level hardly moves. In the chapter's run the lift spans 14.80 to 42.17 % and the level ripples only 0.0047 m peak to peak, against 0.0214 m under feedback.

The ratio of speeds is far beyond the usual design guidance. Cascade is normally recommended when the inner loop is three to ten times faster than the outer one [4], and here the inner loop settles within a few samples against a 1000 sec outer lag. Section 1.4.5 states the rule as three to five times faster and measures on its heater that this range keeps most of the benefit against an upset inside the inner loop and costs a little against one outside it, while a factor of ten costs nothing measurable there. The tank exceeds both statements of the rule by a wide margin.

2.5 Controller Design

2.5.1 Control Algorithm

Both controllers are the same velocity-form proportional-integral-derivative algorithm and differ only in their gains, their limits and what they measure. The increment applied each sample is shown in Equation (2.12).

$$\begin{aligned} e_k &= \mathrm{SP} - \mathrm{PV}_k \\ s_k &= s_{k-1} + \Delta P_k + \Delta D_k \\ c_k &= s_k + I_{k-1} + K_I e_k \Delta t \\ u_k &= \mathrm{clamp}\big(c_k,\, u_{\min}, u_{\max}\big) \\ I_k &= I_{k-1} + K_I e_k \Delta t + (u_k - c_k) \end{aligned} \tag{2.12}$$

where e_k is the error, s_k is the running accumulator, I_k is the integral term and the increments ΔP_k and ΔD_k are selected by the algorithm variant of Equation (2.13), written with the second difference $\delta^2 x_k = x_k - 2x_{k-1} + x_{k-2}$.

$$\begin{aligned} &\text{A:} \quad \Delta P_k = K_c(e_k - e_{k-1}), && \Delta D_k = \tfrac{K_D}{\Delta t}\,\delta^2 e_k \\ &\text{B:} \quad \Delta P_k = K_c(e_k - e_{k-1}), && \Delta D_k = -\tfrac{K_D}{\Delta t}\,\delta^2 \mathrm{PV}_k \\ &\text{C:} \quad \Delta P_k = -K_c(\mathrm{PV}_k - \mathrm{PV}_{k-1}), && \Delta D_k = -\tfrac{K_D}{\Delta t}\,\delta^2 \mathrm{PV}_k \end{aligned} \tag{2.13}$$

where K_c is the proportional gain, $K_I = K_c/\tau_I$ is the integral gain in output units per unit of error per second and $K_D = K_c\tau_D$ is the derivative gain. Type B is the default on the level loop and the only variant available on the flow loop. Opening the valve raises both the inlet flow and the level, so the process gain is positive in both loops and both controllers are reverse acting, with the error computed as setpoint minus measurement and positive gains. With $K_D = 0$ and a setpoint that does not move, all three variants produce the same increment, so none of the results in this chapter depends on which one is selected.

Anti-windup is part of the recursion rather than a stage added after it. The accumulator s_k is never clipped, and the integral term I_k absorbs the whole of the clipping through the back-calculation term $(u_k - c_k)$, which is applied only when $K_I \neq 0$. This is the arrangement of Section 1.6.2, with s_k in the role of S_k there and c_k in that of v_k, except that c_k already holds the current integral increment and the whole clipping is absorbed at once, as with a tracking time constant equal to the sample period. Substituting the fourth line of Equation (2.12) into the fifth gives $c_k = u_{k-1} + \Delta P_k + K_I e_k \Delta t + \Delta D_k$, so with integral action the recursion is algebraically identical to clipping the running sum and none of the results below is affected by the distinction. It matters only at $K_I = 0$, for the reason Section 1.6.2 gives. An output parked at a limit stores only the integral contribution consistent with the bounded command, and can leave the limit when the net increment points into range. Transfer into automatic

and every edit of a gain, a setpoint or a bias re-seed the recursion at the positional law $u = u_{\text{bias}} + K_c e$ with the measurement history filled from the current reading. This prevents a derivative kick, but does not guarantee continuity of valve position: the positional bias and error, plus the next integral increment, may differ from the previous command. It makes a proportional-only setting follow that positional law.

The default level loop tuning is $K_{c1} = 20\,\%/\text{m}$ and $K_{I1} = 0.4$, which is $\tau_{I1} = 50\,\text{sec}$. Internal model control on the identified first-order plant would give $K_c = \tau/(K_p(\tau_c + \theta))$ with $\tau_I = \tau$, so $\tau_c = 720\,\text{sec}$ reproduces the default gain but calls for $\tau_I = 1000\,\text{sec}$ [19], and the SIMC modification, which caps the reset at $4(\tau_c + \theta)$ to improve load rejection on lag-dominant plants, still calls for the full 1000 sec at this τ_c [20]. The default reset is twenty times faster than that, which is a deliberate departure. The tank is slow enough that the internal model control reset would need roughly 3000 sec to clear a load step, and the loop tolerates the aggressive setting because the actuator has no lag of its own to destabilize it.

2.5.2 Structure Definitions

The four structures are shown in Figure 2.2, and the calculations that produce the valve signal are collected in Equations (2.14) to (2.17). In all four the manipulated variable is the same valve lift, the sample period is the same 1 sec and the plant is untouched.

Structure 1 is the plain level loop of Equation (2.14), whose output is the valve lift and whose limits are the valve limits.

$$u_k = \text{PI}_1(L_k^m), \qquad 0 \le u_k \le 100 \tag{2.14}$$

Structure 2 subtracts a trim proportional to the deviation of the measured inlet flow from its design value, as shown in Equation (2.15).

$$u_k = \text{clamp}\left(\text{PI}_1(L_k^m) - K_{\text{ff}}\left(F_{\text{in},k}^m - F_{\text{in},0}\right), 0, 100\right) \tag{2.15}$$

where $F_{\text{in},0} = 2.0\,\text{kg/s} + k_L L^{\text{SP}}$, which is 7.0 kg/s at the default setpoint. The sign is negative because a rise in inlet flow calls for the valve to close.

Structure 3 adds a trim proportional to the deviation of the measured outlet demand from its design value, as shown in Equation (2.16).

$$u_k = \text{clamp}\left(v_k + K_{\text{ff}}\left(F_{\text{out},k}^m - F_{\text{out},0}\right), 0, 100\right) \tag{2.16}$$

where $v_k = \text{PI}_1(L_k^m)$ is the level controller output and $F_{\text{out},0} = 2.0\,\text{kg/s}$. The reference is the fixed 2.0 kg/s of the design point rather than the current demand, so a demand already displaced when the structure is selected leaves a standing trim that the level loop absorbs through its bias.

Structure 4 replaces the direct connection with a cascade, as shown in Equation (2.17).

$$\begin{aligned} F_{\text{in},k}^{\text{SP}} &= \text{clamp}\left(v_k + r\left(F_{\text{out},k}^m - F_{\text{out},0}\right), 0, 40\right) \\ u_k &= \text{PI}_2\left(F_{\text{in},k}^m\right), \qquad 0 \le u_k \le 100 \end{aligned} \tag{2.17}$$

where $r = K_{\text{ff}}/K_{\text{ff}}^0$, $K_{\text{ff}}^0 = 1/K_v(12\,\text{bar}) = 2.8868$ and the trim term is present only when the outlet feedforward option is enabled. Dividing by K_{ff}^0 converts the gain from percent of lift per kg/s into kg/s of flow setpoint per kg/s of demand, where complete feedforward is unity, so the same number in the same field means full compensation in either place. Selecting the cascade also changes the units of the level loop's output, so the application switches K_{c1} from 20 %/m to 6.25 kg/s per m and the bias from 20.2 % to 7.00 kg/s. The two settings are close to the same loop gain expressed in different units, because $20 \times 0.3464 = 6.93$ kg/s per m against the 6.25 the cascade uses.

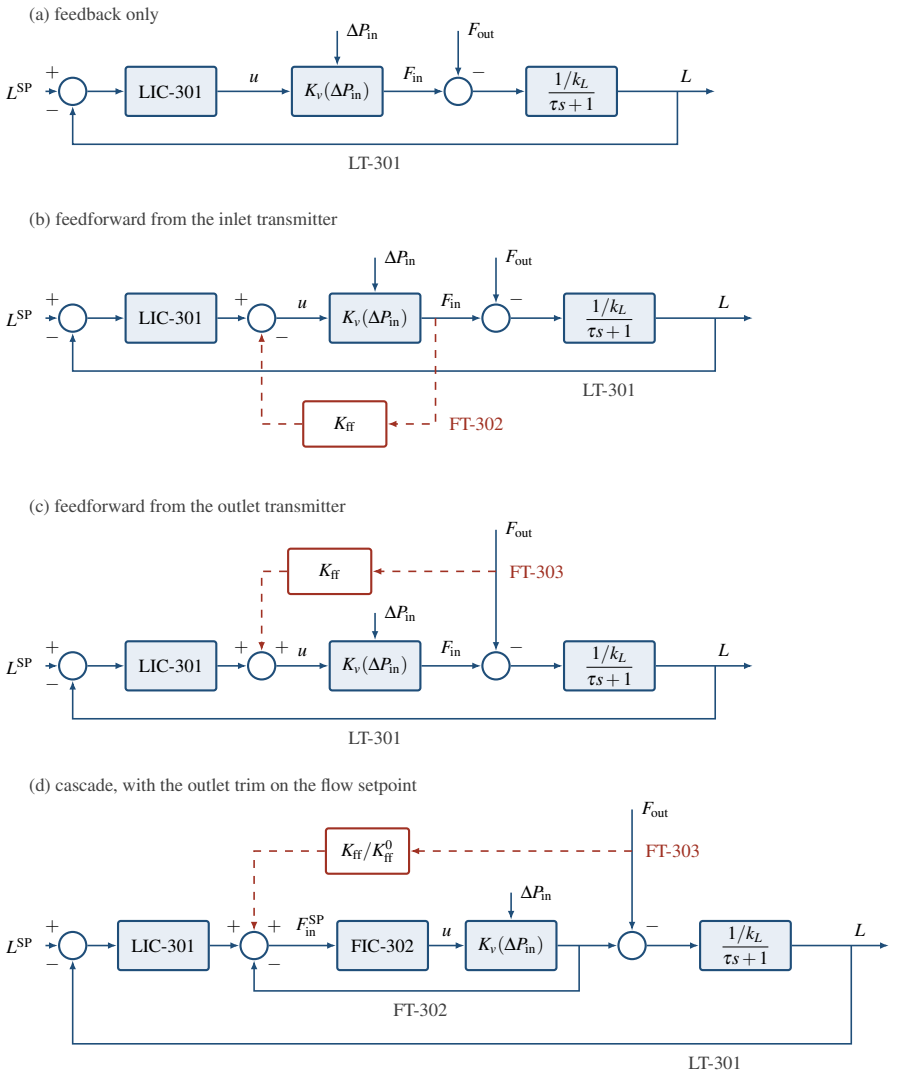


Figure 2.2 The four structures available on the same tank. Solid paths are always active and dashed paths carry the feedforward trim. The supply pressure enters at the valve gain block K_v and the outlet demand enters at the summing junction after it, so (b) taps a signal on the manipulated path while (c) and (d) tap the disturbance.

2.5.3 Inner-Loop Stability Limit

The inner loop is unusual in having no plant dynamics at all, only the one sample of measurement delay identified above. Section 1.4.5 closes an inner loop around a valve lag, with the time constant of Equation (1.7). Here that lag is absent, and the limit comes from the sample delay alone. Writing $F_k^m = K_v u_{k-1}$ and substituting the velocity form gives the characteristic polynomial of Equation (2.18).

$$z^2 + \left[\left(K_{c2} + K_{I2}\Delta t\right)K_v - 1\right]z - K_{c2}K_v = 0 \tag{2.18}$$

The Jury conditions for both roots inside the unit circle reduce to the pair of inequalities in Equation (2.19).

$$K_{c2}K_v < 1 \qquad \text{and} \qquad \left(2K_{c2} + K_{I2}\Delta t\right)K_v < 2 \tag{2.19}$$

For the default $K_{c2} = K_{I2} = 1.0$ at 12 bar the roots are 0.762 and -0.455, so the loop settles in a few samples with a mild alternating component. Holding $K_{c2} = K_{I2} = K$ and $\Delta t = 1$ sec, the second inequality binds and the limit is $K < 2/(3K_v)$, which is 1.925 at 12 bar. The limit moves with the supply pressure, falling to 1.421 at 22 bar and rising to 4.714 at 2 bar, so the sinusoid carries the inner loop in and out of the stable region during every cycle whenever K lies between those two values.

2.5.4 Algebraic Feedback in the Inlet Feedforward

Structure 2 degrades performance, and the mechanism is more specific than the observation that the inlet flow is not the disturbance. Substituting $F^m_{\text{in},k} = K_v u_{k-1}$ into Equation (2.15) shows that the trim closes an algebraic feedback path from the valve to itself, as shown in Equation (2.20).

$$u_k = v_k + K_{\text{ff}} F_{\text{in},0} - K_{\text{ff}} K_v u_{k-1} \tag{2.20}$$

where v_k is the level controller's output. The pole of that recursion sits at $z = -K_{\text{ff}} K_v$. Choosing K_{ff} by Equation (2.11), which is the correct value for the disturbance path, places the pole at exactly $z = -1$ at the design pressure. The default configuration therefore sits precisely on the boundary of a period-two oscillation, and the pressure sinusoid pushes $|K_{\text{ff}} K_v|$ above unity for the half of every cycle in which the drop exceeds 12 bar. The static gain of the same recursion is $1/(1 + K_{\text{ff}} K_v) = 0.5$, so the trim also halves the authority the level controller has over the valve, which is why this structure is worse than plain feedback on a load upset and not merely no better. Section 1.4.7 lists this case among the ways a structure applied to the wrong path does harm.

> **TRY IT IN THE APP** **4 · The Wrong Feedforward**
>
> Press PID auto, choose FF inlet and leave the sin(t/10) swing off. Press Step +10, later Step −10, and run 2 min past the last step. Watch the valve trend move up and down on almost every sample. The chapter reports reversals on 98.0 % of increments with the swing off and a worst level error of 0.3510 m, against 0.2777 m for plain feedback.

2.6 Results

2.6.1 Protocol

Reference trajectories use the current shipped model and controller functions, with twenty Euler steps of 0.05 sec per 1 sec control execution. The ordering reproduces the application: read all transmitters, calculate control, apply the lift, record, then integrate. Selection of automatic control first seeds both recursions from their positional biases. Model time is t. The application labels its first pre-integration sample as elapsed second 1, so exported elapsed time is $t + 1$. Errors use true level, IAE sums absolute error times the sample period, travel sums successive lift changes, and saturation is the fraction of commands at either lift bound.

Each comparison runs 900 sec from the design point, with default noise and LCG seeds 1, 2 and 3. The recurrence is $x_{n+1} = (1664525 x_n + 1013904223) \bmod 2^{32}$, and $x_n / 2^{32}$ supplies the model's Box–Muller draws. Each run draws once for the automatic-control seed and then once per transmitter per execution. Means and half-ranges describe these three runs, not confidence intervals. A noise-free cascade-with-feedforward run is also operated in the built browser application and checked by CSV export. All 900 samples agree to export rounding: 0.00001 m for true level and 0.0001 for lift, flow and pressure.

The steps raise demand from 2 to 12 kg/s and return it to 2 kg/s. Declaring the half-sample phase matters when feedforward nearly cancels the load.

2.6.2 Step Test and Valve Gain

With the swing off, a manual step of the lift from 20.2 % to 30 % identifies $K_p = 0.069$ m per percent and $\tau = 1000$ sec from the export, the values of Equation (2.10). The same step repeated at a mean drop of 4 bar shows the gain fall by $\sqrt{4/12} = 0.58$, the factor by which Equation (2.3) lowers

Table 2.3 Disturbance Scenarios: Demand Steps Halfway Between Executions

Scenario	Pressure swing	Outlet demand
Combined	$10\sin(t/10)$ bar	steps at 100.5, 350.5 s
Pressure only	$10\sin(t/10)$ bar	held at 2 kg/s
Demand only	off	steps at 100.5, 350.5 s

Table 2.4 Combined Scenario: Mean of Three Seeded Runs, With the Worst-Error Half-Range

Structure	max $\lvert e\rvert$ (m)	IAE (m s)	MV travel (%)	Saturated (%)
Feedback only	0.2989 ± 0.0002	145.9	103	0.0
FF from FT-302	0.3590 ± 0.0021	142.0	16295	8.1
FF from FT-303	0.0444 ± 0.0002	12.5	232	0.0
Cascade	0.2929 ± 0.0002	144.8	1219	0.0
Cascade with FF	0.0136 ± 0.0001	3.5	1122	1.1

Table 2.5 Mean Worst True-Level Error for Separate Upsets

Structure	Pressure only (m)	Demand only (m)
Feedback only	0.0184	0.2777
FF from FT-302	0.0162	0.3510
FF from FT-303	0.0183	0.0011
Cascade	0.0039	0.2927
Cascade with FF	0.0039	0.0046

K_v. This is the gain variation the cascade absorbs by measuring the delivered flow, and the static feedforward gain is exact only at the pressure it was computed for.

2.6.3 Combined Scenario

Table 2.4 gives the combined result. The best and worst structures differ by a factor of 26 in worst error with the shipped tunings.

Plain feedback reaches its worst error at model second 316 (exported second 317), while the valve is still ramping against the larger demand. The first seeded run finishes at 1.209 m, outside the ± 0.02 m band. Inlet feedforward has larger worst error and about 158 times the valve travel. It reverses valve direction on 70.0% of consecutive command increments and saturates for 8.1% of samples. With the swing off, its reversal fraction is 98.0%. This is the alternating mode of Equation (2.20), rather than useful anticipation of outlet demand. Its slightly smaller IAE would mislead a comparison that ignored peak error and actuator movement.

Pause. Try Exercise 2.3 before reading on.

Outlet feedforward cuts worst error by a factor of 6.7 and IAE by 11.7. The trim moves the valve by about 28.9% when demand changes, raising travel from 103 to 232%. Cascade alone gives almost the same worst error as feedback, while using 1219% travel. Adding the outlet trim reduces worst error to 0.0136 m, about 3.3 times below outlet feedforward alone.

2.6.4 Separated Scenarios

The separate upsets explain that ranking, as shown in Table 2.5. Cascade divides the pressure error by 4.7, but does not improve the demand error. Outlet feedforward divides demand error by about 250, while leaving pressure error almost unchanged. The two structures combined address both entry points. Inlet feedforward is worse against demand, and its small pressure-error reduction comes with oscillatory valve movement. Figure 2.3 draws the same table as bars, one panel for each upset.

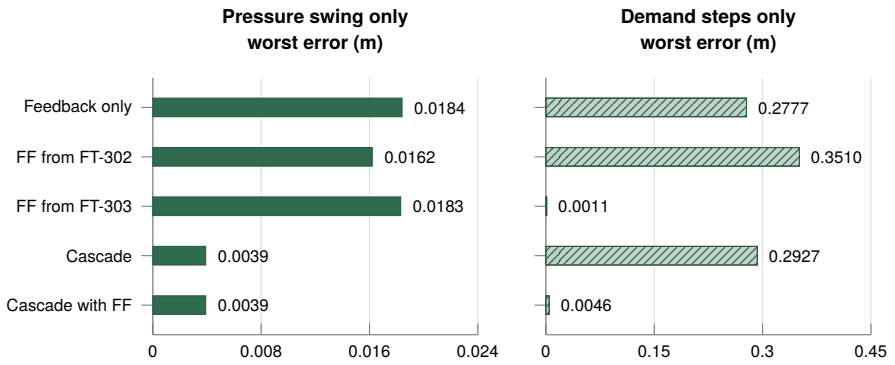


Figure 2.3 The worst level error of Table 2.5 for each structure under one upset at a time, the pressure swing alone (solid bars) and the demand steps alone (hatched bars), each panel on its own scale. Notice that the cascade shortens only the pressure bar and the outlet feedforward only the demand bar, that feedforward from the inlet transmitter lengthens the demand bar, and that only the cascade with the outlet trim is short in both panels.

Table 2.6 Inner Flow Loop Gain Sweep: Means of Three Seeded Runs

K	Roots at 12 bar	Stable below	max $\lvert e \rvert$ (m)	MV travel (%)	Saturated (%)
1.0	0.762, −0.455	44.4 bar	0.0039	254	0.0
1.5	0.701, −0.741	19.8 bar	0.0028	391	0.0
2.0	0.662, −1.047	11.1 bar	0.0744	7599	17.8
3.0	0.614, −1.692	4.9 bar	0.2798	17661	54.3

🕮 **Pause.** Try Exercise 2.1 before reading on.

For the first seeded pressure-only run, over the final 300 s feedback keeps lift between 21.51 and 22.21%, while inlet flow spans 3.04–10.39 kg/s and true level ripples 0.0214 m peak to peak. Cascade sweeps lift through 14.80–42.17%, inlet flow through 5.17–8.90 kg/s, and level through only 0.0047 m. It converts level excursion into actuator travel.

With constant pressure, the ideal trim cancels demand exactly in the noise-free continuous relation. The declared half-sample delay loses $10(0.5)/(\rho A) = 0.001$ m before compensation, and transmitter noise adds to the observed 0.0011 m residual. With swinging pressure the fixed trim is exact only at the mean pressure: a 10 kg/s step needs 21.3% lift at 22 bar but 70.7% at 2 bar, whereas the trim always requests 28.9%. This operating-point mismatch explains the 0.0444 m combined residual [16]. The cascade measures delivered flow rather than predicting it from one valve gain.

2.6.5 Inner-Loop Gain Sweep

The pressure-only test runs for 300 sec, with the same three seeds and $K_{c2} = K_{I2} = K$, not just a change in proportional gain. The pressure boundary solves $(2K_{c2} + K_{I2}\Delta t)K_v = 2$. Raising K from 1.5 to 2.0 increases travel by a factor of 19.4 and worst error by 26.6. The nominal boundary $K = 1.925$ lies between those values. At $K = 2$, mean successive movement is 38.4% above 20 bar but only 1.5% below 4 bar. Stability therefore changes within each pressure cycle. At $K = 1.5$, the frozen-pressure boundary is exceeded briefly near the crest, yet the finite-time response stays small. This is a local stability analysis of a moving operating point, not a guarantee based on the mean pressure.

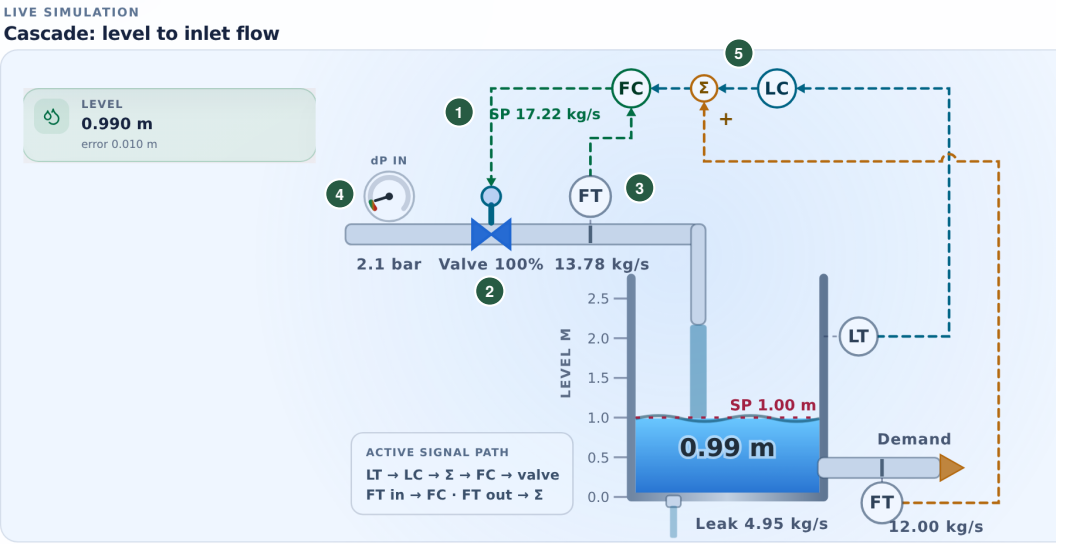


Figure 2.4 Plant view of the current application in a noise-free cascade-with-outlet-feedforward run, with the level tile of the status column set into the view's empty upper left corner, a composite of two parts of one screenshot. Demand rises from 2 to 12 kg/s at model second 100.5 (Table 2.3) while the pressure swing remains on, and the capture is taken at 300 s. The crop shows pressure at about 2.1 bar (4), lift at 100% (2) and, in the tile, level at 0.990 m. The commanded 17.22 kg/s flow setpoint (1) of the flow controller exceeds what the valve can deliver near the trough, as the inlet flow transmitter (3) shows, exposing the actuator constraint while the tank inventory limits the level excursion. The outlet trim joins the level controller's output at the summing junction (5).

> ✦ **DID YOU KNOW?** **Half a Unit Too Much Inner Gain**
>
> From the simulator: on the pressure-only test, raising both inner gains from 1.5 to 2.0 lifts the worst level error from 0.0028 to 0.0744 m, the valve travel from 391 to 7599 % and the time at a valve stop from 0.0 to 17.8 % (Table 2.6). The nominal limit of 1.925 lies between the two.

2.6.6 Application

A cascade run with the outlet trim engaged is shown in Figure 2.4.

2.6.7 Scenarios of the Application

The app has no scenario selector. Its thirteen scenarios are set up by hand from the controls, and the app tracks each as a required unit that combines a history predicate with an explicit learner confirmation. Table 2.7 summarizes the automatic predicates. History is segmented when mode, structure, feedforward state, K_{ff} or either inner gain changes. A demand-step segment needs both an upward and a downward step, 120 s after the last step, and no pressure swing after the first step. A pressure segment needs 200 s of swing since its last demand step. These checks establish activity, not a graded numerical explanation or independently measured learning outcome.

The inner-boundary unit checks proportional gain, whereas the stability sweep in Table 2.6 changes both proportional and integral gains. The noise unit combines a current setting with stored history and does not require that history to have been collected under the current noise value. The metric unit requires confirmation of a comparison but does not calculate whether level error improved. Progress and certificates should be interpreted with those limits.

Table 2.7 Scenarios and Completion Requirements

#	Scenario	Kind	Completion requirements
1	Design point	Experiment	A learner-started session and a manual quiet tail of at least 120 s, with swing off and true level within 0.02 m of SP.
2	Feedback steps	Experiment	Automatic feedback and the demand-step segment defined above.
3	Feedback swing	Experiment	Automatic feedback and the pressure segment defined above.
4	Inlet feedforward	Demonstration	Automatic inlet feedforward and a demand-step segment.
5	Outlet feedforward	Experiment	Automatic outlet feedforward and a demand-step segment.
6	Cascade swing	Experiment	Automatic cascade without feedforward and a pressure segment.
7	Cascade steps	Demonstration	Automatic cascade without feedforward and a demand-step segment.
8	Both structures	Experiment	Cascade with feedforward: one pressure segment and one demand-step segment, which can persist across segments.
9	Feedforward gain	Exploration	Outlet feedforward at gains 1.44, 2.887 and 5.77 (specified tolerances), each with a demand step and 120 s afterward.
10	Inner boundary	Demonstration	Cascade at inner proportional gains 1, 1.5 and 3 (tolerance 0.05), each with at least 200 s of swing.
11	Misleading metric	Demonstration	Cascade with both inner gains within 0.05 of 3, in one segment lasting at least 900 s.
12	Square-root gain	Exploration	One manual segment spanning mean pressure at most 4.5 and at least 21.5 bar, with lift variation at most 0.05%.
13	Noise limit	Demonstration	Current flow-noise setting at least 0.3 kg/s plus a cascade segment with inner proportional gain at least 5 and duration at least 120 s.

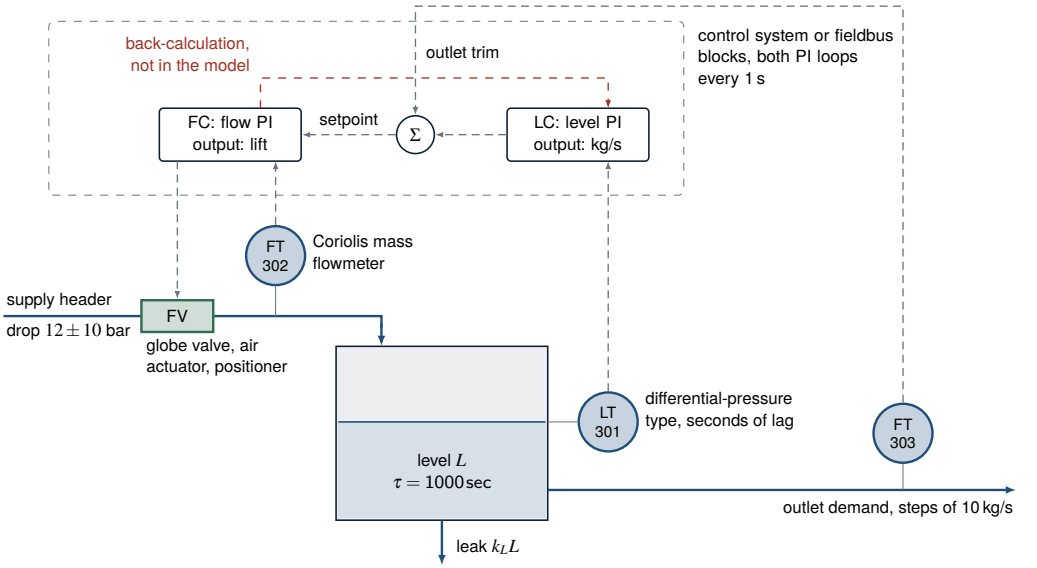


Figure 2.5 The level cascade with its outlet trim on real hardware. The level transmitter LT-301 feeds the level PI, whose output in kg/s is summed with the trim from FT-303 and sent as the setpoint of the flow PI, which reads FT-302 and moves the valve FV. Solid arrows are process streams and dashed gray arrows are signals. Notice the dashed path from the flow controller back to the level controller: a fieldbus sends this back-calculation signal so that the level controller can be told to stop integrating when the valve is at a stop, and the model of this chapter has no such path.

2.7 Technology in Practice

Figure 2.5 draws the loop on real hardware, which adds the lags, the noise and the limits the model leaves out.

What Measures It

A real level transmitter, such as the differential-pressure type that the Assumptions of Section 2.3 name, carries several seconds of lag, little against the 1000 sec level time constant. The flow transmitters matter more, because the inner loop and the outlet trim act on every sample they report. A mass flowmeter on the inlet line would read FT-302 in kg/s, as the model does, and the model's instantaneous reading with white noise of 0.05 kg/s stands for a meter whose own lag is short against the 1 sec sample.

> **TECHNOLOGY** **Sensor: A Coriolis Meter Under the Inner Loop**
>
> A Coriolis meter shakes one or more tubes that carry the liquid at their resonant frequency. Moving mass makes the tubes twist, the phase shift between two points on a tube measures the mass flow, and the frequency gives the density, so the meter reads kilograms per second directly, as FT-302 does in the model [13]. It is linear and accurate and needs no straight pipe ahead of it. Vibration from nearby machines and gas bubbles trapped in bent tubes cause errors in the flow reading, straight tubes are less sensitive at low flow, and the meter is costly in large pipe sizes [13]. An inner loop trusts its meter, so every such error moves the valve.

What Moves It

The final element is the inlet control valve, algebraic in the model: flow follows lift within the sample, linear in the lift and proportional to the square root of the supply drop. A real valve adds its stroking lag inside the inner loop, which the Assumptions name as the effect that lowers the usable inner gain. It is customary in process plants to close yet another loop on the stem with a valve positioner, at the lowest level of the hierarchy, to reduce the effects of friction and nonlinearity [9], so a level cascade on such a valve is three loops deep.

> **TECHNOLOGY** **Final Element: A Pneumatic Valve That Needs Time to Stroke**
>
> A pneumatic actuator moves the stem only as fast as air can fill or vent it. A large actuator takes time to change pressure, so the valve lags its signal. A volume booster, a relay that copies the signal one to one with far more air capacity, shortens the lag but cannot remove it while the air flow is finite [13]. When the pressure drop across a valve changes, the valve also responds more near closed than near open. A flow loop wrapped around the valve confines that distortion to the flow controller, and the level controller sees a more predictable response [13].

Where the Controller Runs

Both controllers would run in a distributed control system or in fieldbus function blocks. Loops there are sometimes nested or cascaded, with cycle times from milliseconds to minutes [21], and the app's 1 sec period sits inside that range. The standard PID function block of a fieldbus carries the chapter's structures as inputs: a cascade input that takes a remote setpoint, a feedforward input with its own gain, and a back-calculation input and output. The flow PID may run inside the transmitter or inside the valve positioner, with no separate controller [13]. In Equation (2.17) the flow setpoint is clamped at 40 kg/s and the lift at 100 %, but no signal passes from the lift clamp back to the level controller. Figure 2.4 shows the valve at that clamp with a setpoint the valve cannot deliver.

What Is Modernizing It

Two developments change where and how a cascade runs. First, control moves into the field: a function block chain propagates the status of every measurement, so a block fed a bad reading can hold its last good value or fail to a set end of its range instead of acting on bad data [13]. Second, the loops of a cascade can close over communication networks, where delay and lost packets eat into the inner loop's margin and schemes that transmit only on events save network load [22].

IN PRACTICE **Keep the Secondary in Cascade, and Tell the Primary**

A secondary controller has three modes: manual, automatic with a local setpoint, and cascade, in which the primary sets its setpoint. Left in automatic, it defeats the cascade as surely as manual defeats any loop [13]. Windup also crosses between the loops. The secondary's anti-windup works in the usual way because it drives the actuator, but the primary must be told when the secondary goes into anti-windup [9]. Fieldbus function blocks pass a back-calculation signal up the chain for this, so a valve that cannot follow, or a selector that drops the primary's output, stops the integral action above it [13].

2.8 Limitations

The plant is a single first-order state and its manipulated path is algebraic, which is the condition under which static feedforward is exactly right and cascade is maximally effective. Both results weaken on a plant with transport delay or valve dynamics, and this application cannot show that weakening because it contains neither.

The performance numbers depend on the default tuning of the application, and that tuning is not optimal for any of the four structures. The level loop's reset of 50 sec against a 1000 sec process is aggressive, and a slower reset would lower the valve travel of every row of Table 2.4 while raising the recovery time. The comparison is between four structures at one tuning, not between four structures each at its own best tuning.

The pressure sinusoid is a single frequency of large amplitude. A single frequency makes the sensitivity argument clean, but it also means the numbers in Table 2.5 say nothing about a broadband supply upset, where the inner loop would attenuate the low-frequency content and pass the content above its own bandwidth. The amplitude is large enough that the square-root nonlinearity contributes harmonics, which is why the measured ripple reduction of 4.6 is not exactly the 3.7 that the ratio of the two linear sensitivities predicts.

Neither the valve position nor the flow setpoint is rate limited, so the cascade is free to sweep the valve 27 % every cycle without penalty. A real installation would either accept the wear or detune the inner loop, and the application offers no way to quantify that trade-off.

The app's random stream is unseeded, so visits are not bit-identical. The reference runs use explicit seeds. Their half-ranges describe only those three realizations, and do not establish the spread over all possible sensor noise.

The model now caps level at the 2.5 m lip and meters spilled mass. Its zero-level floor still clips an infeasible withdrawal without separately metering unmet demand. High-demand dryout therefore needs care in mass-accounting studies, and the results here concern operation near the design level.

OPEN CHALLENGE **An Inner Loop Closed Over a Network**

A recent survey treats cascades whose loops close through a real-time communication network. Delays are then unavoidable in both the inner and the outer loop, and with lost packets they can reduce performance or even cause instability. Schemes that send data only when an event calls for it lighten the network, and cyber-attacks that change the loop's dynamics are part of the problem [22]. On this tank one sample of delay already caps equal inner gains at 1.925 at 12 bar. Progress would be an inner loop whose stability holds with a network's varying delay and losses inside it.

2.9 Discussion

The result that a structure rejects one disturbance and not the other is not a property of this tank. It is a property of where measurements sit relative to where energy and material enter, and it recurs wherever a slow inventory or a slow temperature is held by a fast actuator. A furnace outlet temperature loop that does not cascade to fuel flow will pass every fuel-header pressure swing into the process fluid, and no retuning of the temperature controller will change that, because the temperature controller cannot see the header. Chapter 4 puts flow loops under the fuel and the air of such a heater. A blending

loop without a ratio on the wild stream will chase composition after the fact for the same reason [6], and Chapter 7 blends syrup into water on such a ratio. The measurement is the constraint and the gain is not.

The economic asymmetry follows from the same observation. Adding a flow transmitter and a second controller is a capital project, while changing a gain is a configuration change, so plants under-invest in structure and over-invest in tuning [2]. The measurements in this chapter quantify the return on the added transmitter. The cascade with the outlet trim reaches 0.0136 m of worst error where the default single-loop tuning reaches 0.2989 m, at the reported settings. The comparison does not prove an optimum over all possible tunings, and the default level loop has a loop gain of only 0.0143 at the disturbance frequency. Deciding which variables to measure and control is the decision that bounds every later one [1].

The model of Section 2.3 also supports control methods beyond the four structures implemented. The single state, the algebraic valve law and the two measured disturbances translate directly into a model predictive control formulation, in which one controller coordinates the valve against its 0 % to 100 % travel limits and the level bounds while the square-root pressure dependence enters the prediction model instead of being absorbed as gain variation [23]. Moving horizon estimation or Kalman filtering on the same model would reconstruct the supply pressure drop and the leak coefficient from the three transmitters, replacing the fixed feedforward gain with one that tracks the operating point and removing the 0.0444 m residual the static compensator leaves when the pressure swings. Both methods run on differential and algebraic models of exactly this form in the APMonitor framework [23] and the Gekko Optimization Suite in Python [24], and the exported runs supply the data such an extension would be identified against.

2.10 Conclusion

This chapter placed four control structures on one simulated tank and applied two upsets that enter the plant at different points. The plant is a single mass balance with a square-root valve law, a level-proportional leak, a 1000 sec time constant and three noisy transmitters, and every structure drives the same valve through the same velocity-form proportional-integral algorithm on the same 1 sec sample period. Feedforward from the outlet transmitter is exact rather than approximate on this plant, because the demand and the manipulated flow enter the same integrator and the ideal compensator collapses to the constant 2.887 % per kg/s. Cascade to the inlet flow reaches the supply pressure upset instead, attenuating it by 72.9 % at the 0.1 rad/s swing frequency at which the level loop attenuates none of it. Applying the two upsets separately measures the pairing, because the cascade divides the pressure error by 4.7 and leaves the demand error at 0.2927 m, while the outlet feedforward divides the demand error by about 250 and leaves the pressure error at 0.0183 m. Running both together reaches 0.0136 m on the combined scenario against 0.2989 m for the single loop. This comparison is made at one tuning on a plant with no transport delay and no valve dynamics, and extending it to a delayed manipulated path, a broadband supply upset and a rate-limited valve would show how far the pairing survives. The app lets a reader choose a control structure by measurement rather than by rule of thumb.

KEY POINTS

- Choose a loop structure by where the upset enters the tank, not by tuning.
- The cascade divides the pressure error by 4.7, and outlet feedforward divides the demand error by about 250.
- The inner loop removes 72.9 % of the flow upset at 0.1 rad/s, where the level loop removes none.
- Feedforward from the inlet transmitter puts a pole at $z = -1$ and halves the level controller's authority.

Where else this applies. Flow loops under a heater's fuel and air return in Chapter 4, and inner-loop speed on a fermenter in Chapter 3.

References

[1] S. Skogestad, "Control structure design for complete chemical plants," *Computers & Chemical Engineering*, vol. 28, no. 1-2, pp. 219–234, 2004.

[2] F. G. Shinskey, "Process control: As taught vs as practiced," *Industrial & Engineering Chemistry Research*, vol. 41, no. 16, pp. 3745–3750, 2002.

[3] R. G. Franks and C. W. Worley, "Quantitative analysis of cascade control," *Industrial & Engineering Chemistry*, vol. 48, no. 6, pp. 1074–1079, 1956.

[4] P. R. Krishnaswamy, G. P. Rangaiah, R. K. Jha, and P. B. Deshpande, "When to use cascade control," *Industrial & Engineering Chemistry Research*, vol. 29, no. 10, pp. 2163–2166, 1990.

[5] J. L. Guzmán and T. Hägglund, "Simple tuning rules for feedforward compensators," *Journal of Process Control*, vol. 21, no. 1, pp. 92–102, 2011.

[6] T. Hägglund, "The Blend station – a new ratio control structure," *Control Engineering Practice*, vol. 9, no. 11, pp. 1215–1220, 2001.

[7] 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.

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

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

[10] M. Mehos, H. Price, R. Cable, D. Kearney, B. Kelly, G. Kolb, and F. Morse, "Concentrating solar power best practices study," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/TP-5500-75763, 2020.

[11] W. L. Luyben, "Parallel cascade control," *Industrial & Engineering Chemistry Fundamentals*, vol. 12, no. 4, pp. 463–467, 1973.

[12] Y. Lee, S. Park, and M. Lee, "PID controller tuning to obtain desired closed loop responses for cascade control systems," *Industrial & Engineering Chemistry Research*, vol. 37, no. 5, pp. 1859–1865, 1998.

[13] 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.

[14] D. Gromov, "Analysis of hierarchical structures for hybrid control systems," Ph.D. dissertation, Technische Universität Berlin, Berlin, 2010.

[15] M. Hast and T. Hägglund, "Low-order feedforward controllers: Optimal performance and practical considerations," *Journal of Process Control*, vol. 24, no. 9, pp. 1462–1471, 2014.

[16] E. J. Adam and J. L. Marchetti, "Designing and tuning robust feedforward controllers," *Computers & Chemical Engineering*, vol. 28, no. 9, pp. 1899–1911, 2004.

[17] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[18] ——, "The future of PID control," *Control Engineering Practice*, vol. 9, no. 11, pp. 1163–1175, 2001.

[19] D. E. Rivera, M. Morari, and S. Skogestad, "Internal model control: PID controller design," *Industrial & Engineering Chemistry Process Design and Development*, vol. 25, no. 1, pp. 252–265, 1986.

[20] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[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.

[22] M. Fallahnejad, "A survey of recent breakthroughs in networked cascade control systems," *International Journal of Robust and Nonlinear Control*, vol. 36, no. 5, pp. 2531–2541, 2026.

[23] 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.

[24] 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.

2.11 Exercises

2.1 Concept check · 2 min

With the pressure swing off, the outlet demand steps from 2 to 12 kg/s and back. Plain feedback leaves a worst level error of 0.2777 m and the level-to-flow cascade leaves 0.2927 m (Table 2.5). Why does the cascade not help against this upset?

(a) The inner flow loop is too slow, so it adds a lag in front of the level loop
(b) The inner flow loop runs close to its stability limit at the default gains
(c) The demand enters the balance downstream of FT-302, so the inner loop never sees it
(d) The noise on FT-302 hides the demand step from the inner flow controller

2.2 Concept check · 2 min

Why is the ideal feedforward compensator from the outlet demand to the valve a pure gain on this tank, with no lead, lag or delay?

(a) Because the plant has no transport delay, so every compensator for it is static
(b) Because the integral action of the level loop supplies the missing dynamics
(c) Because the leak makes the tank self-regulating, which removes the pole from the ratio
(d) Because the demand and the inlet flow enter one balance, so both paths share one pole

2.3 Concept check · 1 min

Structure 2 trims the lift by $K_{\mathrm{ff}}(F_{\mathrm{in}}^{m} - F_{\mathrm{in},0})$ with the default $K_{\mathrm{ff}} = 2.887$. What does Equation (2.20) predict at the design pressure?

(a) A pole at $z = -1$, an alternating valve and half the level controller's authority
(b) Exact cancellation of the pressure swing, since 2.887 is the ideal gain $1/K_v$
(c) A pole at $z = +1$, so the lift drifts slowly away from its design value
(d) No change from plain feedback, since FT-302 cannot see the outlet demand

2.4 Theory and calculation · 6 min

The tank of Table 2.2 holds $L = 1.0\,\mathrm{m}$ with $F_{\mathrm{out}} = 2.0\,\mathrm{kg/s}$, $k_L = 5.0\,\mathrm{kg/s}$ per m, $\rho = 1000\,\mathrm{kg/m^3}$, $A = 5.0\,\mathrm{m^2}$, $C_v = 1.0 \times 10^{-4}$ and $g_s = 1.0$.

(a) Find the valve gain K_v at 12 bar and the lift that holds the level. *Checkpoint:* 0.3464 kg/s per percent and 20.21 %.

(b) Find K_p and τ of the linearized model, and the steady level change when the lift steps from 20.2 to 30 % with the swing off. *Checkpoint:* 0.0693 m per percent, 1000 s and 0.679 m.

(c) Find the ideal outlet feedforward gain at 12 bar and at 4 bar, and the factor by which K_p falls at 4 bar. *Checkpoint:* 2.887 and 5.000 % per kg/s, and 0.577.

2.5 Theory and calculation · 8 min

The inner flow loop has $K_{c2} = K_{I2} = K$, $\Delta t = 1\,\mathrm{s}$ and the characteristic polynomial of Equation (2.18). The supply drop follows Equation (2.7) with a mean of 12 bar and an amplitude of 10 bar.

(a) Find the roots at the default $K = 1.0$ and 12 bar. *Checkpoint:* 0.762 and -0.455.

(b) Find the largest stable K at 12 bar, and the drop above which $K = 1.5$ is unstable. *Checkpoint:* 1.925, and 19.75 bar.

(c) For how many seconds of each swing cycle does the drop exceed that boundary for $K = 1.5$? *Checkpoint:* 13.7 s of every 62.8 s.

2.6 Theory and calculation · 6 min

The static outlet feedforward of structure 3 uses the fixed gain $K_{\mathrm{ff}} = 2.887\,\%$ per kg/s while the supply drop swings between 2 and 22 bar.

(a) Find the lift change that passes an extra 10 kg/s at 22, 12 and 2 bar, and compare it with what the trim requests. *Checkpoint:* 21.3, 28.9 and 70.7 %, against 28.9 %.

(b) With the swing off, the demand steps by 10 kg/s halfway between two executions of the 1 s control period. Find the level lost before the trim acts, and the most that can be lost. *Checkpoint:* 0.001 m, and 0.002 m.

(c) Find the gain a feedforward from the supply drop would need at the design point. *Checkpoint:* 0.842 % per bar.

2.7 App challenge · 8 min

In Settings set Level transmitter noise and Flow transmitter noise to 0. Leave the pressure swing on at its default 12 ± 10 bar and the demand at 2 kg/s. For scenario 3, Feedback only, pressure swing: at speed $1\times$ press Reset, select Feedback and press PID auto within the first 5 s, then raise the speed to $10\times$, press Stop once the clock passes 900 s and download the Results file. Repeat for scenario 6, Cascade, pressure swing, selecting Cascade with + outlet FF off. Use the rows from 10 to 15 min of each file.

(a) Read the range, highest minus lowest, of `level_true_m` in each run, and the ratio of the two. *Checkpoint:* 0.0214 m and 0.0047 m, within 0.0005 m, a ratio near 4.6.

(b) Read the lowest and highest `valve_pct` and `inlet_flow_kgs` in each run. *Checkpoint:* Feedback 21.6 to 22.1 % and 3.06 to 10.33 kg/s. Cascade 15.1 to 42.0 % and 5.26 to 8.86 kg/s. Within 0.2 on each.

Compare the ratio of part (a) with the ratio of the sensitivities $|S_1|/|S_2|$ at the swing frequency.

2.8 App challenge · 8 min

In Settings set both transmitter noises to 0, and switch the pressure swing off. For scenario 2, Feedback only, demand steps: at speed $1\times$ press Reset, select Feedback and press PID auto within the first 5 s, then raise the speed to $10\times$, press Step +10 when the clock reads about 100 s and Step -10 at about 350 s, run to 470 s and download the Results file. Repeat with FF outlet (scenario 5, The right feedforward) and with Cascade, + outlet FF off (scenario 7, Cascade, demand steps). For each run find the largest $|1 - \texttt{level_true_m}|$. Accept 0.002 m on each.

2.9 App challenge · 8 min

Scenario 9, Check Kff exactly. In Settings set both transmitter noises to 0, and switch the pressure swing off. At speed $1\times$ press Reset, select FF outlet, set Kff to 1.44 and press PID auto within the first 5 s. Raise the speed to $10\times$, press Step +10 when the clock reads about 100 s, run to 350 s and download the Results file. Repeat with Kff at 2.887 and at 5.77.

(a) For each gain read the lowest and the highest `level_true_m` after the step. *Checkpoint:* 1.44: lowest 0.861 m. 2.887: within 0.002 m of 1 m. 5.77: highest 1.280 m. Within 0.003 m.

(b) Predict the two nonzero errors from the feedback error of 0.2775 m for the same step, taking the error to scale with the share of the load the trim leaves uncompensated. *Checkpoint:* 0.139 m below and 0.277 m above the setpoint.

2.10 Application area · 3 min

In the pressure-only run the cascade sweeps the valve between 14.80 and 42.17 % on every 62.8 s cycle of the swing (Section 2.6).

(a) Estimate the valve travel per day, counting two sweeps per cycle, in full strokes of 100 %. *Checkpoint:* About 750 full strokes per day.

(b) Name the two responses to this wear that the Limitations section gives, and say what the app cannot tell you about them.

2.11 Application area · 4 min

The model takes the leak as linear in level. Suppose the tank drained instead through an orifice, $F_{\text{leak}} = c\sqrt{L}$, with $c = 5\,\text{kg/s}$ per $\sqrt{\text{m}}$, so that the same 5 kg/s leaves at 1 m.

(a) Linearize the leak and find τ and K_p at 1 m and τ at 2 m, with $K_v = 0.3464\,\text{kg/s}$ per percent. *Checkpoint:* 2000 s and 0.139 m per percent at 1 m, and 2828 s at 2 m.

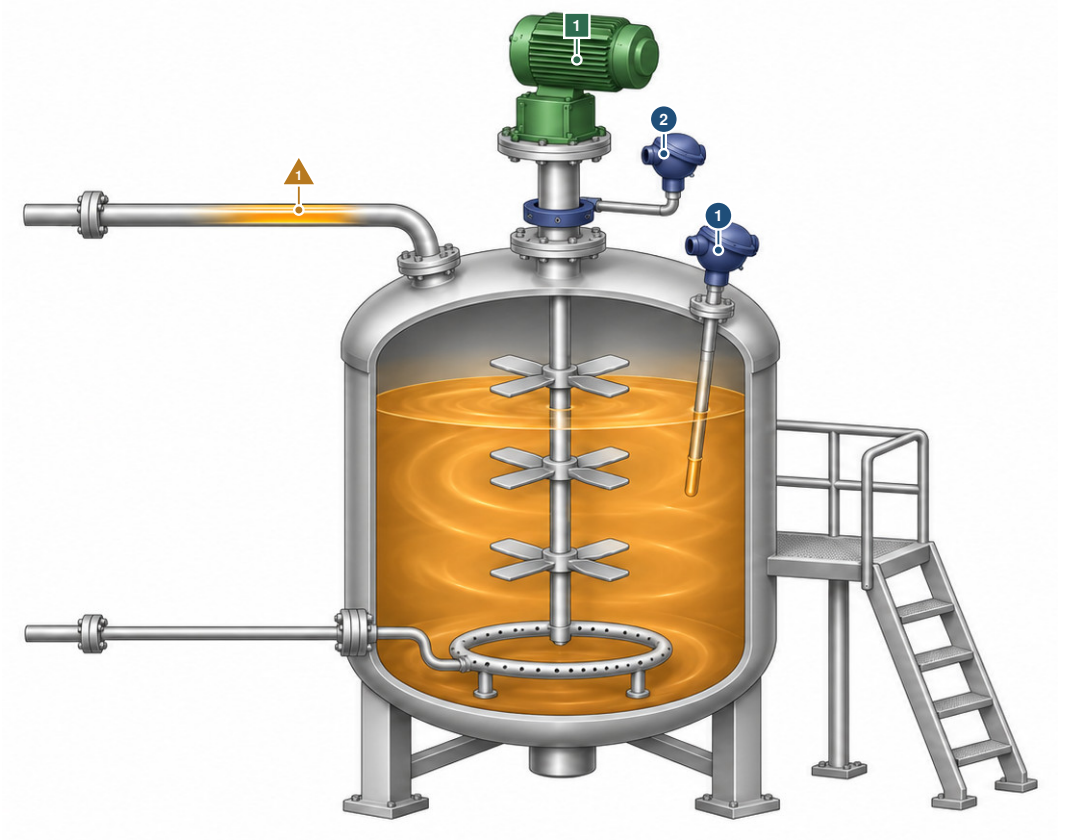
(b) Say whether the ideal outlet feedforward gain changes with the level.

2.12 Review the agent · 5 min

An AI assistant reviews the cascade and replies as follows. “In the 900 s combined scenario, raising both inner gains from 1.0 to 3.0 lowers the worst level error from about 0.29 to about 0.22 m. The tighter flow loop rejects the supply swing faster, so set $K_{c2} = K_{I2} = 3.0$.” Find the flaw in the recommendation and say how you would check it.

CHAPTER 3

Measuring the Inner Loop Speed a Fermenter Oxygen Cascade Requires



The controller moves 1 agitator motor and its drive, through the speed setpoint

The controller reads 1 dissolved oxygen probe in the broth 2 shaft speed sensor (tachometer)

What disturbs it 1 glucose feed rate

Notice the fast speed sensor on the shaft and the slow oxygen probe in the broth: the speed loop closes on the first, and no control structure can speed up the second. Illustration, not to scale.

IN PLAIN TERMS **An Inner Loop Must Be Fast to Help**

The main idea. Yeast in a big stirred tank uses oxygen all the time, and the liquid holds only a few seconds of it. One controller keeps the oxygen steady by changing how fast the stirrer turns, and an inner one makes the motor reach that speed.

The challenge. The motor, the liquid and the oxygen sensor all take several seconds to respond. A slow inner controller makes the pair worse than one controller alone, and nothing can speed up the sensor.

Why it matters. How fast the inner one must be is the main design question for any pair like this.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of the oxygen probe (Section 3.2), the fact on oxygen-starved zones (Section 3.3.6), the fact on outer gain and the try-it segments of scenarios 6 and 8 (Section 3.6) and the key points
Study	65 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	30 min	the first page, the results (Section 3.6), Technology in Practice (Section 3.7), Limitations (Section 3.8) and the open challenges

The control question The agitator drive, the dissolved oxygen in the broth and the membrane probe of a fermenter respond within a factor of two of one another. How fast must the speed loop of an oxygen cascade be before it helps rather than hurts, and which of the three lags can any control structure reach?

Learning objectives

- Compute the oxygen inventory, the liquid-phase time constant and the drive lag of the fermenter from its model.
- Predict the closed-loop time constant of the speed loop from its gain, and state what that estimate leaves out.
- Measure the inner-loop speed at which the cascade stops paying, and compare it with the three-to-five rule.
- Explain why a feedforward from the glucose feed must wait for the culture.
- Tell a capacity limit or a biased probe from a tuning problem by reading the speed margin and the true dissolved oxygen.

Prerequisites Sections 1.4.5 (inner-loop speed), 1.4.3 (feedforward timing), 1.4.2 (entry points) and 1.6.2 (velocity form). Chapter 2 for the two structures on a simpler plant.

The app https://apmonitor.com/apps/bioreactor-do/

See also Volume 19, Chapter 8 for one oxygen loop that sequences agitation, enrichment and gas flow, and Volume 19, Chapter 12 for an oxygen cascade with a meal feedforward. Volume 2, Chapter 2 for the velocity form and back-calculation, Volume 2, Chapter 4 for the same three architectures on steam, and Volume 2, Chapter 3 for an actuator bound that no tuning removes.

3.1 Introduction

Every treatment of cascade control carries the same design rule. The inner loop must be substantially faster than the outer one, usually quoted as three to five times, or the arrangement is not worth building [1, 2]. The rule is rarely tested, because the standard teaching example is a flow loop inside a temperature loop. A flow loop settles in a second and a jacketed vessel settles in ten minutes, so the margin is six hundred to one, and detuning the inner controller by a factor of ten changes nothing visible on the outer trend. The rule is therefore delivered as an assertion rather than as a measurement. Section 1.4.5 states it and measures it once on a heater whose inner loop wraps a valve lag, and this chapter measures it on a plant where the margin is narrow.

Dissolved oxygen in a stirred fermenter is the counterexample. The liquid holds so little oxygen that its own time constant is 12.4 s at the design condition of the plant described here, the agitator drive changes speed with a time constant of 15.4 s, and the membrane probe reports what happened 20 s ago. Those three numbers are within a factor of two of each other, and the speed loop a cascade inserts closes in 2.9 s. The margin is eleven to one against the whole outer chain rather than six hundred to one, the inner tuning is an adjustable input, and the rule can be broken inside twenty simulated minutes. The consequence is a loss rather than a smaller benefit. A cascade whose inner loop settles in 12.3 s answers a feed step with an integrated error of 72.3, against 23.8 for the single loop it replaced.

This chapter targets the misconception that a cascade is structurally safe, so that adding an inner loop can improve performance but cannot degrade it. The measurement in Section 3.6 says otherwise, and it also identifies what the rule of thumb omits. Holding the inner proportional gain fixed and

shortening the inner integral time from 1.20 min to 0.10 min moves the break-even point from an inner time constant of 7.4 s out to 13.4 s, so the deciding quantity is the inner loop's whole closed-loop response and not the proportional gain alone.

The class of problem is any outer loop whose final control element has dynamics and a nonlinearity of its own, and whose own process is fast enough that the element cannot simply be ignored. Variable-speed pumps on flow or pressure, fans on temperature, compressors on suction pressure and smart positioners on valve stems are all the same structure, a fast measured inner loop wrapped around an actuator, so that the outer loop addresses an element that reaches its commanded value. The benefit of the arrangement was quantified against the ratio of primary to secondary dynamics decades ago, and both the size of the benefit and the region where it disappears depend on that ratio rather than on the tuning [2, 3]. Dissolved oxygen belongs to a second class as well, the aeration loops whose actuator power is the largest electricity consumer on the site, where the same measurement lag and the same capacity limits recur in wastewater treatment as in fermentation [4]. Volume 9, Chapter 10 treats such a plant, where a supervisory loop chooses the dissolved-oxygen setpoint itself.

The physical fundamentals underneath the experiments are compact. One oxygen mass balance carries an interphase transfer rate proportional to a concentration driving force against a metabolic uptake rate, with a stored quantity of a few seconds of demand between them, so the loop inherits a time-scale separation between a batch that runs for hours and an inventory that turns over in seconds [1, 5]. The experiments built on that balance cover cascade structure, measurement lag, actuator dynamics and nonlinearity, feedforward timing, anti-windup and capacity constraints, the principles a process dynamics and control course assembles around its cascade lecture [1, 6]. The behavior matters industrially because aeration power dominates the electricity bill of fermentation and wastewater sites alike, through loops that carry the same probe lags and capacity limits [4, 7].

Chapter 2 asks which structure answers which disturbance, using a tank level and two upsets that enter by different routes. This chapter takes the structure as given. The ordinary way to study what the inner loop must be is a pair of transfer functions with an assumed ratio between them, and that ratio is the quantity under study. The fermenter of this chapter has an oxygen holdup, a drive inertia, a probe lag and an inner loop tuning that are each adjustable on their own, so the speed an inner loop must reach becomes something measured on one plant instead of something assumed.

Section 3.2 traces how oxygen transfer in a stirred vessel became a measured quantity and how the loop that holds it was built. Section 3.3 sets out the nine-state fermenter with its transfer correlation, its drive torque balance and its instrument chain. Section 3.4 sorts the three lags of the outer loop by which control structure can reach each one. Section 3.5 gives the two controllers and the carbon-balance feedforward, three architectures that share one output unit. Section 3.6 carries the two central measurements, that a detuned cascade is three times worse than no cascade on the same disturbance, and that the crossover between harm and benefit moves by a factor of 1.8 when the inner integral time alone is changed. The same section runs the scenarios in which a capacity limit, an instrument bias and a derivative term each fail in a way the control error does not report. Section 3.7 sets the loop on real hardware. Sections 3.8 and 3.9 state what the model cannot support and what the error metrics leave out, and Section 3.10 collects the numbers. The correlation, the kinetics and the cascade structure are all standard.

3.2 Background

Aerobic fermentation at scale began without a dissolved-oxygen measurement of any kind. Submerged penicillin culture in the 1940s made the supply of oxygen the limiting design question of an entire industry, and the quantity of engineering interest was a transfer capacity rather than a concentration. The sulfite oxidation method measured how much oxygen an agitated vessel could deliver by letting a fast chemical reaction consume everything that crossed the interface, and correlating the result against impeller power and superficial gas velocity gave the first design equations for agitated gas-liquid contactors [8]. That work fixed the form the industry still uses, a volumetric coefficient $k_L a$ written as

a power law in gassed power per volume and in gas velocity, and it fixed the practice of sizing an agitator from a correlation rather than from a measurement on the vessel being built.

The correlation was refined for thirty years and then audited. A review of the measuring methods and the published results for nonviscous systems showed that the exponents cluster tightly while the leading coefficient does not, and separated coalescing from non-coalescing liquids as the largest single source of the spread [9]. The exponents used in this simulator, 0.70 on power per volume and 0.40 on gas velocity, come from that body of work, and so does the warning attached to them. The leading coefficient for a non-coalescing broth spans more than a factor of two across authors and vessels, so a scale-up that sizes a motor from a bench correlation is choosing the plant's oxygen capacity, its installed power and the loop's process gain in one calculation whose least trustworthy element is that coefficient [10].

Measurement of the dissolved oxygen itself came later, from medical instrumentation. The membrane-covered polarographic electrode reported in 1953 put a platinum cathode behind a permeable membrane so that the current was proportional to oxygen tension rather than to whatever else was in the solution [11]. Fermentation could not use it as built, because a fermenter is steam sterilized and a laboratory electrode is not, and the adaptation took another decade. Steam sterilizable probes with replaceable membranes and internal electrolyte reservoirs made continuous dissolved-oxygen measurement routine on production vessels [12]. The membrane that made the measurement possible is also what makes it slow, because the oxygen has to diffuse through it before the cathode sees any change, and a response time measured in tens of seconds has been a property of the instrument since its invention.

> ⧖ HISTORY **From Clark Cell to Optical Probe, Around 2005**
>
> Dissolved oxygen in bioreactors has traditionally been measured with Clark-type electrodes. Oxygen is reduced at a cathode with four electrons exchanged for each molecule, so the current is linear in the concentration [13]. A federal field manual for water-quality data used such amperometric sensors as its primary method until luminescence-based optical sensors became more common, around 2005 [14]. Its 2020 revision states that optical sensors have replaced the amperometric type and makes them the standard field method, while the older sensor remains acceptable [15]. Stirred fermenters changed more slowly. A review of 2013 found Clark-type electrodes still dominating in conventional larger-scale systems, the kind of probe this chapter models [13].

Closed-loop control followed once the measurement existed. Early practice fixed an airflow and an agitator speed from the batch recipe and accepted whatever the culture did to the reading, which is the manual operation scenarios 1 and 2 of this application reproduce. Automatic control through agitation speed was reported for a fermenter in the mid 1980s and is now the standard arrangement, because speed enters the transfer correlation through the power term with an exponent about twice that of the gas velocity term and is therefore the stronger actuator [16]. The two remaining levers are vessel back pressure, which raises the equilibrium concentration through Henry's law at negligible compressor duty, and airflow, which raises the gas velocity term while making foam. The industrial split that survives is agitation and pressure as the controlled actuators with airflow on a recipe schedule.

The fed-batch itself has a separate lineage that matters for the results in this chapter. Feeding a concentrated sugar to a substrate-limited culture rather than charging it at the start became standard once the respiratory capacity of yeast was understood as the quantity being protected, with overfeeding rather than oxygen starvation as the immediate failure mode [17]. In such a culture the feed pump sets the growth rate and dissolved oxygen limits nothing until it falls very low, which is why the biased probe of scenario 11 costs no product at all and still constitutes a serious fault. Product formation is written in the growth-associated and non-growth-associated form introduced for lactic acid fermentation, which closes the carbon balance rather than asserting a yield [18]. Fuel ethanol is the product because it is a commodity manufactured at very large scale with margins thin enough that the net energy balance of the process has been debated in the open literature [19], and because aeration

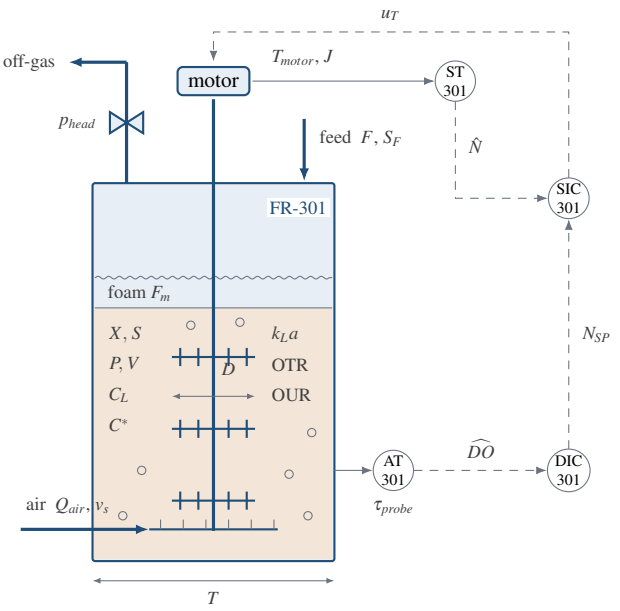


Figure 3.1 Fermenter FR-301 with its agitator drive, sparger, dissolved-oxygen probe and the two controllers of the cascade. Broth of volume V holds cells X, glucose S and ethanol P with dissolved oxygen C_L against an equilibrium concentration C^* set by the headspace pressure p_{head}. Three impellers of diameter D on a shaft of inertia J turn at speed N against motor torque T_{motor}, setting the transfer coefficient k_La and therefore the transfer rate OTR against the culture's uptake rate OUR. Glucose feed F at concentration S_F is the principal disturbance. AT-301 reports the dissolved oxygen $\widehat{DO}$ through a lag τ_{probe} to DIC-301, whose output is a speed demand N_{SP} for SIC-301, which reads tachometer ST-301 and writes the drive torque command u_T.

is by a wide margin the largest mechanical energy demand of a stirred production fermenter [7]. The agitator in this simulator draws 88 kW at the design condition and 281 kW at its speed limit on the same charge, and no control specification anywhere in the application mentions either number.

3.3 Process Model

Vessel FR-301 is a 4.0 m diameter stirred fermenter of 105 m^3 maximum working volume, fitted with three six-blade Rushton turbines of 1.40 m diameter on one shaft and a ring sparger below the bottom impeller. The design charge, from which nine of the thirteen scenarios start, is a fed-batch about half way through, 65 m^3 of broth carrying 15 g/L of cells and 40 g/L of ethanol at 32 °C, with a 550 g/L glucose syrup metered in at 24 L/min, air at 0.40 vvm and a headspace held at 0.5 barg. Dissolved oxygen is measured by a membrane probe and reported as a percentage of air saturation at one atmosphere, which is what a probe calibrated in air before the batch actually reads. Every symbol used below appears on Figure 3.1.

Table 3.1 separates what the loop holds from what it moves and from what moves underneath it. Time is in minutes throughout, because one time base must serve both a twelve-second oxygen transient and a fifteen-hour batch. Concentrations are g/L for the culture and mg/L for dissolved oxygen, volumes are m^3, agitator speed is integrated in rev/s and displayed in rev/min, torque is N m and pressure is Pa. Four annunciations bound normal operation, dissolved oxygen below 15 %, foam filling the headspace, the agitator at its speed limit, and the broth reaching the 105 m^3 working volume, after which the vessel overflows at the feed rate and the dilution terms carry cells and ethanol

Table 3.1 Variables of the Dissolved-Oxygen Loop. The Nominal Column Is the Design Charge That Opens Nine of the Thirteen Scenarios.

Symbol	Role	Nominal	Range
DO (AT-301)	CV, measured	25.0 %	5–90 % SP
N_{SP}	MV	64.1 rev/min	21–93 rev/min
$\hat{N}$ (ST-301)	inner CV	64.1 rev/min	0–93 rev/min
u_T	inner MV	43.5 %	0–100 %
F feed	DV, measured	24 L/min	0–90 L/min
Q_{air}	operator set	0.40 vvm	0.05–1.2 vvm
p_{head}	operator set	0.5 barg	0–1.5 barg
probe bias	DV, unmeasured	0 %	±20 %
probe noise	DV, unmeasured	0 %	0–4 % 1σ
X	state	15 g/L	≥ 0
S	state	0.047 g/L	≥ 0
P	state	40 g/L	≥ 0
V	state	65 m^3	≤ 105 m^3
C_L	state	1.65 mg/L	0–C^*
N	state	1.068 rev/s	0–1.55 rev/s
C_m	state	1.65 mg/L	≥ 0
F_m foam	state	0.22	0–1
A antifoam	state	0	≥ 0

out with the overflow. None of the four stops the run. Each latches and is recorded, so an excursion outside normal operation leaves a record. Nine states are integrated.

3.3.1 Oxygen Balance

The liquid oxygen balance is Equation (3.1), and every other equation in the model exists to supply one of its three terms.

$$\frac{\mathrm{d}C_L}{\mathrm{d}t} = k_L a\,(C^* - C_L) \;-\; \mathrm{OUR} \;-\; \frac{F}{V}\,C_L \tag{3.1}$$

where C_L is the dissolved oxygen concentration in mg/L, $k_L a$ is the volumetric transfer coefficient in min^{-1}, C^* is the equilibrium concentration at the working pressure in mg/L, OUR is the oxygen uptake rate in mg/(L·min), F is the glucose feed in m^3/min and V is the broth volume in m^3. Every term carries units of mg/(L·min). The dilution term is three orders of magnitude below the other two at every condition the application reaches and is kept only so that the balance closes exactly.

Equation (3.1) has the form of a first-order lag, although the size of its two large terms relative to the stored quantity produces different behavior. At the design charge the 25 % setpoint is 1.65 mg/L, which over 65 m^3 is 107 g of dissolved oxygen in the whole vessel. The culture consumes 39.6 mg/(L·min), which over the same volume is 2.57 kg every minute. The entire dissolved inventory is therefore consumed and replaced every 2.5 s. A tank holds minutes of throughput and a jacketed reactor holds minutes of thermal energy, while this vessel holds two and a half seconds of oxygen, so a one percent mismatch between the two rates moves the dissolved oxygen by one percent of its value every two and a half seconds. Dissolved oxygen therefore collapses rather than drifts when control is lost.

3.3.2 Oxygen Transfer and the Agitator

The transfer coefficient is written in the literature power-law form rather than directly in agitator speed, as shown in Equation (3.2).

$$k_L a = a\left(\frac{P_g}{V}\right)^{\alpha} v_s^{\beta}, \qquad \frac{P_g}{V} = \frac{2\pi f_g k_T N^3}{V} \tag{3.2}$$

where P_g is gassed shaft power in W, v_s is the superficial gas velocity in m/s, a, α and β are the correlation constants, f_g is the gassed-to-ungassed power ratio and $k_T = N_P \rho D^5 n_{imp}/2\pi$ is the ungassed load-torque coefficient in N m per $(\text{rev/s})^2$, built from the impeller power number N_P, the broth density ρ, the impeller diameter D and the number of impellers n_{imp}. Writing the correlation this way rather than as a fit in speed is deliberate, because airflow, working volume and gassing then act through the same equation and a change in any of them moves the loop gain with no control setting having been touched. Since gassed power per volume goes as N^3 at a fixed aeration condition, Equation (3.2) gives $k_L a \propto N^{3\alpha} = N^{2.1}$, and 2.1 is the process gain exponent of the outer loop. At the design point 64.1 rev/min delivers 1350 W/m^3 at a superficial gas velocity of 0.0345 m/s for a $k_L a$ of 4.84 min^{-1}, which by Equation (3.1) is a liquid-phase time constant $1/k_L a$ of 12.4 s.

The superficial gas velocity and the gassed power ratio are given by Equation (3.3).

$$v_s = \frac{Q_{air} V}{60 A_t}, \qquad f_g = \frac{1}{1 + k_g \text{Fl}}, \qquad \text{Fl} = \frac{Q_{air} V}{60 N D^3} \tag{3.3}$$

where Q_{air} is the airflow in vessel volumes per minute, $A_t = \pi T^2/4$ is the vessel cross-section in m^2, Fl is the aeration number and k_g is the gassing coefficient. Sparged gas forms cavities behind the blades and unloads the impellers, so the same speed costs less torque when the air is on. The reciprocal form is used rather than the customary linear fit in Fl because a linear fit returns a negative power ratio at the aeration numbers a 65 m^3 fermenter actually runs at. At the design condition $\text{Fl} = 0.148$ and $f_g = 0.849$.

Two multipliers ride on top of Equation (3.2). Antifoam divides the coefficient by $1 + k_{AF} A$, and a foam blanket multiplies it by $1 - 0.4\,\text{clamp}((F_m - 0.8)/0.4, 0, 1)$, which removes up to 40 % once the headspace is more than four fifths full and is ramped in over the last fifth so that the right-hand side stays smooth for the integrator. A fouled sparger multiplies it by a further 0.75.

The equilibrium concentration and the reading are Equation (3.4).

$$C^* = C_{ref}(T_b)\,\frac{p_{atm} + p_{head}}{p_{atm}}, \qquad DO\,\% = \frac{100\,C_L}{C_{ref}(T_b)} \tag{3.4}$$

where C_{ref} is the oxygen solubility of the broth under one atmosphere of air, taken as a cubic fit in broth temperature for clean water multiplied by a salting-out factor of 0.92. At 32 °C that gives $C_{ref} = 6.59$ mg/L. The reading of Equation (3.4) divides by C_{ref} and not by C^*, because the probe was calibrated in air at one atmosphere before the batch and nothing afterward changes the calibration. That convention is why back pressure raises the driving force at low compressor cost and why the reading can exceed 100 %. Taking the head from 0.5 barg to 1.5 barg moves C^* from 9.84 mg/L to 16.34 mg/L and multiplies the driving force in Equation (3.1) by about 1.8.

3.3.3 Culture Kinetics and Fed-Batch Balances

Growth is Monod on glucose with non-competitive inhibition by the ethanol the culture is making and a Monod term in dissolved oxygen, as shown in Equation (3.5).

$$\mu = \mu_{max}\,\frac{S}{K_S + S}\,\frac{K_P}{K_P + P}\,\frac{C_L}{K_O + C_L} \tag{3.5}$$

where μ is the specific growth rate in min^{-1}, K_S is the glucose half-saturation, K_P is the ethanol inhibition constant and K_O is the oxygen half-saturation. The oxygen term couples the dissolved-oxygen loop to the batch outcome, and its half-saturation of 0.35 mg/L is 5.3 % of air saturation, well below the 15 % the plant alarms at. The rates that follow from μ are Equations (3.6) and (3.7).

$$\text{OUR} = 1000 X \left(\frac{\mu}{Y_{X/O}} + m_O \right) \tag{3.6}$$

$$q_P = \alpha_{LP}\mu + \beta_{LP}, \qquad q_S = \frac{\mu}{Y_{X/S}} + m_S + \frac{q_P}{Y_{P/S}} \tag{3.7}$$

where X is the biomass in g/L, $Y_{X/O}$ is the cell yield on oxygen, m_O is the oxygen maintenance coefficient, q_P and q_S are the specific product and substrate rates in g/g/min, α_{LP} and β_{LP} are the growth-associated and non-growth-associated product coefficients [18], $Y_{X/S}$ and $Y_{P/S}$ are the cell and product yields on glucose and m_S is the glucose maintenance coefficient. The factor of 1000 in (3.6) carries g/L of cells into mg/L of oxygen. The maintenance term is the part of the demand that does not stop when growth does, which is why a stalled culture still needs the agitator. Writing q_S with three separate sinks in (3.7) closes the carbon balance, so substrate the culture cannot consume appears as rising residual glucose rather than disappearing from the balance.

The four culture states are Equation (3.8).

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= \mu X - \frac{F}{V}X, & \frac{\mathrm{d}S}{\mathrm{d}t} &= \frac{F}{V}(S_F - S) - q_S X \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= q_P X - \frac{F}{V}P, & \frac{\mathrm{d}V}{\mathrm{d}t} &= F \end{aligned} \tag{3.8}$$

where S_F is the glucose concentration in the feed. The disturbance this loop faces for the whole batch is in the first of those four. Oxygen demand is proportional to cell mass through Equation (3.6) and cell mass only ever grows, so the load on the loop is a ramp that runs from inoculum to harvest, 18.3 mg/(L·min) at 5 g/L of cells, 39.6 at 15 and 65.8 at 34.

3.3.4 Agitator Drive, Probe and Foam

The shaft speed follows from a torque balance rather than from a command that is directly obtained, as shown in Equation (3.9).

$$2\pi J \frac{\mathrm{d}N}{\mathrm{d}t} = T_{motor} - f_g k_T N^2, \qquad \tau_{drive} = \frac{2\pi J}{2 f_g k_T N} \tag{3.9}$$

where J is the effective drive inertia in kg m^2, lumping the rotating mass, the entrained broth and the drive's own current limit, and N is in rev/s. Linearizing the quadratic load about the running speed gives the drive time constant on the right, which is inversely proportional to speed. The drive is slowest when it is turning slowly, 23.5 s at the 43 rev/min of an early batch, 15.4 s at the design 64 rev/min and 11.6 s at the 86 rev/min of a late one. A fixed-gain outer controller cannot track that variation, while the speed loop removes it. Setting the derivative in (3.9) to zero whenever the speed rests on a stop with the torque imbalance pushing further outward keeps the fixed-step integrator from crossing a limit and coming back, and a clip applied after each completed step catches anything left over.

The probe is a first-order lag on the true concentration and is a state of the model rather than a filter in the controller, as shown in Equation (3.10).

$$\tau_{probe} \frac{\mathrm{d}C_m}{\mathrm{d}t} = C_L - C_m, \qquad \widehat{DO} = \frac{100\,C_m}{C_{ref}} + b + n(t) \tag{3.10}$$

where C_m is the probe state in mg/L, b is a calibration bias in percentage points and n is electrical noise added after the membrane. The noise is a Gaussian sequence taken through a first-order filter of 0.05 min (3 s) and scaled so that the standard deviation of the stationary sequence equals the set noise level. The tachometer is a separate first-order filter of time constant τ_N on the true shaft speed, with no bias and no noise.

Foam is one lumped height as a fraction of the available headspace, generated by gas velocity, power input and the protein a growing culture releases, as shown in Equation (3.11).

$$\begin{aligned} \frac{\mathrm{d}F_m}{\mathrm{d}t} &= G_F - \frac{F_m}{\tau_F} \\ G_F &= \frac{k_F}{1 + k_{AF,f} A} \left(\frac{v_s}{v_{s,ref}} \right)^{1.5} \frac{X}{K_{fx} + X} \sqrt{\frac{P_g/V}{(P/V)_{ref}}} \end{aligned} \tag{3.11}$$

where k_F is the generation rate constant, $v_{s,ref}$ and $(P/V)_{ref}$ are reference conditions, K_{fx} is the biomass half-saturation for foaming, $k_{AF,f}$ is the antifoam suppression strength and τ_F is the collapse time. Antifoam is a first-order washout, $\mathrm{d}A/\mathrm{d}t = -A/\tau_{AF}$, dosed as a unit step on the state, and each dose also knocks the foam down to at most 0.15 at once. The foam state is clamped to the interval $[0,1]$ after every step, and reaching 1 annunciates a foam-out. Equation (3.11) gives the airflow actuator a consequence that its own range does not show, which is the subject of scenario 10.

3.3.5 Constants, Initialization and Numerics

Table 3.2 gives every constant the model uses, and all of them are displayed and adjustable. The task here is not to recover a hidden constant but to compare four time constants the application reports, so full disclosure of the constants is part of the design.

Every scenario opens at a steady state of the oxygen subsystem. The initializer solves a coupled fixed point, because the feed sets the growth rate and the growth rate sets the oxygen demand while the oxygen in turn limits the growth. The residual glucose of a substrate-limited fed-batch is the root of $(F/V)(S_F - S) = q_S(\mu(S))X$, which is unique because the supply falls and the demand rises with S, and the dissolved oxygen is the root of Equation (3.1) set to zero, which is unique because transfer falls and uptake rises with C_L. Both are found by bisection to machine precision in sixty steps. Six passes alternate between them, after which the transfer correlation of Equation (3.2) is inverted for the agitator speed that balances the demand at the dissolved-oxygen setpoint, and that speed is clamped to $[N_{min}, N_{max}]$. The final dissolved oxygen is then re-solved at whichever speed survived the clamp, so a scenario whose agitator has run out opens at the resulting lower value rather than at the setpoint. The correlation inversion is itself a fixed point, since the gassed power ratio depends on the speed being solved for, and it converges in three or four passes because f_g is a mild function of speed. The initial torque is set to the load torque at the initial speed, so the shaft starts balanced. The first verification test the application runs on itself checks that the derivatives of C_L, N, C_m, F_m and A are all below 10^{-9} at the opening state.

States advance by classical fourth-order Runge-Kutta on a fixed step of 0.002 min (0.12 s), which places about a hundred steps across the 12.4 s oxygen lag. DIC-301 and SIC-301 share an execution period declared as 0.006 min. The engine runs a controller at the first step at or after each period, so both execute every third step, 0.006 min (0.36 s). The historian is declared at $1/0.018$ samples per simulated minute and for the same reason records every ninth step, 0.018 min (1.08 s). Noise is drawn from a seeded generator with seed 20260829, so scenario 12 repeats to the last digit. The engine and its three clocks are those of Section 1.6.1.

3.3.6 Assumptions

The model is detailed on the oxygen path and simplified elsewhere, and each simplification has a domain where it fails.

The broth is one perfectly mixed lump. A 100 m^3 vessel with three impellers carries a real dissolved-oxygen gradient from the sparger to the surface, and cells circulating through such a vessel experience substrate and oxygen fluctuations on a scale of seconds [20]. The probe sits at one height in that gradient. This is the largest single omission in the application, and it is one reason a plant runs a higher setpoint than the kinetics alone would justify.

> **+ DID YOU KNOW? Oxygen-Starved Zones at 22 Cubic Meters**
>
> Fed-batch cultures of a bacterium at 22 m^3 accumulated formate, a sign of zones short of oxygen. The study, published in 2001, suggests that moving zones of oxygen limitation explain part of the lower biomass yield at large scale [20]. One probe at one height sees none of them.

The gas phase carries no state. Oxygen leaves the sparger at its inlet composition and Equation (3.4) applies that composition everywhere, so the driving force is overstated near the top of a tall

Table 3.2 Constants of the Fermenter Model in the Units the Application Reports. Every Constant Is Displayed and Adjustable.

Symbol	Meaning	Value	Unit
T	vessel inside diameter	4.0	m
T_b	broth temperature	32	°C
—	broth solubility factor	0.92	—
V_{full}	maximum working volume	105	m^3
N_P	impeller power number	5.0	—
D	impeller diameter	1.40	m
n_{imp}	number of impellers	3	—
ρ	broth density	1050	kg/m^3
N_{min}	minimum agitator speed	21	rev/min
N_{max}	maximum agitator speed	93	rev/min
T_{rated}	rated motor torque	30	kN m
J	effective drive inertia	60000	$kg\,m^2$
$f_{g,nom}$	drive schedule calibration	0.78	—
k_g	gassed-power coefficient	1.2	—
a	transfer coefficient	2.0×10^{-3}	SI, s^{-1} basis
α	power exponent	0.70	—
β	gas velocity exponent	0.40	—
μ_{max}	maximum growth rate	0.55	h^{-1}
K_S	glucose half-saturation	0.05	g/L
K_P	ethanol inhibition	75	g/L
K_O	oxygen half-saturation	0.35	mg/L
$Y_{X/S}$	cell yield on glucose	0.45	g/g
$Y_{X/O}$	cell yield on oxygen	1.00	g/g
$Y_{P/S}$	ethanol yield on glucose	0.48	g/g
m_S	glucose maintenance	0.030	g/g/h
m_O	oxygen maintenance	0.015	g/g/h
α_{LP}	growth-associated ethanol	1.2	g/g
β_{LP}	non-growth ethanol	0.05	g/g/h
S_F	glucose in the feed	550	g/L
k_F	foam generation rate	0.024	min^{-1}
τ_F	foam collapse time	20	min
K_{fx}	foaming half-saturation	15	g/L
$v_{s,ref}$	reference gas velocity	0.035	m/s
$(P/V)_{ref}$	reference power input	1500	W/m^3
τ_{AF}	antifoam residence time	60	min
k_{AF}	antifoam k_La penalty	0.30	—
$k_{AF,f}$	antifoam suppression	4.0	—
τ_{probe}	probe time constant	20	s
τ_N	tachometer filter	0.5	s
—	critical dissolved oxygen	15	%
—	high foam alarm	0.75	—
X_0	biomass at start	15	g/L
P_0	ethanol at start	40	g/L
V_0	broth volume at start	65	m^3

vessel where the gas has already been stripped. A real plant closes the same balance on an off-gas analyzer, which is also the measurement a better feedforward than the one implemented here would use.

The transfer correlation has fixed exponents and no upper limit. Equation (3.2) says nothing about broth viscosity, which climbs through a fermentation and can halve the coefficient by the end, and nothing about impeller flooding, where the gas rate outruns the impeller's ability to disperse it and

k_La collapses instead of rising [9, 10]. Both set real limits that the model draws as a smooth curve.

The drive is a first-order torque balance. The 15 s response of Equation (3.9) is right for a machine of this size, although a real variable-frequency drive reaches speed under a ramp limit rather than an exponential, so a large speed change takes longer than the model says and a small one arrives sooner. The cascade argument survives that change while the exact crossover in Section 3.6 would move.

The kinetics are unstructured with constant yields. Real Crabtree-positive yeast switches metabolism with both glucose and oxygen [17], so neither the ethanol yield nor the oxygen demand per gram of sugar is truly constant. The values in Table 3.2 are a fair average over the middle of a batch and are not a prediction of a lag phase, a diauxic shift or a stationary phase. The oxygen demand modeled here is that of an aerated propagation-stage vessel rather than that of the anaerobic main fermenter of a fuel ethanol plant.

Broth temperature is a parameter. There is no temperature loop, although aerobic metabolism releases roughly 460 kJ of heat per mole of oxygen consumed, about 14 MJ per kilogram, so oxygen transfer and cooling duty are the same constraint seen twice and on many plants the cooling runs out first.

The probe is a pure lag with no dead time. A real membrane probe responds more like two lags in series with a small transport delay, and a loop tolerates a lag better than a delay of the same size, so the probe penalties measured in Section 3.6 are optimistic.

3.4 Dynamic Analysis

3.4.1 Lag Structure of the Outer Loop

Between the dissolved-oxygen controller's output and its input sit three first-order lags, and at the design condition they are within a factor of two of one another. The drive of Equation (3.9) contributes 15.4 s, computed from $2\pi J/(2f_g k_T N)$ with $k_T = 13\,481\,\mathrm{N\,m}$ per $(\mathrm{rev/s})^2$ and $f_g = 0.849$ at $N = 1.068\,\mathrm{rev/s}$. The liquid phase of Equation (3.1) contributes $1/k_La = 12.4\,\mathrm{s}$. The probe of Equation (3.10) contributes 20 s, and 60 s once the membrane has been in the vessel for six batches. The disturbance is not instant either, because a feed change reaches the pump immediately and reaches the cells through a glucose pool that takes about a minute to fill. On the cascade run of Section 3.6 the oxygen demand reaches 63 % of its change 1.3 min after the feed moves and 90 % of it at 4.3 min.

The three lags are not equally reachable by a control structure, and the comparison in Section 3.6 follows from that inequality. The drive lies between the controller output and the shaft, so a loop closed on a shaft measurement encloses it. The oxygen holdup lies between the shaft and the true dissolved oxygen, so no loop encloses it, because it is the process itself. The probe lies on the far side of everything, so no rearrangement of controllers touches it. A cascade reaches the actuator, a feedforward reaches the disturbance, and nothing reaches the measurement.

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

3.4.2 Closed-Loop Response of the Inner Loop

The inner loop closes torque onto speed. Linearizing the load torque in (3.9) gives an inner process gain from percent of rated torque to speed, and proportional feedback divides the drive lag by the loop gain, as shown in Equation (3.12).

$$K_{p,N} = \frac{60\,T_{rated}/100}{2f_g k_T N}, \qquad \tau_{inner} = \frac{\tau_{drive}}{1 + K_{c,N}K_{p,N}} \tag{3.12}$$

where $K_{p,N}$ is in rev/min per percent of rated torque and $K_{c,N}$ is the inner proportional gain. The second expression is the proportional-only form of Equation (1.7), with the drive lag in place of the valve lag and $K_{p,N}$ in place of the valve gain. At the design condition $K_{p,N} = 0.736\,\mathrm{rev/min}$ per percent, so the shipped inner gain of 6 gives $\tau_{inner} = 2.85\,\mathrm{s}$ against a drive lag of 15.4 s and an

oxygen holdup of 12.4 s. Equation (3.12) is the number the application reports and it is deliberately incomplete, because it counts the inner proportional gain and not the inner integral time. Section 3.6 measures what that omission costs, and the answer is a factor of 1.8 on the very threshold the equation exists to locate.

The inner loop removes three separate effects. It removes the 15.4 s of drive lag. It removes the nonlinearity, because torque and speed are related by a square root in Equation (3.9), so the open-loop drive gain at 90 rev/min is only 39 % of what it is at 40 rev/min at the design volume and airflow (the square root alone gives 44 %, and the gassed power ratio rising with speed gives the rest), and a fixed outer gain has to survive both. And it removes every disturbance that enters at the shaft, which includes gassing, broth viscosity, a fouling impeller and a dragging bearing.

3.4.3 Steady-State Error of the Open-Loop Drive

The single-loop architecture is not a cascade with the inner controller removed but a drive running an open-loop torque schedule, as shown in Equation (3.13).

$$T_{motor} = f_{g,nom}\, k_T\, N_{dem}^2 \tag{3.13}$$

where $f_{g,nom}$ is the gassed power ratio the drive was commissioned at. Equation (3.13) represents a variable-frequency drive running without an encoder, commissioned at one gassing condition and retaining that calibration indefinitely. Setting (3.13) equal to the true load torque $f_g k_T N^2$ gives the steady-state error directly, since the shaft settles at $N = N_{dem}\sqrt{f_{g,nom}/f_g}$ with f_g evaluated at the settled speed. At the design airflow the commissioned 0.78 stands against a gassed power ratio of 0.855 at the settled speed, and a demand of 70.0 rev/min delivers 66.88 rev/min, a ratio of 0.955. Raising the airflow from 0.40 to 0.95 vvm at the same demand unloads the impellers further and the shaft runs up to 72.45 rev/min, an 8.3 % move in the manipulated variable caused by a disturbance the dissolved-oxygen controller was never told about. Both tests return 70.00 rev/min in cascade.

3.5 Controller Design

3.5.1 Velocity Form PID Algorithm

Both controllers are the velocity-form algorithm of Section 1.6.2, with the derivative on the measurement. The outer dissolved-oxygen controller DIC-301 is Equations (3.14) and (3.15).

$$\begin{aligned} e_k &= DO_{\mathrm{SP},k} - \widehat{DO}_k \\ \Delta N_k^P &= K_c\,(e_k - e_{k-1}) \\ \Delta N_k^I &= \frac{K_c}{\tau_I}\, e_k\, \Delta t_c \end{aligned} \tag{3.14}$$

$$\Delta N_k^D = -\,K_c \tau_D\, \frac{\widehat{DO}_k^f - 2\widehat{DO}_{k-1}^f + \widehat{DO}_{k-2}^f}{\Delta t_c} \tag{3.15}$$

where e_k is the error in percentage points of dissolved oxygen, K_c is the outer proportional gain in rev/min per percent, and τ_I and τ_D are the integral and derivative times in minutes. Both loops act direct, because more agitator speed and more motor torque each raise their own controlled variable. The derivative term reads a second difference of a measurement that has already been filtered, with $\widehat{DO}_k^f = \gamma\widehat{DO}_{k-1}^f + (1-\gamma)\widehat{DO}_k$ and $\gamma = \exp(-\Delta t_c/\tau_f)$, the filter constant τ_f defaulting to 0.05 min. The three increments are combined and the demand is clipped to the speed limits, as shown in

Equation (3.16).

$$\begin{aligned} s_k &= s_{k-1} + \Delta N_k^P + \Delta N_k^D \\ c_k &= N_{ff,k} + s_k + I_{k-1} \\ N_{SP,k} &= \min(N_{max},\ \max(N_{min},\ c_k)) \\ I_k &= I_{k-1} + \Delta N_k^I + \frac{\Delta t_c}{\tau_t}\left(N_{SP,k} - c_k\right) \end{aligned} \tag{3.16}$$

where Δt_c is the realized 0.006 min controller execution period, $N_{ff,k}$ is the feedforward speed of Section 3.5.3, $[N_{min}, N_{max}] = [21, 93]$ rev/min and $\tau_t = \tau_I$ is the tracking time of the back-calculation. In the notation of Equations (1.10) to (1.12), s_k is the running sum S_k, c_k is the output before limiting v_k, $N_{ff,k}$ is the bias u_b and τ_t is T_t. The back-calculation corrects the integral alone and never truncates the running sum [6]. At the shipped $\tau_I = 0.6$ min each execution removes $1/100$ of the clipped excess, so an integral wound against a speed limit unwinds over about τ_I rather than in one execution. Nothing rate-limits the output, and its only bound is that clamp. Manual mode and bumpless transfer work as Section 1.6.2 describes. The manual tracking also runs on whichever controller is idle, which is what makes a change of architecture mid-run bumpless as well.

The inner speed controller SIC-301 is proportional-integral with no derivative term, as shown in Equation (3.17).

$$\begin{aligned} e_{N,k} &= N_{SP,k} - \hat{N}_k, \qquad \Delta u_{N,k}^I = \frac{K_{c,N}}{\tau_{I,N}}\, e_{N,k}\, \Delta t_c \\ s_{N,k} &= s_{N,k-1} + K_{c,N}\left(e_{N,k} - e_{N,k-1}\right) \\ c_{N,k} &= s_{N,k} + I_{N,k-1} \\ u_{T,k} &= \min\bigl(100,\ \max\bigl(0,\ c_{N,k}\bigr)\bigr) \\ I_{N,k} &= I_{N,k-1} + \Delta u_{N,k}^I + \frac{\Delta t_c}{\tau_{I,N}}\left(u_{T,k} - c_{N,k}\right) \end{aligned} \tag{3.17}$$

with $T_{motor} = (u_T/100)\, T_{rated}$, where $\hat{N}$ is the filtered tachometer reading and u_T is clamped to 0–100 % of rated torque with the same back-calculation of (3.16), here with the tracking time $\tau_{I,N}$. Both loops execute on the same 0.006 min period. Tuning is bounded to $0 \le K_c \le 8$ rev/min per percent, $0.1 \text{ min} \le \tau_I \le 20$ min, $0 \le \tau_D \le 1$ min, $0 \le K_{c,N} \le 25$ percent of torque per rev/min and $0.01 \text{ min} \le \tau_{I,N} \le 3$ min.

3.5.2 A Common Output Unit Across Architectures

The outer controller's output means an agitator speed demand in rev/min in all three architectures, which is what makes the comparison in Section 3.6 fair. In the single loop that demand goes to Equation (3.13), in the cascade it is the setpoint of Equation (3.17), and in the feedforward architecture the same cascade is retained while N_{ff} in Equation (3.16) becomes nonzero. One set of tuning constants therefore carries across the whole ladder. The choice is not free, because in many cascades the outer output changes units when the structure changes and the gains have to change with it. The steam-heated exchanger of Volume 2, Chapter 4 offers the same ladder, a single loop, a cascade to steam flow and a heat-balance feedforward with a feedback trim.

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

3.5.3 Feedforward From a Carbon Balance

The feedforward is calculated rather than fitted. At quasi steady state every gram of glucose the pump delivers is consumed, the share of it that becomes new cells follows from Equation (3.7) with the maintenance terms dropped, and that new cell mass has to be paid for in oxygen at $Y_{X/O}$, as shown in

Equation (3.18).

$$\mathrm{OUR}_{ff} = \frac{1000\,K_{ff}}{Y_{X/O}} \cdot \frac{F S_F / V}{1/Y_{X/S} + \alpha_{LP}/Y_{P/S}}$$
$$k_L a_{req} = \frac{\mathrm{OUR}_{ff}}{C^* - C_{L,\mathrm{SP}}} \tag{3.18}$$

where K_{ff} is an operator confidence factor defaulting to 1 and $C_{L,\mathrm{SP}}$ is the setpoint expressed as a concentration. Equation (3.2) is then inverted for the speed that delivers $k_L a_{req}$, and that speed is passed through a first-order lag of time constant τ_{ff} and enters (3.16) as N_{ff}. Entering as the controller bias rather than as a summed output means saturation, anti-windup and bumpless transfer all account for the term with no extra arithmetic. A change of K_{ff} is itself treated as a transfer. The lag state is shifted by the step the new gain makes in the calculation and I in (3.16) takes back the same step, so a feed change already in transit keeps its timing and N_{SP} does not jump.

Dropping the maintenance terms from both halves of (3.18) is deliberate, because a feedforward that requires the biomass concentration requires a measurement that is rarely available online. The result is a little high, asking for 66.5 rev/min where the plant needs 64.1 rev/min at the design feed and 83.7 rev/min where it needs 82.2 rev/min at 40 L/min, errors of 3.8 % and 1.8 %. Errors of that size are typical of a calculated feedforward, and the feedback trim removes them. The lag τ_{ff} is a different matter entirely, and Section 3.6 measures what happens without it.

Pause. Try Exercise 3.3 before reading on.

3.5.4 Implementation

The app runs on the framework of Section 1.6, with manual and automatic modes and bumpless transfer in every architecture. Integrated error is scored against the true dissolved oxygen rather than the probe reading, so a biased probe scores badly although its trend is flat. Runs export one row every 1.08 s of simulated time for offline analysis, carrying the true dissolved oxygen, the probe state and the biased noisy signal the controller acted on as three separate series, which is enough to repeat every comparison in Section 3.6. Twelve built-in verification checks cover the opening steady state, the oxygen and carbon balances, the exponents of Equation (3.2), the sign of every control action and the four time constants of Section 3.4. Figure 3.2 shows scenario 5 after the feed step is rejected.

3.6 Results

Table 3.3 lists thirteen scenarios in instructional order. Each one selects a starting charge, a mode, an architecture, both sets of tuning constants, the disturbance settings and whatever timed events it needs. Every integrated absolute error quoted below is in percent-DO minutes accumulated over a twenty-minute run with the feed step at $t = 4$ min, except where another window is stated, and every travel figure is the total distance the speed demand moved in rev/min over the same window.

3.6.1 Open-Loop Response

Scenario 1 opens four hours into a fed-batch, with 5 g/L of cells in 48 m^3 on an 8 L/min feed and the agitator speed demand parked at 42.7 rev/min, where transfer balanced uptake when the run started. No input changes for twenty minutes. The single-loop drive runs the open-loop torque schedule of Equation (3.13), so the shaft sags to 41.4 rev/min within a minute and $k_L a$ falls by 6 %. Cell mass grows from 5.00 to 5.32 g/L, but the uptake term of Equation (3.1) stays near 18.3 mg/(L·min), because the feed pump rather than the cell mass sets the growth rate of a substrate-limited fed-batch. The dissolved oxygen falls from 25.0 to a minimum of 16.0 % at 15 min and reads 16.1 % at twenty minutes. The fall is arrested by the plant rather than by any control action. As the dissolved oxygen drops, the driving force $C^* - C_L$ widens until transfer meets uptake again, and by sixty minutes the reading has recovered to 18.1 % as the glucose pool runs down and the uptake eases. The plant is

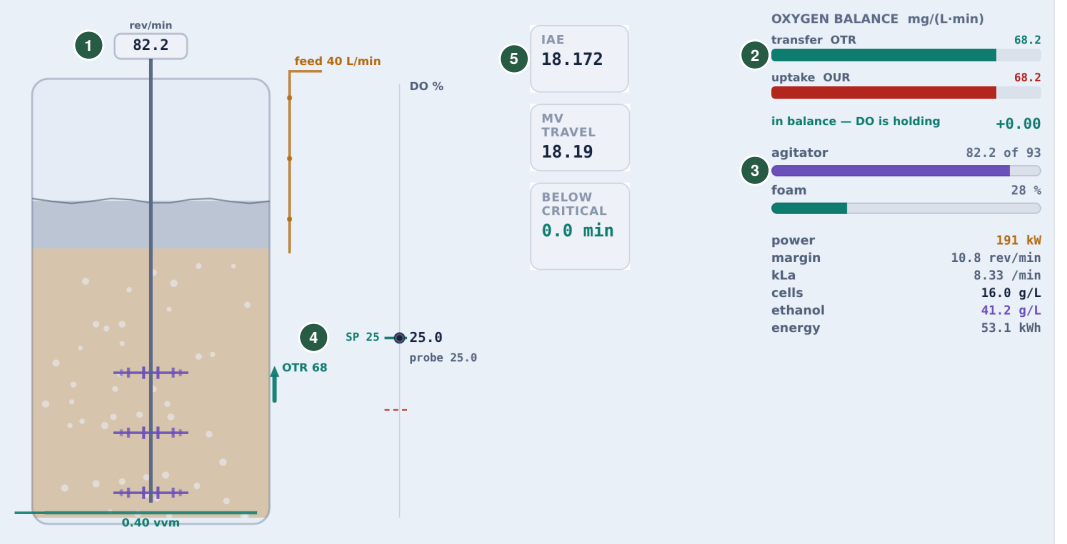


Figure 3.2 Scenario 5 twenty simulated minutes after the start, with the cascade in automatic and the feed step at $t = 4$ min rejected: the plant view of the capture with three scorecard tiles of the same capture pasted into its empty middle (a composite). (1) The agitator speed, moved from the design 64.1 rev/min of Table 3.1 to 82.2 rev/min. (2) Transfer and uptake, level at 68.2 mg/(L·min). (3) The agitator bar, with 10.8 rev/min of margin and 191 kW of shaft power listed below it, against the 88 kW of the design condition (Section 3.2). (4) The setpoint marker, where the probe is back at the 25 % setpoint. The true dissolved oxygen dipped 9.34 % below it on the way (Table 3.4). (5) The pasted tiles: an integrated error of 18.17, 18.19 rev/min of demand travel and no time below the critical value.

Table 3.3 The Thirteen Scenarios and the Mechanism Each One Isolates

#	Scenario	What it isolates
1	Early batch, open loop	a variable with no inventory, self-regulation
2	Late batch, open loop	identical CV, transformed MV
3	Single loop on a feed step	three lags in series, gain limit
4	Single loop at excessive gain	actuator travel as a cost
5	Cascade at the same tuning	what an inner loop removes
6	Cascade with detuned inner loop	the crossover from benefit to harm
7	Probe at 60 s	a lag no structure reaches
8	Feedforward from the feed	wrong in size against wrong in time
9	Feed raised at the speed stop	capacity, not tuning
10	Airflow against back pressure	a constraint that is not a range
11	Probe biased 14 % high	a perfect trend at the wrong value
12	Probe noise with derivative on	differentiating a slow measurement
13	Graded challenge, three faults	the whole ladder under time

stable at whatever value it arrives at rather than at a chosen one, so control supplies the selected operating point rather than stability. With the cascade's speed loop holding the shaft at 42.7 rev/min, the same twenty minutes stay between 24.7 and 25.3 %.

Scenario 2 repeats the experiment ten hours later at 34 g/L of cells in 92 m^3 on a 58 L/min feed, where the oxygen demand is 3.6 times larger. The dissolved oxygen reaches a minimum of 15.2 % at 4.4 min and reads 16.5 % at twenty minutes, against 16.0 and 16.1 % in scenario 1, which is nearly

Table 3.4 Integrated Absolute Error, Peak Deviation and Demand Travel on the Same 24 L/min to 40 L/min Feed Step, Read at $t = 20$ min With $\tau_I = 0.6$ min Throughout.

K_c	Single loop IAE / peak / travel	Cascade IAE / peak / travel	+ Feedforward IAE / peak / travel
0.6	23.8 / 11.83 / 25	18.2 / 9.34 / 18	5.9 / 2.90 / 18
1.8	10.6 / 8.82 / 44	6.0 / 6.19 / 18	2.4 / 2.43 / 18
3.2	12.0 / 5.18 / 95	**3.5** / 4.82 / 23	**2.1** / 2.13 / 22
4.5	57.1 / 21.02 / 683	11.7 / 3.93 / 152	10.1 / 2.69 / 137

the same excursion. The demand does not set the excursion. The open-loop drive loses about the same fraction of its speed at both ends of the batch, from 86.0 to 83.0 rev/min here against 42.7 to 41.4 rev/min early, and the same fractional loss of $k_L a$ costs the same few points of dissolved oxygen. The changes underneath are not visible on that trace. The agitator is at 83.0 rev/min instead of 41.4, drawing 188 kW instead of 23, with 10.0 rev/min of margin left instead of 51.6. The same batch run the same way is eight times as expensive to aerate and has spent more than four fifths of its actuator range.

3.6.2 Three Architectures on One Feed Step

Table 3.4 carries the central comparison. The feed steps from 24 to 40 L/min at $t = 4$ min from the design charge, the integral time is held at 0.6 min throughout, the inner loop is the shipped fast one, and every figure is read at $t = 20$ min.

Three results come out of Table 3.4. At the shipped tuning the cascade improves the peak from 11.83 to 9.34 % and the integrated error from 23.8 to 18.2, a modest gain, because it removed 15.4 s from a chain that still contains 12.4 s of liquid and 20 s of probe. The larger difference is in what the outer loop can then be tuned to. The single loop reaches its best integrated error of 10.6 at $K_c = 1.8$ and spends 44 rev/min of travel getting there, while the cascade reaches 3.5 at $K_c = 3.2$ for 23 rev/min, a factor of three on the error and a factor of two on the wear. And the single loop's best peak sits at a different gain from its best integrated error, which is the usual sign that a loop is being pushed against a lag rather than tuned against a process. Figure 3.3 draws the integrated error and the travel of Table 3.4 as bars.

Scenario 4 is what happens past that point, with K_c at 4.5 and the integral time also shortened to 0.5 min. At the design operating point this tuning is unstable. The oscillation grows from the start, the demand swings between its speed limits, and the dissolved oxygen reaches 23.77 % above the setpoint at 3.9 min and 21.47 % below it at 4.3 min, just after the feed step. The demand is close to a square wave and the actual speed is a rounded version of it running behind, so every correction arrives while the previous one is still in transit. Once the feed step has moved the operating point to 82 rev/min, where the drive lag and the oxygen holdup are both shorter, the same tuning is stable and the oscillation decays to ± 1.1 % over the last four minutes. The run ends with an integrated error of 72.5, 1.06 min below the critical value and 859 rev/min of demand travel against 25 at the shipped tuning. The significant number is the travel rather than the error, because 859 rev/min of demand movement in twenty minutes is gearbox and mechanical-seal wear spent to obtain a worse result than a smaller gain delivers. A tuning that is unstable at one point of the batch and stable at another is the operating-point dependence of the three lags seen from the outer loop.

> ✦ **DID YOU KNOW?** **One Outer Gain, Two Very Different Bills**
>
> From the simulator: at an outer gain of 4.5 on the same feed step, the single loop moves its speed demand 683 rev/min for a peak deviation of 21.02 %, while the cascade moves it 152 rev/min for 3.93 % (Table 3.4). The same gain costs the drive far less once a speed loop sits under it.

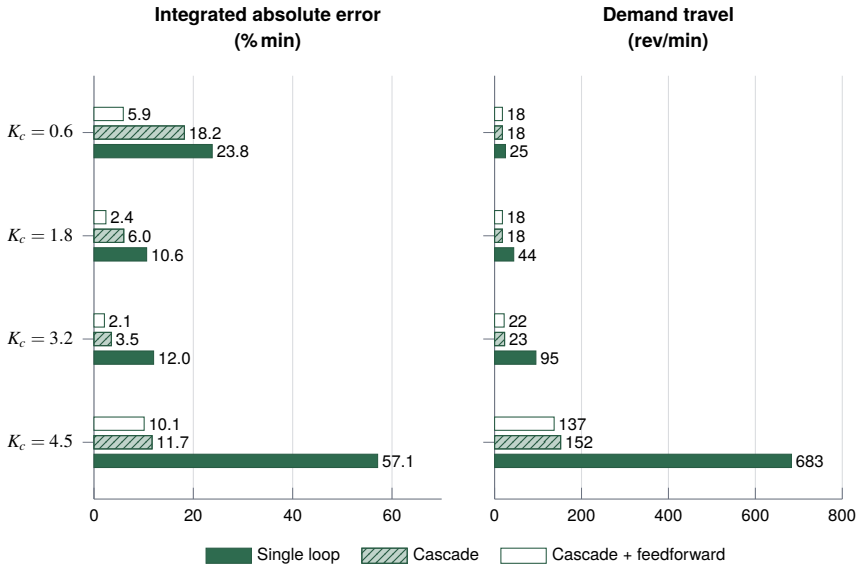


Figure 3.3 Table 3.4 redrawn: integrated absolute error and demand travel of the three architectures at four outer gains on the same feed step, the single loop in solid bars, the cascade in hatched bars and the cascade with feedforward in open bars, each panel on its own scale. Notice that the single loop's error rises again above $K_c = 1.8$ while its travel climbs at every step, and that the cascade keeps its travel between 18 and 23 rev/min up to $K_c = 3.2$.

3.6.3 Inner Loop Speed Requirement

Scenario 6 changes one thing about scenario 5. The inner gain is cut from 6 to 0.35 and the inner integral time is stretched from 0.10 to 1.20 min, so that by Equation (3.12) the speed loop settles in 12.3 s instead of 2.9 s. The detuned inner loop still takes the shaft to its setpoint. The result is worse than no cascade at all. Peak deviation is 14.84 % against the single loop's 11.83 %, the integrated error is 72.3 against 23.8, and the batch spends 2.76 min under the critical dissolved oxygen where the single loop spent 0.53. Detuning the inner loop turned a 23 % improvement into a threefold degradation, because an inner loop slower than the process it sits inside is not a cascade but an extra lag with an extra integrator in it.

> **TRY IT IN THE APP** **6 · The Same Cascade With a Slow Inner Loop**
>
> Load the scenario, press Run in Auto and let twenty simulated minutes pass. Watch the IAE and Below critical tiles. The chapter reports an integrated error of 72.3 and 2.76 min under the critical value, against 23.8 and 0.53 for the single loop of scenario 3, from a speed loop that still reaches its setpoint.

Sweeping the inner gain locates the crossover, and repeating the sweep at a different inner integral time shows what the rule of thumb leaves out. Figure 3.4 carries both sweeps with the outer controller held at $K_c = 0.6$ throughout.

The upper sweep holds the inner integral time at the 1.20 min scenario 6 ships and walks $K_{c,N}$ through 0.15, 0.25, 0.35, 0.50, 0.75, 1.0, 1.5, 2.0, 4.0 and 6.0, giving integrated errors of 172, 102, 72.3, 52.6, 38.3, 30.5, 23.4, 20.3, 18.2 and 18.2. It crosses the single loop's 23.8 at an inner time constant of about 7.4 s, which is 0.6 times the 12.4 s oxygen holdup the loop is nested inside. Below that speed the cascade is a liability. Above it the improvement continues until roughly 4 s and then stops. Going from 6.2 to 3.9 s takes the integrated error from 20.3 to 18.2, and the last step to 2.8 s

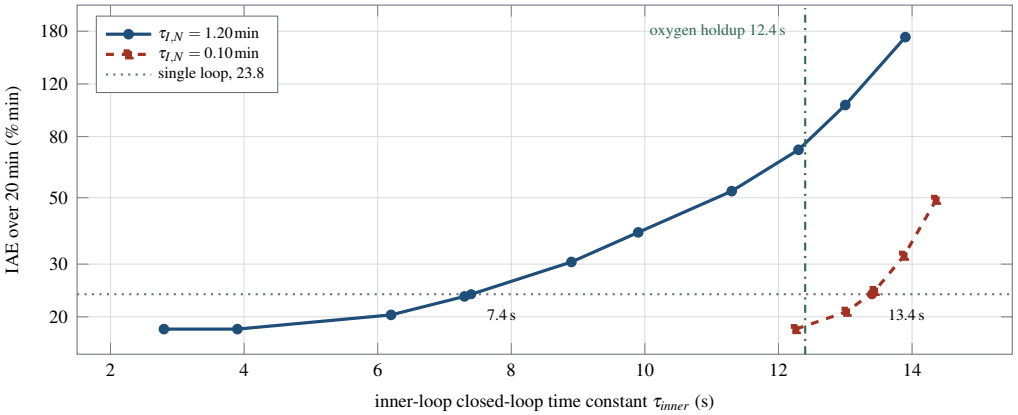


Figure 3.4 Where a cascade stops paying, measured on the scenario 5 feed step with the dissolved-oxygen controller held at $K_c = 0.6$ and $\tau_I = 0.6$ min. The inner gain $K_{c,N}$ is walked from 0.15 to 6.0 and the abscissa is the closed-loop time constant Equation (3.12) returns for it. Solid line with circles: $\tau_{I,N} = 1.20$ min. Dashed line with squares: $\tau_{I,N} = 0.10$ min. Below the dotted horizontal line the cascade beats the single loop and above it the cascade is worse than no cascade. The dash-dotted vertical line marks the oxygen holdup. The crossover moves from 7.4 s to 13.4 s when only the inner integral time is changed.

changes it by less than 0.01. The inner loop has become invisible to the outer one, which is what a cascade is trying to achieve.

The second sweep identifies what the conventional rule omits. Restoring the inner integral time to 0.10 min and repeating with $K_{c,N}$ at 0.10, 0.15, 0.20, 0.25 and 0.35 gives 49.2, 32.0, 24.4, 20.8 and 18.2. Equation (3.12) assigns those same gains time constants of 14.36, 13.88, 13.44, 13.02 and 12.26 s, so the crossover has moved out to about 13.4 s, nearly twice as slow as before, and everything available is collected by 12.3 s. A fast integral term is doing work the proportional term was doing in the first sweep. The number that decides whether a cascade pays is the inner loop's whole closed-loop response and not the proportional gain alone, a quantity that a rule expressed in gains does not contain.

Stated as a rule the measurement supports, a cascade stops paying at about the moment its inner loop becomes as slow as the fastest dynamics in the outer process. Breaking even needs an inner loop about 1.7 times faster than that fastest element, 7.4 s against the 12.4 s holdup, while collecting the whole benefit needs about three times faster, near 4 s. The conventional three to five is under-specified rather than wrong, because the whole outer chain is 12.4 s of liquid plus 20 s of probe and three to five times faster than that 32 s total is 6.5 to 11 s, which brackets the measured crossover only because the probe dominates the sum. A comparison against the wrong time constant errs by a factor of three. Two practical consequences follow. At the crossover the cascade is actively harmful rather than neutral, so the rule has no safe side to err toward at the slow end, and inner speed beyond the point where it stops mattering costs nothing, which is the argument for specifying an inner loop generously rather than exactly.

The inner measurement imposes its own limit. Running scenario 5 with the tachometer filter at 0.5, 1, 2, 4 and 6 s gives an integrated error of 18.2 every time with 18 rev/min of demand travel. At 8 s the loop goes unstable, at an integrated error of 73.9 and 120 rev/min of demand travel with a shaft that moves 570 rev/min in response to its own filtered measurement. An inner loop with a gain of 6 and a measurement filtered at 8 s is not a slow cascade but an unstable one, and the failure appears in the outer loop as an oxygen excursion that resembles a tuning problem. The tachometer is specified at 0.5 s for that reason, and half a second against twenty seconds of dissolved-oxygen probe states

the cascade argument in two numbers.

3.6.4 Probe Lag

Scenario 7 triples the probe time constant from 20 to 60 s and changes nothing else. The cascade's peak deviation goes from 9.34 to 12.01 % and the batch now spends 0.77 min under the critical value where before it spent none. The single loop goes from 11.83 to 14.12 %. The penalty is 2.7 points for the cascade and 2.3 for the single loop, which within the resolution of the experiment is the same penalty paid twice, and the cascade absorbed none of it. Set beside the 15.4 s of drive lag the same cascade removed outright, the asymmetry is the central result of the comparison.

A second-order effect deserves attention because it appears beneficial. A slow probe is also a filter. Pushing the cascade to $K_c = 3.2$ with the probe at 10 s makes the loop oscillate, at an integrated error of 38.7 for 696 rev/min of demand travel, where the shipped 20 s probe gives 3.5 for 23 and the aged 60 s probe gives 3.4 for 18. The slow probe delivers stability by discarding the information that would have made the loop fast, so a loop tuned against an aged probe becomes unstable when a new one is fitted. No control-system diagnostic reports that the membrane has aged, because a slower probe makes the trend appear calmer, and the record that would catch it is a calibration sheet carrying a response time.

3.6.5 Feedforward Magnitude and Timing

Scenario 8 adds the carbon-balance term of Equation (3.18) to the same cascade. Peak deviation falls to 2.90 % and the integrated error to 5.9, against 9.34 and 18.2 for the cascade alone and 11.83 and 23.8 for the single loop, and demand travel is unchanged at 18 rev/min, so the improvement came free of actuator wear. Setting K_{ff} to zero before the run, or carrying that setting through a restart, reproduces scenario 5 to within 1.6×10^{-13} percentage points of dissolved oxygen, which is the cleanest way to separate what the calculation contributes from what the structure contributes. Made during a run, the change is bumpless. The bias falls to the 21 rev/min speed floor in one execution and the integral takes back the same step, so the speed demand does not move and the loop continues as the plain cascade.

Setting the feedforward lag τ_{ff} to zero is the more instructive mistake. The agitator reaches the new speed the instant the feed changes while the oxygen demand is still a minute or more behind, and the dissolved oxygen overshoots 22.86 % upward with an integrated error of 15.1 and twice the travel, so a correct but early feedforward acts as a disturbance in its own right, the early case of Section 1.4.3. Matching τ_{ff} to the glucose pool at 0.8 min brings the error back to 5.9 and makes it nearly symmetric, at 2.90 % low and 1.75 % high. A feedforward admits two errors, a static calculation from incorrect data and a dynamic mismatch between when the term acts and when the disturbance arrives. Feedback corrects the first and only the lag corrects the second.

> **TRY IT IN THE APP** **8 · Feedforward From the Feed the Operator Is Already Setting**
>
> Run the scenario as loaded, then press Reset, switch to Advanced, set Feedforward lag to 0 and run again. Watch which way the dissolved oxygen leaves its setpoint after the feed step. The chapter reports a dip of 2.90 % and an integrated error of 5.9 with the lag at 0.8 min, and 22.86 % upward with 15.1 without it.

3.6.6 Transfer Capacity Limits

Scenario 9 raises the feed from 58 to 88 L/min at 5 min late in the batch, with the agitator already at 86.0 rev/min and 7.0 rev/min of margin. The agitator reaches 93.0 rev/min at 5.6 min and stays there with zero margin, the dissolved oxygen falls to 6.0 % by twenty minutes and keeps sinking slowly, to 3.2 % at three hours, and the batch spends 14.8 of the first 20 minutes below the critical value. No tuning changes any of it. A higher gain, a shorter integral time or the feedforward moves the time below critical by less than 0.4 min, because oxygen uptake is asking for 89.7 mg/(L·min) and the

Table 3.5 Vessel Back Pressure at the Design Charge, Holding 25 % Dissolved Oxygen After the 24 L/min to 40 L/min Feed Step.

p_{head}	C^*	Agitator	Shaft power
0 barg	6.59 mg/L	93.0 rev/min, pinned	281 kW
0.5 barg	9.84 mg/L	82.2 rev/min	191 kW
1.0 barg	13.09 mg/L	70.6 rev/min	118 kW
1.5 barg	16.34 mg/L	63.0 rev/min	83 kW

drive cannot transfer that much at this volume and airflow. A demand pinned against its limit while the error stays large is the signature of a capacity problem rather than a tuning one.

The cost of the fault is not primarily lost product. Growth in a substrate-limited fed-batch is set by the feed pump rather than by the oxygen, so the titer changes little for a long while and the unconsumed sugar accumulates instead. Residual glucose goes from 0.04 to 1.00 g/L at twenty minutes, 2.85 at an hour and 7.53 g/L at three hours. That is the first clear symptom of the fault and it appears on the culture trend rather than on the oxygen trace.

Airflow appears to be the remedy and fails. Scenario 10 applies the same 88 L/min feed at 4 min, and at 10 min, with the agitator pinned at its speed stop, more than doubles the air from 0.40 to 0.95 vvm. The instant the flow changes, the gas velocity term of Equation (3.2) lifts k_La from 9.53 to 11.83 min^{-1}, a 24 % gain, and the gassed power falls from 269 to 223 kW because the extra gas unloads the impellers through Equation (3.3). Then the foam of Equation (3.11) arrives. The headspace is full 3.5 min after the change, the sparged gas then passes through foam rather than through broth, and twenty minutes after the change k_La is back to 9.37 min^{-1}, below where it started. The dissolved oxygen ends at 5.6 % with 25.8 of the 30 minutes below the critical value. The airflow range extends to 1.2 vvm, while the effective limit arrived at 0.95 vvm on a different variable.

Back pressure succeeds where airflow does not, because it raises the equilibrium concentration in Equation (3.4) rather than the transfer coefficient. Table 3.5 sweeps it at the design charge on the same feed step.

At atmospheric pressure the plant cannot hold its own setpoint at the design charge once the feed has stepped, and it burns 281 kW failing to, while one and a half bar of head holds it for 83 kW, and the compressor duty that buys the difference is a rounding error beside a 200 kW agitator. On the capacity disturbance of scenario 9, with 1.5 barg set before the run starts, the same lever holds 25.0 % dissolved oxygen at 80.7 rev/min with 12.3 rev/min of margin at twenty minutes. Run for three hours, the pinned batch makes 83.38 g/L of ethanol for 792 kWh of agitator energy while the pressurized one makes 85.31 g/L for 466 kWh, with residual glucose at 0.05 g/L instead of 7.53 and half a minute below the critical value instead of 175. Raised at the 5 min feed step instead, the pressure gives the same titer for 476 kWh and no time below the critical value. Both batches reach the 105 m^3 working volume at 149 min, so their last half hour is made under overflow rather than fed-batch. The pressurized batch returns 2.3 % more product for 41 % less electricity, and the lever is a pressure setpoint rather than a controller tuning.

3.6.7 Instrument Faults

Scenario 11 puts 14 percentage points of positive bias on the probe of Equation (3.10) and changes nothing else. The controller holds the reading at 25.0 % and the true dissolved oxygen at 11.0 %, below the 15 % critical value for the whole run, and the trend remains flat. The batch loses nothing. After four hours it has made 49.24 g/L of ethanol against 49.25 with an honest probe, from the same 21.3 g/L of cells, and it used 279 kWh of agitator energy against 325 because it was holding a lower real dissolved oxygen.

The reason is structural. The feed pump sets the growth rate in a substrate-limited fed-batch, and oxygen becomes the limit only once it can no longer support the growth the feed supplies [17].

At this charge that crossover is at a true dissolved oxygen of 3.5 %, not 15, so the bias consumed margin rather than product, and it stays invisible until a second fault such as a fouled sparger removes that margin. This is the one fault in the collection that a better controller makes strictly worse. The defenses are outside the loop, a second probe, an off-gas balance that would fail to close, or calibration discipline, and the same conclusion has been reached for aeration control in wastewater treatment, where a biased sensor is invisible to every diagnostic built on control error [21].

Scenario 12 fails the other way, costing no product but a large amount of actuator wear. The probe carries 1.5 % of electrical noise, the feed is steady, the vessel is otherwise at steady state, and derivative action is switched on at 0.15 min with no filter. Twenty idle minutes cost 13 943 rev/min of demand travel. Filtering the derivative at 0.05 min drops that to 2350 and at 0.20 min to 1076, although the filtered runs pay for it in integrated error, 34.3 and 30.0 against 25.4. Removing the derivative altogether gives 613 rev/min with the best integrated error of the four at 21.4. Dropping the noise from 1.5 to 0.5 % takes the 613 to 204, and removing it entirely leaves 0.25 rev/min of travel. Derivative action supplies anticipation, and on a 20 s probe it has nothing to anticipate, because a real change is no longer predictive by the time it reaches the derivative term and everything the term responds to promptly is noise. The general test is whether the measurement is faster than the quantity the derivative is asked to predict, and a dissolved-oxygen probe fails it on both counts.

3.6.8 Combined Fault Scenario

Scenario 13 applies three faults over 45 min to a cautious single loop ($K_c = 0.25$, $\tau_I = 2.0$ min), a feed increase from 24 to 44 L/min at $t = 6$, a probe membrane fouling that triples the probe lag at $t = 18$ and a quarter of the sparger plugging at $t = 28$. The case is passed when the true dissolved oxygen stays within ± 4 % of the setpoint over the last third of the run, from 29.7 min, and the agitator finishes off its speed limit. With the shipped settings the batch spends 26.0 of the 45 minutes below the critical value, ends at 9.1 %, burns 137 kWh and leaves the band by up to 15.9 %, nearly four times the band, although the agitator ends 5.5 rev/min below its limit. The case therefore fails on the band alone. Tuning alone removes most of the error and then meets a capacity limit. The feedforward architecture with a fast inner loop ($K_{c,N} = 8$, $\tau_{I,N} = 0.08$ min) and $K_c = 2.6$ at $\tau_I = 0.5$ min cuts the integrated error from 471 to 158 and the time below critical to 4.32 min, but the shaft reaches its speed limit at 28.3 min, just after the sparger fault, and finishes with no margin at all. Adding 1.0 barg of back pressure to that same tuning, which starts the plant at the 1.0 barg steady state with the agitator at 55.1 rev/min, gives a worst deviation of 0.19 % over the judged window, inside the band with 3.8 points to spare, an integrated error of 7.1, 0.17 min below critical, 9.0 rev/min of margin still in hand and 110 kWh against the 162 kWh the tuning-only configuration spent. The rule reads only the judged window and the final margin, so the back pressure with $K_c = 2.6$ and $\tau_I = 0.5$ min on the single loop also passes, at 2.4 %, after a 28.7 % swing and 696 rev/min of demand travel early in the run, while neither the pressure (12.5 %) nor the tuning (10.6 %) passes alone. The ranking of the contributions is the result to retain. The inner loop and the architecture buy the response, the feedforward buys the feed disturbance, the pressure buys the capacity, and no measure recovers the probe lag.

3.7 Technology in Practice

Figure 3.5 draws the cascade and its feedforward on real hardware, where the speed loop lives inside the drive.

What Measures It

The outer loop reads a membrane probe, AT-301, which the model of Equation (3.10) reduces to one first-order lag of 20 s with a bias and a noise term. The inner loop reads the shaft. A tachometer is typically a small generator whose voltage is proportional to speed, often on a 0 to 10 V scale [22], and the model's ST-301 is that reading through a 0.5 s filter with no bias and no noise.

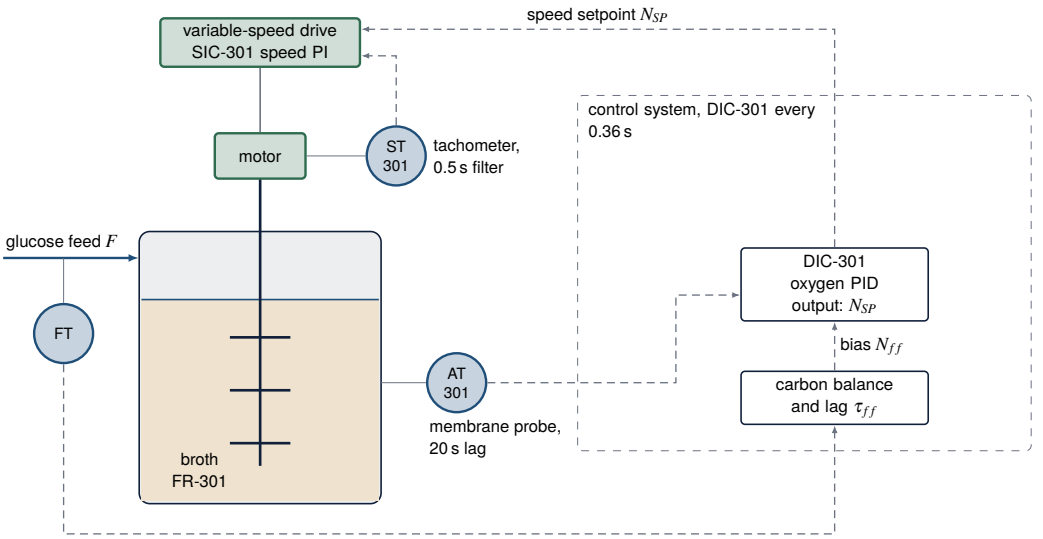


Figure 3.5 The oxygen cascade and its feedforward on real hardware. The membrane probe AT-301 feeds the oxygen controller DIC-301, whose output is a speed setpoint for the speed loop SIC-301, which runs inside the variable-speed drive, reads the tachometer ST-301 and sets the motor torque. The measured glucose feed enters DIC-301 as its bias through the carbon balance and its lag. Solid arrows are process streams and dashed gray arrows are signals. Notice how the slow probe and the fast tachometer close their loops in two different devices.

> **TECHNOLOGY** **Sensor: A Membrane Electrode That Consumes What It Reads**
>
> A Clark-type electrode reduces oxygen at a cathode behind a membrane, so it consumes the oxygen it measures. Its signal then depends on the flow past the membrane, where a depleted boundary layer forms, and its response is close to first order with a time constant that is generally not small against $1/k_La$ [13]. One probe on a 15 L cell-culture vessel measured 49.6 s, and microbial fermentations run transfer coefficients more than ten times higher [23]. A field manual asks for about one foot per second of water across such a membrane, and lists a fouled membrane as the cause of an instrument too slow to react [14]. On a trend that fault looks like a calmer loop, the change from 20 s to 60 s of scenario 7.

What Moves It

The final element is the agitator motor on its variable-speed drive. The model's drive is a torque balance with a 15.4 s lag at the design speed, clamped to 0 to 100 % of rated torque and to 21 to 93 rev/min. Real drives carry the same limits as settings: speed limiting, torque limiting, current limiting, and limits on acceleration and deceleration [22]. The last of these is the ramp that the Assumptions of Section 3.3.6 say a real drive follows, so a large speed change takes longer than the model shows.

> **TECHNOLOGY** **Final Element: A Drive That Closes Its Own Speed Loop**
>
> A variable-speed drive that reads a tachometer on the motor shaft throttles power until the shaft turns at the commanded speed. That built-in loop makes the drive the secondary controller of a cascade, taking its speed setpoint from the process controller. In the direct-current drives that rely on this feedback, a lost tachometer signal reads as zero speed and the motor runs away to full speed [22]. An alternating-current drive sets speed through its output frequency and knows the speed only approximately without feedback, and precise control, above all at low speed where slip is a larger share, may need a speed sensor [22]. Equation (3.13) is such a drive without its sensor.

Where the Controller Runs

DIC-301 would run in a distributed control system or a programmable logic controller, and SIC-301 inside the drive itself, which takes the speed demand as its setpoint. A host system can often update a networked drive's limits while the motor runs [22]. Loops in such systems are sometimes nested or cascaded, with cycle times from milliseconds to minutes [24], and the app's 0.36 s execution period for both controllers sits inside that range. The feedforward of Section 3.5.3 enters DIC-301 as its bias, so it belongs in the same controller as the oxygen loop.

What Is Modernizing It

Three developments touch this loop. First, optical probes measure the lifetime of a quenched luminescence, consume no oxygen, need no flow past the sensor and hold their calibration longer than amperometric ones [14]. Second, soft sensors compute the oxygen uptake from the probe reading, a characterized transfer coefficient and the solubility, which removes the need for an off-gas analyzer for that purpose [23]. Third, such soft sensors commonly run in the supervisory systems above the control level, because scripts written into the control system itself are the exception, and in an industrial plant they must first pass a functional and risk assessment [25].

> **IN PRACTICE** **Inspect the Membrane, Check the Zero, Then Calibrate**
>
> A federal field manual sets a routine for membrane probes. Inspect the membrane for wrinkles, tears and bubbles under it before use. Read the probe in a fresh zero-oxygen sodium sulfite solution, and if more than 0.2 mg/L persists, replace the membrane and the electrolyte. Let a new membrane condition for two to six hours, overnight for greater stability, before calibrating. Calibrate at one point, full saturation in air or in air-saturated water, with the sensor and its thermistor at the same temperature, and expect an amperometric sensor to need recalibration every week or two [14]. The zero check is the one that exposes a constant offset like the 14 points of scenario 11.

3.8 Limitations

The application is not a design tool. With one perfectly mixed lump, no gas-phase state, fixed correlation exponents, unstructured kinetics and no cooling model, the application cannot support a claim about a real vessel's oxygen capacity, its agitator sizing or the yield a real batch would make, and Section 3.3.6 says where each of those simplifications breaks. The upsets are idealized as well. A feed step from 24 to 40 L/min arriving in one instant would not occur under a fed-batch recipe, and a sparger that loses a quarter of its holes at once isolates a loss of transfer capacity from everything else. The twenty-minute scenarios compress a batch that runs fifteen to sixty hours, so no number in the application is a batch yield or a real cycle time.

The control scope is one dissolved-oxygen loop with one cascade and one feedforward. There is no respiratory quotient control, no substrate feedforward from an off-gas balance, no model predictive control and no supervisory optimizer choosing the feed profile, and a modern plant would have at least the off-gas balance. Airflow and back pressure are operator settings rather than controlled variables, and on a real vessel both would be loops of their own. The cell-culture vessel of Volume 19, Chapter 8 sequences agitation, oxygen enrichment and gas flow on one dissolved-oxygen loop. The feed profile is an input rather than a decision the application optimizes, although the real question on a fed-batch is what feed trajectory maximizes titer against the oxygen the vessel can transfer. There is no electricity price and no product price, so the reported metrics state what a decision cost in kilowatt hours and in grams per liter without saying which way the decision should go. Nothing fails outright, because the critical dissolved oxygen, the foam-out, the speed limit and the working volume annunciate and the run continues past the point where a real batch would have been shut down. Perfect repeatability is itself an idealization, because a fixed integration step and a fixed execution period give one trajectory on every machine while a distributed control system scanning about once a second gives a slightly different one every batch.

The model of this chapter also supports control and estimation methods beyond the implemented PID, cascade and feedforward loops. The nine states with the speed limit, the torque limit and the foam constraint form a natural plant for Model Predictive Control, which would coordinate agitation, airflow and back pressure as one constrained multivariable problem and could weight the agitator electricity against the dissolved-oxygen target as an economic objective. The same structure supports state and parameter estimation, because a moving horizon estimator or a Kalman filter driven by the probe signal and the measured shaft power can reconstruct the oxygen uptake rate and the biomass concentration that no instrument measures directly, a soft sensor that addresses the biased probe of scenario 11 from outside the loop. The APMonitor framework implements MPC and moving horizon estimation for differential algebraic models of exactly this form [26], and the Gekko Optimization Suite provides the same methods in Python [27], so the constants of Table 3.2 transfer directly to a graduate study of advanced control.

> ⚠ **OPEN CHALLENGE** **Gradients That One Probe Cannot See**
>
> A large stirred tank develops gradients when its power input is limited. Computational fluid dynamics can predict them, but one turbulent simulation may take hours, too slow for routine optimization or control. A 2026 study trained surrogate models on 760 such simulations of a 170 m^3 vessel with four impellers and predicted the three-dimensional substrate field in milliseconds, at a test R^2 near 0.91. Its simulations were single-phase, without experimental validation, for one geometry at one scale [28]. Progress would be a validated model that carries the gas phase, so that a dissolved-oxygen probe could be read against a known gradient.

3.9 Discussion

The dissolved-oxygen setpoint is an economic decision as much as a control decision, and no control-error metric reports the cost. Agitator power goes as the cube of speed through Equation (3.2), while transfer goes as the 2.1 power, so the last few points of dissolved oxygen are the expensive ones. At the design charge, holding 10 % costs 72.9 kWh over an hour, 25 % costs 86.0 and 50 % costs 118.8 kWh, and all three make the same ethanol to within 0.01 g/L, 42.41 g/L. That is 63 % more electricity for no additional product, committed by the choice of setpoint. Aeration is the dominant mechanical energy demand of a stirred production fermenter [7], and on a commodity fuel whose net energy balance has been debated [19] the difference between 52.0 kWh and 90.4 kWh per tonne that scenario 9 measures is not a rounding error.

Two quantities that no control-error metric contains are therefore reported beside it. Actuator travel is a maintenance budget, and the 859 rev/min of scenario 4 and the 13 943 rev/min of scenario 12 represent gearbox, mechanical seal and drive wear rather than abstract numbers. Speed margin is capacity, and no tuning creates more of it. The application reports six competing numbers rather than one composite score for that reason, because a run that scores well on error by pinning the agitator has exhausted its capacity margin rather than solved the problem.

> ⚠ **OPEN CHALLENGE** **A Soft Sensor Inherits Its Probe's Faults**
>
> A soft sensor joins hardware signals to a model to predict a quantity no instrument reads, such as the oxygen uptake of a culture. A critical review notes that a faulty input very probably gives a faulty output, yet studies that detect sensor faults in bioprocesses, or tolerate them, are rare [25]. An uptake estimate built on the biased probe of scenario 11 would start from a reading 14 points high. Progress would be estimators that test their own inputs, for example against a redundant probe, before the plant acts on them.

3.10 Conclusion

This chapter used an aerated fed-batch fermenter to measure how fast the inner loop of a cascade has to be. The simulator integrates nine states, the four culture balances, the dissolved oxygen, the agitator speed, the probe, foam and antifoam, with oxygen transfer from a Van 't Riet power law of

exponents 0.70 and 0.40. The vessel holds 107 g of dissolved oxygen against a 2.57 kg per minute uptake, so four lags decide the loop, a drive at 15.4 s, a liquid phase at 12.4 s, a probe at 20 s and a speed loop at 2.9 s. On one feed step the single loop reaches an integrated error of 10.6 for 44 rev/min of demand travel, the cascade 3.5 for 23 and the calculated feedforward 2.1 for 22. Detuning only the inner loop takes the cascade to 72.3 against the single loop's 23.8, and sweeping the inner gain places the crossover between benefit and harm at 7.4 s, which is 0.6 times the oxygen holdup, while the same sweep at a shorter inner integral time places it at 13.4 s. Tripling the probe lag costs 2.7 points of peak deviation in cascade and 2.3 in the single loop, which is the boundary of what any structure buys.

A fuller model would add a gas-phase state with an off-gas balance, a compartment model for the mixing gradient, a ramp limited drive and a cooling loop so that the two capacity constraints can be studied together. As it stands, the fermenter measures a cascade design rule instead of quoting it.

KEY POINTS

- A cascade helps only when its inner loop is fast enough, and a slow inner loop makes the oxygen loop worse than no cascade at all.
- Detuning only the inner loop takes the integrated error to 72.3 against the single loop's 23.8.
- The crossover between benefit and harm sits at an inner time constant of 7.4 s, 0.6 times the oxygen holdup, and at 13.4 s with a shorter inner integral time.
- A cascade reaches the actuator, a feedforward reaches the disturbance, and nothing reaches the measurement.

Where else this applies. Oxygen cascades return with a meal feedforward in Volume 19, Chapter 12 and under a supervisory setpoint in Volume 9, Chapter 10.

References

[1] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle, III, *Process Dynamics and Control*. John Wiley & Sons, 2016.

[2] P. R. Krishnaswamy, G. P. Rangaiah, R. K. Jha, and P. B. Deshpande, "When to use cascade control," *Industrial & Engineering Chemistry Research*, vol. 29, no. 10, pp. 2163–2166, 1990.

[3] G. K. McMillan, "Effect of cascade control on loop performance," in *1982 American Control Conference*, 1982, pp. 363–368.

[4] L. Åmand, G. Olsson, and B. Carlsson, "Aeration control – a review," *Water Science and Technology*, vol. 67, no. 11, pp. 2374–2398, 2013.

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

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

[7] D. Humbird, R. Davis, and J. D. McMillan, "Aeration costs in stirred-tank and bubble column bioreactors," *Biochemical Engineering Journal*, vol. 127, pp. 161–166, 2017.

[8] C. M. Cooper, G. A. Fernstrom, and S. A. Miller, "Performance of agitated gas-liquid contactors," *Industrial & Engineering Chemistry*, vol. 36, no. 6, pp. 504–509, 1944.

[9] K. Van't Riet, "Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels," *Industrial & Engineering Chemistry Process Design and Development*, vol. 18, no. 3, pp. 357–364, 1979.

[10] F. Garcia-Ochoa and E. Gomez, "Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview," *Biotechnology Advances*, vol. 27, no. 2, pp. 153–176, 2009.

[11] L. C. Clark, Jr., R. Wolf, D. Granger, and Z. Taylor, "Continuous recording of blood oxygen tensions by polarography," *Journal of Applied Physiology*, vol. 6, no. 3, pp. 189–193, 1953.

[12] M. J. Johnson, J. Borkowski, and C. Engblom, "Steam sterilizable probes for dissolved oxygen measurement," *Biotechnology and Bioengineering*, vol. 6, no. 4, pp. 457–468, 1964.

[13] T. V. Kirk and N. Szita, "Oxygen transfer characteristics of miniaturized bioreactor systems," *Biotechnology and Bioengineering*, vol. 110, no. 4, pp. 1005–1019, 2013.

[14] S. A. Rounds, F. D. Wilde, and G. F. Ritz, "Dissolved oxygen (ver. 3.0)," U.S. Geological Survey, Techniques of Water-Resources Investigations, book 9, chap. A6, sec. 6.2, 2013. [Online]. Available: https://pubs.usgs.gov/twri/twri9a6/twri9a62/twri9a6_6.2_ver3.pdf

[15] U.S. Geological Survey, "Dissolved oxygen," U.S. Geological Survey, Techniques and Methods, book 9, chap. A6.2, 2020.

[16] T. A. Clark, T. Hesketh, and T. Seddon, "Automatic control of dissolved oxygen tension via fermenter agitation speed," *Biotechnology and Bioengineering*, vol. 27, no. 10, pp. 1507–1511, 1985.

[17] B. Sonnleitner and O. Käppeli, "Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis," *Biotechnology and Bioengineering*, vol. 28, no. 6, pp. 927–937, 1986.

[18] R. Luedeking and E. L. Piret, "A kinetic study of the lactic acid fermentation. Batch process at controlled pH," *Journal of Biochemical and Microbiological Technology and Engineering*, vol. 1, no. 4, pp. 393–412, 1959.

[19] A. E. Farrell, R. J. Plevin, B. T. Turner, A. D. Jones, M. O'Hare, and D. M. Kammen, "Ethanol can contribute to energy and environmental goals," *Science*, vol. 311, no. 5760, pp. 506–508, 2006.

[20] S.-O. Enfors, M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jürgen, E. Krüger, T. Schweder, G. Hamer, D. O'Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. McFarlane, A. Nienow, T. Kovacs, C. Trägårdh, L. Fuchs, J. Revstedt, P. C. Friberg, B. Hjertager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H.-Y. Lin, P. Neubauer, R. van der Lans, K. Luyben, P. Vrabel, and Å. Manelius, "Physiological responses to mixing in large scale bioreactors," *Journal of Biotechnology*, vol. 85, no. 2, pp. 175–185, 2001.

[21] O. Samuelsson, G. Olsson, E. Lindblom, A. Björk, and B. Carlsson, "Sensor bias impact on efficient aeration control during diurnal load variations," *Water Science and Technology*, vol. 83, no. 6, pp. 1335–1346, 2021.

[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] M. Pappenreiter, B. Sissolak, W. Sommeregger, and G. Striedner, "Oxygen uptake rate soft-sensing via dynamic kLa computation: Cell volume and metabolic transition prediction in mammalian bioprocesses," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 195, 2019.

[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.

[25] V. Brunner, M. Siegl, D. Geier, and T. Becker, "Challenges in the development of soft sensors for bioprocesses: A critical review," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 722202, 2021.

[26] 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.

[27] 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.

[28] V. Puig-I-Laborda, J. Carrasco Muriel, A. Castañeda Tena, T. Groves, J. le Nepvou de Carfort, U. Krühne, K. V. Gernaey, C. Haringa, I. Puig-de Dou, and L. K. Nielsen, "Machine-learning surrogates enable rapid prediction of 3-D substrate gradients in industrial bioreactors," *Nature Communications*, vol. 17, p. 11636, 2026.

3.11 Exercises

3.1 ○ Concept check · 1 min

In the cascade of this chapter, what does the dissolved-oxygen controller DIC-301 manipulate?

(a) The motor torque command u_T, in percent of rated torque

(b) The glucose feed rate F, in liters per minute

(c) The agitator speed setpoint N_{SP}, in revolutions per minute

(d) The vessel back pressure p_{head}, in bar gauge

3.2 Concept check · 2 min

Three lags lie between the output of DIC-301 and its measurement: the agitator drive (15.4 s), the oxygen holdup of the broth (12.4 s) and the membrane probe (20 s). Which of them does a speed loop closed on the tachometer remove?

(a) The probe lag, because the tachometer reads far faster than the probe

(b) The oxygen holdup, because the agitator speed sets the transfer coefficient

(c) All three, as long as the speed loop is three to five times faster

(d) The drive lag only, since only the drive lies inside the speed loop

3.3 Concept check · 2 min

The carbon-balance feedforward of Equation (3.18) is computed from the glucose feed and drops the maintenance terms. Why is it not computed from the cell concentration X, which sets the uptake through Equation (3.6)?

(a) The feed rate is measured, and X would need a measurement rarely available online

(b) The maintenance terms are larger than the growth terms, so X adds nothing

(c) An off-gas analyzer already supplies the uptake, so X is not needed

(d) Dropping X makes the calculation exact, so no feedback trim is needed

3.4 Theory and calculation · 5 min

At the design charge the broth holds $65\,\mathrm{m^3}$ at 25 % dissolved oxygen, with $C_{ref} = 6.59\,\mathrm{mg/L}$, $C^* = 9.84\,\mathrm{mg/L}$ at 0.5 barg, an uptake of 39.6 mg/(L·min) and $k_La = 4.84\,\mathrm{min^{-1}}$.

(a) Find the dissolved oxygen concentration C_L and the mass of dissolved oxygen in the vessel. *Checkpoint:* 1.65 mg/L and 107 g.

(b) Find the oxygen the culture consumes per minute and the time in which it would consume the whole dissolved inventory. *Checkpoint:* 2.57 kg/min and 2.5 s.

(c) Find the liquid-phase time constant and check that transfer balances uptake at this state. *Checkpoint:* 12.4 s, and a transfer rate of 39.6 mg/(L·min).

3.5 Theory and calculation · 8 min

The agitator drive has $J = 60\,000\,\mathrm{kg\,m^2}$, $k_T = 13\,481\,\mathrm{N\,m}$ per $(\mathrm{rev/s})^2$ and a rated torque of 30 kN m. At the design point $N = 1.068\,\mathrm{rev/s}$ and $f_g = 0.849$.

(a) Find the open-loop drive time constant from Equation (3.9). *Checkpoint:* 15.4 s.

(b) Find the inner process gain $K_{p,N}$ and the closed-loop time constant of Equation (3.12) at the shipped $K_{c,N} = 6$ and at the detuned $K_{c,N} = 0.35$ of scenario 6. *Checkpoint:* 0.736 rev/min per percent, 2.85 s and 12.3 s.

(c) Find the inner gain at which Equation (3.12) gives the measured break-even time constant of 7.4 s, and say what the equation omits. *Checkpoint:* $K_{c,N} \approx 1.47$.

3.6 Theory and calculation · 6 min

Back pressure raises C^* through Equation (3.4). Take $C_{ref} = 6.59\,\mathrm{mg/L}$, $p_{atm} = 1.013\,25\,\mathrm{bar}$, a setpoint of 25 % and the uptake of 68.2 mg/(L·min) that the culture reaches after the scenario 5 feed step.

(a) Find C^* at 0.5 and 1.5 barg, and the ratio of the two driving forces $C^* - C_L$. *Checkpoint:* 9.84 mg/L, 16.35 mg/L and 1.79.

(b) Find the transfer coefficient that holds the setpoint at each pressure. *Checkpoint:* $8.33\,\mathrm{min^{-1}}$ and $4.64\,\mathrm{min^{-1}}$.

(c) At 0.5 barg the agitator turns at 82.2 rev/min for 191 kW. Use $k_La \propto N^{2.1}$ and a shaft power proportional to N^3 to estimate the speed and power at 1.5 barg, and compare with Table 3.5. *Checkpoint:* About 62.2 rev/min and 83 kW.

3.7 App challenge · 8 min

Load scenario 3, Single-loop DO control on a feed step. The glucose feed steps from 24 to 40 L/min at 4 min. Use the default seed and run each case for 20 min. For each run read the IAE and MV travel tiles of the scorecard and the last row of the exported column `do_max_error_pct`, the largest deviation of the true dissolved oxygen.

(a) Run the scenario as shipped, then set DO proportional gain to 3.2, press Reset and run again. *Checkpoint:* IAE 23.8, peak 11.83 %, travel 25 rev/min. At 3.2: 12.0, 5.18 % and 95 rev/min.

(b) Load scenario 5, Cascade: a speed loop under the DO loop, and repeat both runs. *Checkpoint:* IAE 18.2, peak 9.34 %, travel 18 rev/min. At 3.2: 3.5, 4.82 % and 23 rev/min.

(c) Explain why the cascade improves the IAE by a quarter at the shipped gain and by a factor of more than three at the higher gain.

Accept an IAE within 0.1 and a peak within 0.02 %.

3.8 App challenge · 8 min

Load scenario 6, The same cascade with a slow inner loop. Leave the Speed loop integral time at 1.20 min and the outer tuning as shipped. Use the default seed and run each case for 20 min, pressing Reset between runs.

(a) Set Speed loop gain to 1.0, then 1.5, then 2.0, and read the IAE tile after each run. *Checkpoint:* 30.5, 23.4 and 20.3, within 0.1.

(b) Compute the inner time constant for each gain from Equation (3.12) with $\tau_{drive} = 15.4\,\text{s}$ and $K_{p,N} = 0.736\,\text{rev/min}$ per percent. *Checkpoint:* 8.9 s, 7.3 s and 6.2 s.

(c) The single loop gives an IAE of 23.8 on the same step. Interpolate linearly for the inner time constant at which the cascade breaks even, and compare it with the 12.4 s oxygen holdup. *Checkpoint:* About 7.4 s, 0.6 times the holdup.

3.9 App challenge · 8 min

Load scenario 8, Feedforward from the feed the operator is already setting. The feed steps from 24 to 40 L/min at 4 min. Use the default seed and run each case for 20 min. For each run read the IAE and MV travel tiles and the extremes of the exported column `do_error_pct`, the true dissolved oxygen less the setpoint.

(a) Run the scenario as shipped ($K_{ff} = 1$, $\tau_{ff} = 0.8\,\text{min}$). *Checkpoint:* IAE 5.9, −2.90 % and +1.75 %, travel 18 rev/min.

(b) Set Feedforward lag to 0, press Reset and run again. *Checkpoint:* IAE 15.1, +22.86 % at 4.25 min, travel 38 rev/min.

(c) Press the reload button, set Feedforward gain to 0 and run again. Compare with scenario 5. *Checkpoint:* IAE 18.2, the scenario 5 value.

Accept an IAE within 0.1 and an error within 0.05 %.

3.10 Application area · 4 min

At the design charge, an hour at a dissolved-oxygen setpoint of 10 % costs 72.9 kWh of agitator energy, 25 % costs 86.0 kWh and 50 % costs 118.8 kWh, and all three make the same ethanol to within 0.01 g/L.

(a) Find the extra energy of the 25 % and the 50 % setpoints over the 10 % one. *Checkpoint:* 18 % and 63 %.

(b) The oxygen half-saturation is 5.3 % of air saturation. Give two reasons from the chapter for not running at 10 %.

3.11 Application area · 4 min

In scenario 11 the probe reads 14 percentage points high. The controller holds the reading at 25.0 % while the true dissolved oxygen is 11.0 %. After four hours the batch has made 49.24 g/L of ethanol against 49.25 with an honest probe, for 279 kWh of agitator energy against 325.

(a) Find the energy saved, and explain why the product barely changed. *Checkpoint:* 14 %.

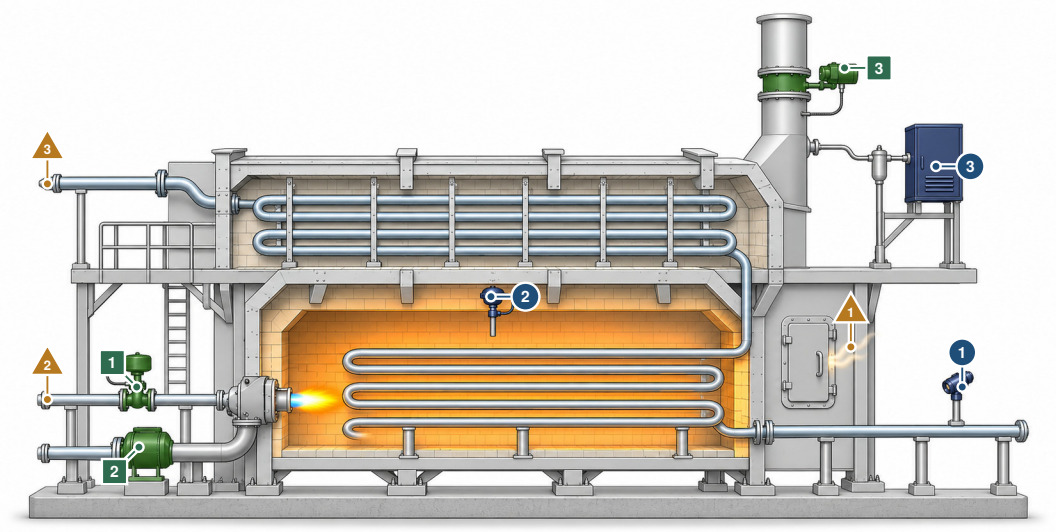
(b) Name two checks outside the loop that would expose the bias, and explain why a better controller makes this fault worse.

3.12 Review the agent · 5 min

An AI assistant is asked to cut wear on the agitator drive and replies as follows. “Detune the speed loop to $K_{c,N} = 0.35$ with $\tau_{I,N} = 1.20\,\text{min}$. Equation (3.12) gives 12.3 s. The outer process is 12.4 s of liquid plus 20 s of probe, 32 s in all, so the inner loop is still 2.6 times faster than the outer chain, close to the three-to-five rule, and the cascade keeps most of its benefit with gentler torque.” Find the flaw and say how to check it.

CHAPTER 4

Fired Heater Simulator for Cross-Limiting, Oxygen Trim and an Optimum Held by a Constraint



The controller moves 1 fuel gas valve 2 combustion air fan, through its flow setpoint 3 stack damper

The controller reads 1 oil outlet temperature 2 draft at the arch 3 stack oxygen and carbon monoxide analyzer

What disturbs it 1 air leaking in at a sight door 2 heating value of the fuel gas 3 oil flow to the heater

Notice that air leaking in at the door reaches the stack analyzer without passing the burner, so the analyzer counts air the flame never had. Illustration, not to scale.

IN PLAIN TERMS **Air Goes First, Then Find the Limit**

The main idea. A fired heater burns gas to warm oil flowing through its tubes. Its fuel valve answers in seconds, but the fan that blows in the air is six times slower.

The challenge. Too little air leaves unburned fuel in a hot box, which can explode, so two simple rules let the air move first. The best mix makes more carbon monoxide, a poison gas, than the permit allows, and a sensor in the chimney that can drift must find that limit.

Why it matters. The same rules guard any process where one mistake is a danger and the other is only a cost.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of combustion control and the fact on air leaks (Section 4.2), the try-it segments of scenarios 2 and 8 and the fact on draft (Section 4.5) and the key points
Study	35 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 4.5), Technology in Practice (Section 4.6), Limitations (Section 4.7) and the open challenge

The control question The forced-draft fan of a fired heater answers six times more slowly than its fuel valve, and too little air is a hazard while too much is a cost. How does a pair of selectors keep the burners lean through a load step, and why does a carbon monoxide override, not an oxygen setpoint, find the best legal operating point?

Learning objectives

- Compute the burner and stack air ratios, the stack oxygen and the carbon monoxide from the model, with and without tramp air.
- Explain how the selectors of cross-limiting make the air lead on an increase and follow on a decrease, and what the outlet pays for it.
- Predict the burner air ratio after an unmeasured heating value change and the trim that restores it.
- Locate the efficiency peak and the permit on a walk of the oxygen setpoint, and explain how a high-select override holds the optimum.
- Tell a drifting analyzer and tramp air from a real shortage of air, and judge what a trim clamp and an override each protect against.

Prerequisites Sections 1.4.4 and 1.4.6 (ratio with a trim, cross-limiting), 1.4.3 (static feedforward), 1.6.2 (velocity form) and 1.6.4 (scenarios and exported files).

The app https://apmonitor.com/apps/fired-heater/

See also Volume 4, Chapter 2 and Volume 4, Chapter 6 for protective overrides on a limit with a hazardous side. Volume 9, Chapter 8 for an optimum on a constraint that moves, and Volume 9, Chapter 4 for a heating value no controller measures. Volume 11, Chapter 3, Volume 11, Chapter 14 and Volume 11, Chapter 13 for refinery heaters, their fuel gas and an air trim on a drifting analyzer.

4.1 Introduction

Ratio control is usually taught as a symmetric problem, in which the wild stream moves and the controlled stream follows it at a fixed multiple [1, 2]. On a burner the ratio is not symmetric. Air in excess of what the fuel needs costs efficiency, while air short of it leaves unburned fuel in a hot enclosed space, and a burner management system exists precisely because that condition can end in an explosion [3]. The two flows also differ in speed. A fuel control valve reaches a new flow in about two seconds and a forced-draft fan with its damper takes about twelve, so a ratio that simply slaves the air to the measured fuel runs short of air on every load increase for as long as the fan lags.

The standard answer is cross-limiting. A low selector holds the fuel setpoint to what the measured air can burn and a high selector holds the air setpoint up to what the measured fuel needs, so the air leads on the way up and follows on the way down [3, 4]. Section 1.4.6 compares the three arrangements on the water heater of Chapter 1, and this chapter builds them on a refinery heater. Cross-limiting is a pair of selectors enforcing an inequality, and the same structure reappears one level higher. Efficiency on a heater has an interior maximum in excess air, but on the unit modeled here the maximum sits where carbon monoxide is more than four times the permit. The economic operating point is therefore the lowest excess air at which carbon monoxide stays legal, an active constraint whose position moves with burner condition, fuel and load. Holding a process on an active constraint of unknown position is a selector problem as well [5, 6]. Volume 9, Chapter 8 meets the same situation in a data hall, where the lowest-cost feasible point sits on a constraint that moves with the weather.

The simulator is aimed at two misconceptions. The first is that ratio control is symmetric, which the fuel-first scheme disproves in the first minute of the first load step. The second is that an optimum is a setpoint to be found and held. The class of problem is general: wherever one direction of error is a hazard and the controlled flows respond at different speeds, as in fuel and air on boilers, kilns and furnaces, reactant feeds to exothermic reactors and chemical dosing, the answer is an asymmetric limit rather than a better tuning, and wherever an optimizing loop sits over a safety-relevant constraint the answer is an override rather than a narrower setpoint.

This chapter works with a fired heater simulator in which the combustion scheme, the oxygen trim and its protections can be changed while the heater runs, and it measures the consequence of each change in air ratio, carbon monoxide, outlet temperature and fuel. Cross-limiting, oxygen trim and override control are standard methods. What the model adds is a pair of separate burner and stack air ratios, so that tramp air and analyzer faults behave as they do on a real heater, and a carbon monoxide override written on the logarithm of the concentration with external reset feedback. The scenarios run from a load step that goes fuel-rich to an analyzer that drifts the firebox rich over half an hour. Section 4.2 reviews combustion control on fired equipment. Section 4.3 sets out the model, its two air ratios and its assumptions, and Section 4.4 the temperature loop, the three fuel and air schemes, the trim with its override and the draft loop. Section 4.5 gives the measured results scenario by scenario, through the graded challenge. Section 4.6 describes the real instruments, fan and control platform, Section 4.7 states the limitations, Section 4.8 draws the conclusions, and the exercises follow in Section 4.9.

4.2 Combustion Control on Fired Equipment

Fuel and air were first coupled by a linkage, then metered and trimmed by an analyzer, as the history segment of this section tells. Oxygen tells the operator how much air is left over, while carbon monoxide tells the operator whether the burners were short of it.

> **HISTORY From a Shared Linkage to a Carbon Monoxide Trim, 1961 and 1984**
>
> Simple burners have long kept fuel and air in proportion mechanically. One rotary actuator moves the fuel gas valve through a cam and the air damper through a linkage, and changing the ratio is a task for a skilled mechanic. A more automated approach puts a flow loop on one line and a flow transmitter on the other [7]. Trimming the ratio needed a measurement of what the flame achieved. In 1961 a high-temperature galvanic cell was described whose voltage is nearly linear in the logarithm of the oxygen pressure between 600 and 750 °C [8]. By 1984 a steamship operator reported automatic combustion air trim by measuring carbon monoxide, unburned hydrocarbons and opacity in the flue gas, with fleetwide fuel savings of 3.5 percent [9].

Process heaters add their own problems to the boiler practice. The heat is delivered mostly by radiation in the firebox and partly by convection in a separate bank [10]. The draft is natural, held by a stack damper, so air leaks in through sight doors and casing seams, and that tramp air reaches the stack oxygen analyzer without ever passing a burner. Field practice treats a heater that reads normal oxygen with high carbon monoxide as a leaking heater before it treats it as a control problem [11]. Efficiency depends on excess air, stack temperature and ambient conditions in ways that have been measured on operating heaters [12], and the value of each percent of excess oxygen removed is what pays for the analyzers. Volume 11, Chapter 3 places a fired heater at the end of a fouling preheat train, where the fuel valve saturates during a cleaning outage and the crude charge rate has to give way.

> **DID YOU KNOW? A Leak Can Cost a Tenth of the Fuel**
>
> The draft of a furnace draws cold air in through cracks and badly closed doors, and that air leaves at flue gas temperature. In a high-temperature furnace that is not properly maintained, the extra fuel can reach 10 % of the fuel input [13].

4.3 Process Model

The heater warms 100 kg/s of oil from 253 to 360 °C, absorbing 29.96 MW. At design it fires 0.785 kg/s of fuel gas (2828 kg/h) with 15.36 kg/s of air, a burner air ratio of 1.150, and with 3 % tramp air the stack reads 3.00 % wet oxygen. Carbon monoxide is 41 ppm, the stack is at 355 °C, the bridgewall at 794 °C, and the efficiency is 81.16 %. The arrangement is shown in Figure 4.1.

The variables are classified in Table 4.1. Six states are integrated: the fuel flow F_f, the burner air flow F_a, the heat reaching the process Q_f, the bridgewall temperature T_{bw}, the outlet temperature T_{out}

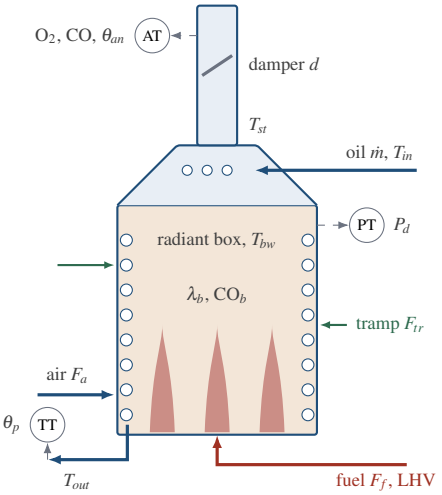


Figure 4.1 Natural-draft process heater. Fuel F_f of heating value LHV and air F_a from the forced-draft fan burn at a burner air ratio λ_b in the radiant box, whose bridgewall temperature is T_{bw}. Tramp air F_{tr} leaks in through sight doors and casing and joins the flue gas. The oil enters the convection bank at T_{in} with flow $\dot{m}$ and leaves the radiant coil at T_{out}, measured by TT after a delay θ_p. The stack damper d holds the arch draft P_d measured by PT, and the analyzer AT reads stack oxygen and carbon monoxide after a sample-line delay θ_{an}.

Table 4.1 Variables at the Design Point

Symbol	Role	Nominal	Range
T_{out}	CV, outlet temperature	360 °C	SP 330–390
O_2	CV, stack oxygen (wet)	3.0 %	SP 0.5–6
CO	constraint, stack	41 ppm	limit 200
P_d	CV, arch draft	−25 Pa	SP −60 to −5
D	MV, firing demand	73.8 %	0–100
ρ	MV, air ratio trim	0 %	±8 %
d	MV, stack damper	44 %	5–100
F_f, F_a	states, flows	0.785, 15.36 kg/s	—
Q_f, T_{bw}, T_{out}, P_d	states	—	—
$\dot{m}$	DV, process flow	100 %	60–120
T_{in}	DV, inlet temperature	253 °C	230–275
LHV	DV, heating value	100 %	85–115
F_{tr}	DV, tramp air at −25 Pa	3 %	0–15
b, $\dot{b}$	DV, analyzer bias, drift	0	±2, ±0.2/min
φ	DV, burner fouling	0	0–1
w	DV, wind gusts (1σ)	0 Pa	0–30

and the arch draft P_d. Everything about the combustion itself is algebraic in the two flows and the draft.

4.3.1 Two Air Ratios

The burners see only the air that passes through them, while the stack analyzer sees the air that reaches the stack. The two ratios are defined in Equation (4.1).

$$\lambda_b = \frac{F_a}{F_f\,\mathrm{AFR}}, \qquad \lambda_t = \frac{F_a + F_{tr}}{F_f\,\mathrm{AFR}} \tag{4.1}$$

where AFR is the stoichiometric air per kilogram of fuel. For gaseous hydrocarbons the air needed per megajoule is nearly constant, so both LHV and AFR scale with the heating value disturbance,

AFR $= 17.0\,(\mathrm{LHV}/47.0)$. A richer gas therefore needs more air per kilogram, and a ratio set in kilograms for the old gas is short of air on the new one. Tramp air flows through fixed openings and scales with the square root of the draft, Equation (4.2).

$$F_{tr} = \frac{\tau_{tr}}{100}\,(15\,\mathrm{kg/s})\sqrt{\frac{\max(0,-P_d)}{25\,\mathrm{Pa}}} \tag{4.2}$$

with τ_{tr} the leak in percent of design air at the design draft.

4.3.2 Carbon Monoxide, Heat Release and Stack Loss

Carbon monoxide leaving the burners follows an exponential correlation above a knee, Equation (4.3).

$$\mathrm{CO}_b = \min\left[50000,\ \mathrm{CO}_0 + 1000\exp\left(\frac{\lambda_k + 0.08\varphi - \lambda_b}{w}\right)\right] \tag{4.3}$$

with the floor $\mathrm{CO}_0 = 30\,\mathrm{ppm}$, the clean-burner knee $\lambda_k = 1.04$, the width $w = 0.025$ and burner fouling φ moving the knee up. The fraction of fuel energy not released is the fuel that finds no oxygen plus a small carbon monoxide penalty, and the stack loss is the flue gas heated from ambient to stack temperature, Equations (4.4) to (4.6).

$$u_b = \min[1,\ \max(0, 1-\lambda_b) + k_{CO}\min(\mathrm{CO}_b, 20000)] \tag{4.4}$$

$$Q_{rel} = \mathrm{LHV}\,F_f\,(1-u_b)$$

$$T_{st} = T_{in} + \Delta T_{ap}\left(\frac{F_{fl}}{F_{fl,0}}\right)^{0.6} \tag{4.5}$$

$$Q_{abs} = Q_{rel} - F_{fl}c_{p,g}(T_{st} - T_{amb}) - \varepsilon_c\,\mathrm{LHV}\,F_f \tag{4.6}$$

where $F_{fl} = F_f + F_a + F_{tr}$ is the flue gas flow, $F_{fl,0}$ its design value, ΔT_{ap} the stack approach at design flow and ε_c the casing loss. The efficiency is $\eta = Q_{abs}/(\mathrm{LHV}\,F_f)$. Below the knee, carbon monoxide and unburned fuel cost efficiency, and above it every kilogram of extra air is heated to stack temperature, so the efficiency has an interior maximum. Carbon monoxide at the stack is diluted by tramp air, $\mathrm{CO}_{st} = \mathrm{CO}_b\,(F_f + F_a)/F_{fl}$.

The stack oxygen analyzer sees λ_t. For a methane-like fuel burning with λ times stoichiometric air, the wet oxygen fraction is given by Equation (4.7).

$$y_{O_2} = \frac{200\,(\lambda_t - 1)}{1 + 9.52\,\lambda_t}\ \% \tag{4.7}$$

4.3.3 Dynamics

The fuel and air flow loops are represented by their closed-loop responses. The fuel path is a 2 s lag, and the air path is a 12 s lag behind a 2 s damper dead time with a stroking limit of 5 % of fan capacity per second, Equations (4.8) and (4.9).

$$\tau_f\frac{\mathrm{d}F_f}{\mathrm{d}t} = F_f^{SP} - F_f \tag{4.8}$$

$$\frac{\mathrm{d}F_a}{\mathrm{d}t} = \mathrm{clamp}\left(\frac{F_a^{SP}(t-\theta_a) - F_a}{\tau_a},\ \pm 0.05\,F_{a,max}\right) \tag{4.9}$$

The heat reaches the process and the bridgewall through a firebox and refractory lag, and the outlet temperature follows an energy balance on the coil, Equations (4.10) and (4.11).

$$\tau_{fb}\frac{\mathrm{d}Q_f}{\mathrm{d}t} = Q_{abs} - Q_f, \qquad \tau_{fb}\frac{\mathrm{d}T_{bw}}{\mathrm{d}t} = T_{bw}^{*} - T_{bw} \tag{4.10}$$

$$C_{th}\frac{\mathrm{d}T_{out}}{\mathrm{d}t} = Q_f - \dot{m}\,c_p\,(T_{out} - T_{in}) \tag{4.11}$$

Table 4.2 Model Parameters, With Hidden Parameters Not Exposed in the Interface

Symbol	Meaning	Value	Class
LHV_0, AFR_0	design fuel gas	47.0 MJ/kg, 17.0	visible
Q_{max}	fired duty at 100 %	50 MW	visible
$\dot{m}_0$, c_p	process flow, heat capacity	100 kg/s, 2.8	visible
τ_p, θ_p	coil lag, outlet delay	90, 15 s	visible
τ_{fb}	firebox and refractory lag	40 s	visible
f_r	radiant fraction	0.60	visible
$c_{p,g}$	flue gas heat capacity	1.15 kJ/(kg K)	visible
ΔT_{ap}	stack approach	100 K	visible
ε_c	casing loss	1.5 %	visible
T_{amb}	ambient	20 °C	visible
λ_k, w	CO knee, width	1.04, 0.025	visible
CO_0	CO floor	30 ppm	visible
k_{CO}	heat lost per ppm CO	4.5×10^{-6}	hidden
τ_f	fuel loop response	2 s	visible
τ_a, θ_a	air loop response, dead time	12, 2 s	visible
—	fan stroking rate	5 %/s	visible
$F_{a,max}$	fan capacity	25 kg/s	visible
θ_{an}, τ_{an}	analyzer delay, lag	12, 6 s	visible
D_0	stack draft	360 Pa	visible
$P_{c,0}$, $P_{d,0}$	convection, damper loss	150, 125 Pa	visible
τ_D	draft response	2 s	visible
—	CO permit	200 ppm	visible

with $C_{th} = \tau_p \dot{m}_0 c_p$ and $T^*_{bw} = T_{amb} + Q_{rel}(1 - f_r)/(F_{fl} c_{p,g})$. The arch draft is the stack draft less the convection and damper losses, each proportional to the square of the flue gas flow ratio $r = F_{fl}/F_{fl,0}$, Equation (4.12).

$$\tau_D \frac{\mathrm{d}P_d}{\mathrm{d}t} = -D_0 + \left[P_{c,0} + P_{d,0}\, e^{4.5(0.5 - d/100)}\right] r^2 + P_w - P_d \tag{4.12}$$

The damper loss falls exponentially with opening, so a percent of stroke is worth the same fraction of the loss, 4.5 %, anywhere in its range. A plain butterfly characteristic with loss proportional to the inverse square of the open area, matched at 50 %, has 2.6 times the gain at 30 % as at 50 %, each taken at the flue gas flow that the opening holds at −25 Pa, against 1.5 times for the exponential characteristic. An earlier version of the model with the butterfly characteristic made the draft loop unstable at turndown. P_w is a colored wind disturbance.

The outlet thermocouple reads T_{out} after a 15 s delay with colored noise of 0.1 °C. The analyzers read after a 12 s sample-line delay and a 6 s lag. These two delays and the 2 s damper dead time are sampled at each controller execution, so a delayed value is held for up to one 0.5 s scan. The oxygen reading carries its bias, a drift that accumulates at the drift rate, and 0.03 % of noise. The fuel-rich flag is raised whenever $\lambda_b < 1$ and a positive-pressure flag whenever $P_d > 0$.

The parameters are listed in Table 4.2. Each scenario starts from a true steady state computed for its own load, inlet temperature, tramp air and trim setting, so no transient is left over from initialization.

4.3.4 Assumptions

The whole firebox is one burner air ratio. Real heaters have many burners, and the worst-adjusted burner sets the carbon monoxide.

Carbon monoxide is an exponential correlation above a knee rather than a kinetic model. The knee on a real heater moves with burner condition and must be found by test, which is why the economic optimum is approached with an override rather than computed [10].

The fuel is methane-like, with stoichiometric air proportional to heating value. Hydrogen-rich refinery fuel gas breaks that proportionality, because hydrogen needs less air per megajoule than methane.

The fuel and air flow loops are closed-loop lags. Valve characteristics are not modeled, and only the stack damper has a characteristic.

No burner management system is modeled. The fuel-rich flag marks a condition on which a real system would trip the heater, and the simulation continues so that the event can be studied.

Analyzer faults are a bias, a drift and noise. Cell failures, calibration checks and cross-sensitivity are outside the model.

4.4 Implemented Control System

The controller runs every 0.5 s and the physics is integrated by fourth-order Runge-Kutta at 0.1 s, on the three clocks of Section 1.6.1.

4.4.1 Outlet Temperature and Feedforward

The temperature controller TIC is the framework's velocity-form PID of Section 1.6.2, executed every $\Delta t = 0.5\,\mathrm{s}$ on the error $e_k = T_{SP} - T_{out,m}$, Equation (4.13).

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= D_{ff} + s_k + I_{k-1}, \qquad D_k = \mathrm{clamp}(c_k, 0, 100) \\ I_k &= I_{k-1} + \frac{K_c}{\tau_I}\, e_k \Delta t + \frac{\Delta t}{\tau_t}\,(D_k - c_k) \end{aligned} \tag{4.13}$$

where $\Delta u_k^P = K_c(e_k - e_{k-1})$ is the proportional increment and Δu_k^D the change in the derivative term. In the symbols of Equations (1.10) to (1.12), s_k is S_k, c_k is v_k, the applied firing demand D_k is u_k and the feedforward D_{ff} is the output bias u_b. The demand is clipped to 0 to 100 % of the 50 MW fired duty, and only the integral is corrected, by back-calculation with tracking time $\tau_t = \tau_I$. The derivative, off by default, acts on the measurement through a 5 s filter. The defaults are $K_c = 1.0\,\%/^{\circ}\mathrm{C}$ and $\tau_I = 110\,\mathrm{s}$, with 1200 s at the top of the slider meaning proportional only. The fuel setpoint clamp and the cross-limiting low selector act downstream of D_k and are not fed back, so the integral keeps integrating while the fuel waits for the air. The feedforward is the static inverse of the heat balance as a deviation from design, the static feedforward of Section 1.4.3, Equation (4.14).

$$D_{ff} = D_0 \left[\frac{(\dot m/\dot m_0)(T_{SP} - T_{in})}{360 - 253} - 1 \right] \tag{4.14}$$

with $D_0 = 73.3\,\%$ the design demand. It enters as the output bias, and switching it on or off moves the integral by the same amount so the demand does not bump.

4.4.2 Three Fuel and Air Schemes

The demand is converted to a fuel mass rate $F_{dem} = D\,Q_{max}/(100\,\mathrm{LHV}_0)$ at the design heating value, because the controller does not know the actual one. The ratio applied is $R = \mathrm{AFR}_0\,\lambda_0\,(1 + \rho/100)$, with $\lambda_0 = 1.15$ and the trim ρ. The three schemes are given by Equations (4.15) to (4.17).

$$\text{fuel leads:} \quad F_f^{SP} = F_{dem}, \quad F_a^{SP} = R\,F_{f,m} \tag{4.15}$$

$$\text{air leads:} \quad F_a^{SP} = R\,F_{dem}, \quad F_f^{SP} = F_{a,m}/R \tag{4.16}$$

$$\begin{aligned} \text{cross-limited:} \quad F_f^{SP} &= \min[F_{dem},\,(1+\delta)F_{a,m}/R] \\ F_a^{SP} &= \max\left[R\,F_{dem},\,(1-\delta)R\,F_{f,m}\right] \end{aligned} \tag{4.17}$$

where the subscript m marks a measured flow and $\delta = 2\,\%$ is the margin by which fuel may run ahead of air. The cross-limited scheme is Equation (1.8) written in kilograms, with the trimmed ratio R in

place of λ_0. The fuel setpoint is clamped to 12 to 100 % of its maximum and the air setpoint to 20 to 100 % of fan capacity.

🕮 **Pause.** Try Exercise 4.1 before reading on.

4.4.3 Oxygen Trim and the Carbon Monoxide Override

The trim multiplies the ratio, as Section 1.4.4 recommends. It is a high select between an oxygen PI controller and a carbon monoxide PI controller, written in external reset form [14]. Each controller's integral is a first-order filter of the trim actually applied, with its own integral time, Equations (4.18) to (4.20).

$$F_{O_2} \leftarrow F_{O_2} + \frac{\Delta t}{\tau_{I,O}} \left(\rho_{k-1} - F_{O_2}\right)$$
$$\rho_{O_2} = K_{c,O}\, e_{O_2} + F_{O_2} \tag{4.18}$$
$$F_{CO} \leftarrow F_{CO} + \frac{\Delta t}{\tau_{I,CO}} \left(\rho_{k-1} - F_{CO}\right)$$
$$\rho_{CO} = K_{c,CO} \ln \frac{\mathrm{CO}_m}{\mathrm{CO}_{SP}} + F_{CO} \tag{4.19}$$
$$\rho_k = \mathrm{clamp}[\max(\rho_{O_2}, \rho_{CO}), \pm\rho_{max}] \tag{4.20}$$

where $e_{O_2} = \mathrm{O}_{2,SP} - \mathrm{O}_{2,m}$. The selected controller behaves as an ordinary PI, the unselected one sits below it by the difference in proportional terms, and neither winds up because both filter the clamped trim that was applied. This form replaced a version built from two tracking PID objects that deadlocked with a steady oxygen error, because tracking wiped the integral increment of whichever controller had been selected on the previous scan. The override acts on the logarithm of carbon monoxide because carbon monoxide is exponential in the air ratio, Equation (4.3), so one e-fold of carbon monoxide is always about the same change in air ratio and the loop gain is nearly constant. The defaults are $K_{c,O} = 2.0$ percent of ratio per percent oxygen, $\tau_{I,O} = 60\,\mathrm{s}$, $K_{c,CO} = 2.0$ percent of ratio per e-fold, $\tau_{I,CO} = 60\,\mathrm{s}$, an override setpoint of 150 ppm below the 200 ppm permit, and a trim authority $\rho_{max} = 8\,\%$. Because the selection is a maximum, carbon monoxide can only ever add air. In normal operation the carbon monoxide controller is the unselected one and waits below the oxygen trim, much as the protective override of Volume 4, Chapter 2 rests on its clamp through healthy operation.

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

4.4.4 Draft

The draft controller PIC is a reverse-acting PI on the arch pressure, opening the damper when the arch rises toward atmospheric, with $K_c = 0.4\,\%/\mathrm{Pa}$, $\tau_I = 3\,\mathrm{s}$, a setpoint of −25 Pa and an output range of 5 to 100 %, on the same algorithm as the TIC with $\tau_t = \tau_I$. The characterized damper holds the process gain between 12.6 and 5.4 Pa per percent of stroke from 60 to 120 % of design flue gas flow, where a butterfly would span 28 to 3.9 Pa per percent.

4.5 Results

Table 4.3 lists the scenarios and Table 4.5 the measured outcomes, each read from the app at its default seed (Section 1.6.1). Steady-state values are ten-minute averages of the logged run, and peaks, crossing times and fuel-rich times are taken at the 0.1 s physics step. The exported file of a run holds both air ratios, the analyzer readings beside their true values and the selected trim controller (Section 1.6.4). Table 4.4 gives the requirement that completes each scenario. The comparisons in cases 3, 5, 7 and 8 start from Reset, which keeps the reader's settings, and a setting changed before Run gives the same run as that setting carried through Reset.

Table 4.3 Scenario Sequence

#	Scenario	Teaching point
1	Fire it by hand	the thermal lag
2	Fuel leads	a load step goes rich
3	Cross-limited	the safe direction
4	Trim toward the peak	an optimum on a constraint
5	A fuel-gas swing	an unmeasured disturbance
6	Drift, wide-open trim	trusting an analyzer
7	Drift, clamped, overruled	bounding the trim
8	Tramp air	two air ratios
9	Draft in the wind	margin to atmosphere
10	Challenge	graded run

Table 4.4 Scenarios and Completion Requirements

#	Scenario	Kind	Required to complete
1	Fire it by hand	Experiment	In Manual, hold the firing demand at 73.8 % for 1 min, step it to 78.8 % and hold it for 5 min, then return the outlet to 360 °C by hand and hold it within ±0.5 °C for 1 min.
2	Fuel leads	Demonstration	Run with fuel leading through the +20 % process flow step at 60 s until the burners go fuel-rich ($\lambda_b < 1$), and on to 3 min.
3	Cross-limited	Experiment	Run 4 min cross-limited through the same step, Reset and run 4 min with air leading, then step the process flow down 20 % and follow the fuel and air for 30 s.
4	Trim toward the peak	Experiment	With the override off, hold the oxygen setpoint at 2.0 % for 3 min and then at 1.0 % or below for 3 min, switch the CO override on with the setpoint below 1.5 %, and keep the override in control for 5 min.
5	A fuel-gas swing	Experiment	Run 10 min through the +6 % heating value step at 60 s with the oxygen trim on, then Reset and run 5 min with the trim off.
6	Drift, wide-open trim	Demonstration	Run with the analyzer drifting 0.1 % oxygen per minute from 60 s until the reading exceeds the true oxygen by 1 %, and on to 30 min.
7	Drift, clamped, overruled	Experiment	With the override off, run until the trim reaches its −8 % clamp and on to 30 min, then Reset and run 30 min with the CO override on.
8	Tramp air	Experiment	With the sight doors opened at 60 s, keep the CO override in control for 5 min, then Reset and run 15 min with the override off.
9	Draft in the wind	Experiment	In 30 Pa gusts, run 1 min at −25 Pa, then hold the arch draft setpoint at −10 Pa for 10 min, at −5 Pa for 5 min and at −40 Pa for 5 min.
10	Challenge	Challenge	Through the load step at 2 min, the fuel-gas swing at 10 min and the analyzer drift from 15 min: never fuel-rich, CO never above 200 ppm, a run-average efficiency of at least 80.5 %, and the outlet within ±1 °C over the final 10 min, judged at 30 min with the model parameters and disturbances as given and no extra disturbance button pressed.[a]

[a] With the shipped settings (fuel leads, no override, ±25 % trim) the run fails, with 3.7 s fuel-rich and 118 s over the permit, although its 80.70 % efficiency and 0.22 °C worst outlet error meet their criteria. With the suggested settings (cross-limited, override on, ±8 % trim) it passes with no time rich or over the permit, a 173 ppm peak 27 ppm under the permit, 80.70 % efficiency 0.20 points above the target and a 0.21 °C worst outlet error.

4.5.1 A Load Step in the Firebox

The thermal side is slow. A 5 % firing step in manual gives 6.9 °C at the outlet once settled, about 1.4 °C per percent, of which 6.5 °C has arrived after five minutes. The outlet barely moves for 20 s, and 63 % of the change has arrived after about 143 s. The 2 s fuel valve is faster than that chain by a factor of about seventy.

The combustion side is fast, and the scheme decides what happens in it. With the air slaved to the measured fuel, a 20 % process flow step with the feedforward on sends the fuel ahead of the fan. Burner λ falls to 0.990, the firebox is fuel-rich for 3.7 s and carbon monoxide peaks near 7300 ppm.

Table 4.5 Measured Results

Case and condition	Result
1, 5 % firing step	6.9 °C, 63 % after 143 s
2, fuel leads, +20 % load	λ_b 0.990, 3.7 s rich, CO 7270 ppm
3, cross-limited, same step	$\lambda_b \geq 1.136$, CO 50 ppm
2 and 3, outlet dip	4.6 against 5.8 °C
3, air leads, outlet dip	6.3 °C
4, O_2 3.0 / 2.5 / 2.0 %	η 81.16 / 81.73 / 82.23 %
4, O_2 2.0 %	CO 210 ppm
4, O_2 1.2–1.5 %	η 82.51–82.55 %, CO $\geq$ 660 ppm
4, override on	O_2 2.15 %, CO 150 ppm, η 82.09 %
5, LHV +6 %, no trim	λ_b 1.085, CO 190 ppm
5, LHV +6 %, trim	trim +6.0 %, λ_b 1.150
6, drift, ±25 %	over permit at 12 min, rich at 26 min
6, end of run	true O_2 0.3, reading 3.2 %, 187 s rich
7, drift, ±8 % clamp	λ_b 1.055, CO 500 ppm
7, clamp and override	CO 150 ppm, reading 5.1 %
8, tramp air, override off	O_2 3.0 %, λ_b 1.061, CO 420 ppm
8, tramp air, override on	O_2 3.4 %, λ_b 1.090, CO 150 ppm
9, 30 Pa gusts, SP −25 / −10 / −5 Pa	positive 0 / 2 / 86 s

Cross-limited, the same step never takes burner λ below 1.136 and carbon monoxide peaks at 50 ppm. The price is paid in temperature, because the heat now waits for the fan: the outlet dips 5.8 °C against 4.6 °C for fuel leads (Table 4.5), and Figure 4.2 shows the heater 65 s after the step. Air leads is safe as well, with a 6.3 °C dip, but on a 20 % load decrease, pressed at 240 s to return the flow to design, it makes the fuel wait for the fan too. Over the next ten minutes its temperature error integrates to 1725 °C s, about twice the 810 and 898 °C s of fuel leads and cross-limiting. Cross-limiting behaves as air leads on the way up and as fuel leads on the way down, which is the rule stated as two selectors.

> **TRY IT IN THE APP** **2 · Fuel Leads, and a Load Step Goes Rich**
>
> Load the scenario and press Run. At the 60 s marker the process flow rises 20 % and the fuel moves before the air. Watch the burner line of the Burner air ratio λ strip dip below one, then read the Time fuel-rich and Peak CO tiles: about 3.7 s rich and a peak near 7300 ppm. Scenario 3 runs the same step cross-limited.

4.5.2 The Optimum and the Permit

Walking the oxygen setpoint down with the override off, the steady-state efficiency, averaged over ten minutes, is 81.16 % at 3.0 % oxygen, 81.73 % at 2.5 and 82.23 % at 2.0, where carbon monoxide is already 210 ppm, over the permit. The efficiency peak, 82.55 % at 1.37 % oxygen with carbon monoxide near 890 ppm, is flat within 0.04 points between 1.2 and 1.5 % oxygen, where carbon monoxide is 660 ppm and above, so the true economic optimum lies on the carbon monoxide constraint and not at the efficiency maximum. With the override on and the oxygen setpoint pushed to 1.5 % or below, the heater settles at 2.15 % oxygen, 150 ppm and 82.09 %, burning about 32 kg/h less fuel than at 3 %. The operating point is found by the constraint, and it moves when the constraint moves, which a fixed oxygen setpoint cannot do.

4.5.3 An Analyzer That Cannot Be Trusted

A heating value increase of 6 % is invisible to every measurement except the temperature and the oxygen. Without trim the burner ratio settles at 1.085 with carbon monoxide near 190 ppm, close enough to the permit that thermocouple noise, passed to the fuel by the temperature loop, carries it over for 70 s in the first ten minutes. With trim the trim rises about 6.0 % and restores $\lambda_b = 1.150$. Either way the transient dips λ_b to about 1.08 before the analyzer and trim respond. Volume 11, Chapter 14 follows such swings to their source, a refinery fuel gas header, and feeds the Wobbe index of the gas forward to the burners.

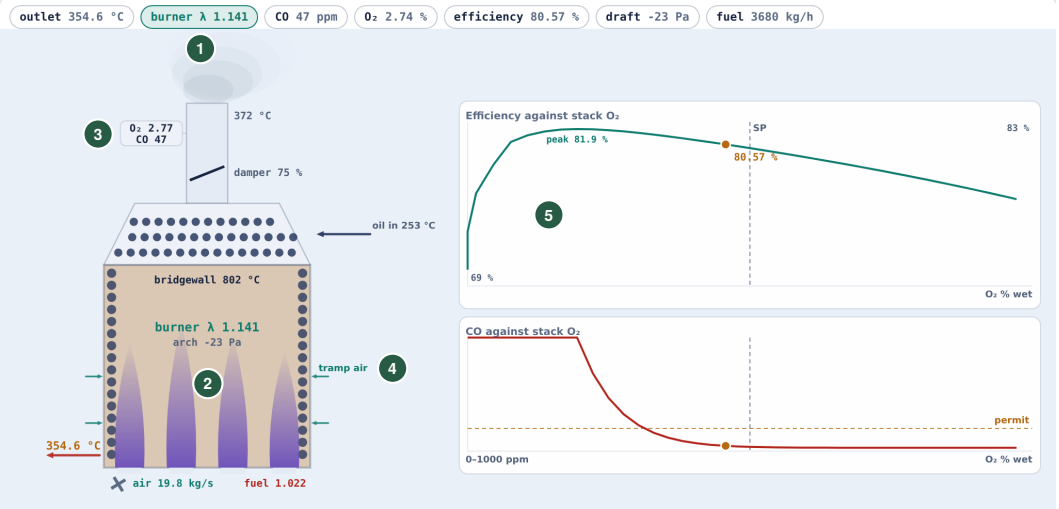


Figure 4.2 The fired heater running scenario 3 with the fuel and air cross-limited, 65 s after a 20 % process flow increase at 60 s, cropped to the plant view. (1) The row of readings, with the outlet at 354.6 °C, the burner air ratio at 1.141 and carbon monoxide at 47 ppm. (2) The four burners, with flames colored by the burner air ratio, between the wall coils. (3) The stack with its damper and the analyzer's oxygen and carbon monoxide readings. (4) Tramp air entering at the sight doors. (5) Efficiency against stack oxygen at the present conditions, with the operating point, the oxygen setpoint and the efficiency peak left of the setpoint, above carbon monoxide against stack oxygen, which rises steeply below the setpoint and crosses its permit.

A trim that follows its analyzer follows its analyzer's faults. With the authority at 25 % and no override, an oxygen reading drifting 0.1 % per minute takes carbon monoxide past the permit about 12 min after the drift begins, and the burners first go rich about 26 min after it begins. By the end of the half hour the true oxygen is 0.3 % against a reading of 3.2 %, and the heater has spent 187 s fuel-rich. Clamping the trim at ±8 % bounds the damage: the trim reaches the clamp about 17 min in, burner λ bottoms at 1.055 and carbon monoxide settles near 500 ppm. The override removes it. Carbon monoxide takes the trim about 11 min in and holds near its 150 ppm setpoint, at 145 ppm after 30 min, while the oxygen reading climbs to 5.1 %. A drifting analyzer takes the Claus plant of Volume 11, Chapter 13 outside its permit in the same way, with its air demand reading on target.

Tramp air produces a different signature from the same instruments. With the sight doors opened and the override off, the stack oxygen settles at 3.0 %, its setpoint, while the burner ratio falls to 1.061 and carbon monoxide settles near 420 ppm, because the trim has removed burner air to make up for the air leaking in. With the override on, carbon monoxide holds at 150 ppm, λ_b at 1.090 and the stack oxygen reads 3.4 %, each averaged over the last ten minutes of a half hour. High oxygen, normal carbon monoxide and an active override are the combination that identifies a leak, which is how field troubleshooting describes it [11].

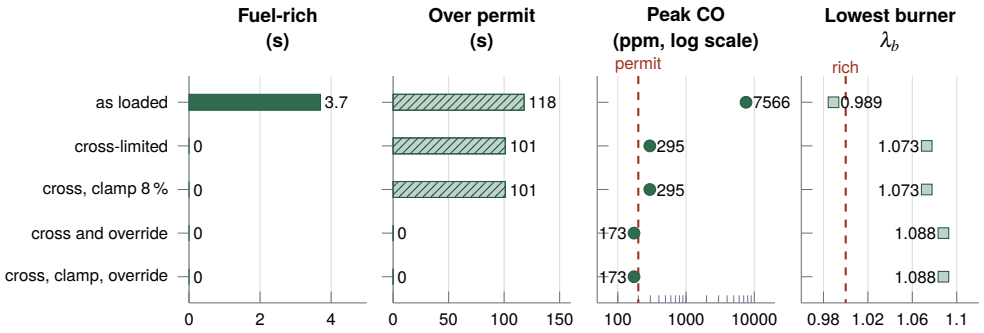


Figure 4.3 Table 4.6 redrawn: the five challenge configurations over 30 min, with time fuel-rich and time over the permit in bars, and peak carbon monoxide (round dots, logarithmic scale) and lowest burner air ratio (square dots) against the dashed permit and stoichiometric lines. Notice that cross-limiting alone ends the fuel-rich time but leaves carbon monoxide over the permit, and only the rows with the override come to rest left of the permit line.

Table 4.6 Challenge Outcomes Over 30 min

Configuration	Rich (s)	Over permit (s)	Peak CO	min λ_b
as loaded	3.7	118	7566	0.989
cross-limited	0	101	295	1.073
cross, clamp 8 %	0	101	295	1.073
cross and override	0	0	173	1.088
cross, clamp, override	0	0	173	1.088

TRY IT IN THE APP **8 · Tramp Air the Burners Never See**

Run the scenario as loaded for half an hour. The sight doors open at 60 s, the burner and stack lines part on the λ strip, and the CO override holds carbon monoxide near 150 ppm. Then Reset, switch CO override off and run again. The stack oxygen sits on its 3.0 % setpoint while carbon monoxide settles near 420 ppm.

With 30 Pa wind gusts the arch at a −25 Pa setpoint peaks at −10.5 Pa in fifteen minutes. At −10 Pa it goes positive for about 2 s, and at −5 Pa for about 86 s, counted at the physics step. Five other noise seeds give 4 to 12 s and 71 to 97 s, so the ordering is robust and the counts are not.

DID YOU KNOW? **The Last Few Pascals of Draft**

From the simulator: in 30 Pa gusts the arch spent no time above atmospheric with its draft setpoint at −25 Pa, about 2 s at −10 Pa and about 86 s at −5 Pa (Table 4.5). Other noise seeds change the counts but not their order.

4.5.4 The Graded Challenge

The challenge runs 30 min with a 20 % load step at 2 min, a 6 % fuel-gas swing at 10 min and an analyzer drifting from 15 min, starting with fuel leading, no override and a 25 % trim. It is met with no time fuel-rich, carbon monoxide never over the permit, a run-average efficiency of at least 80.5 % and the outlet within 1 °C for the last ten minutes. Table 4.6 reports the measured outcomes, and Figure 4.3 draws them.

Cross-limiting removes the fuel-rich excursion but not the drift. The clamp alone changes nothing within the half hour, because the drift drives the trim down from the +5.3 % the fuel-gas swing required, and it has reached only −0.3 % when carbon monoxide crosses the permit at 28 min and

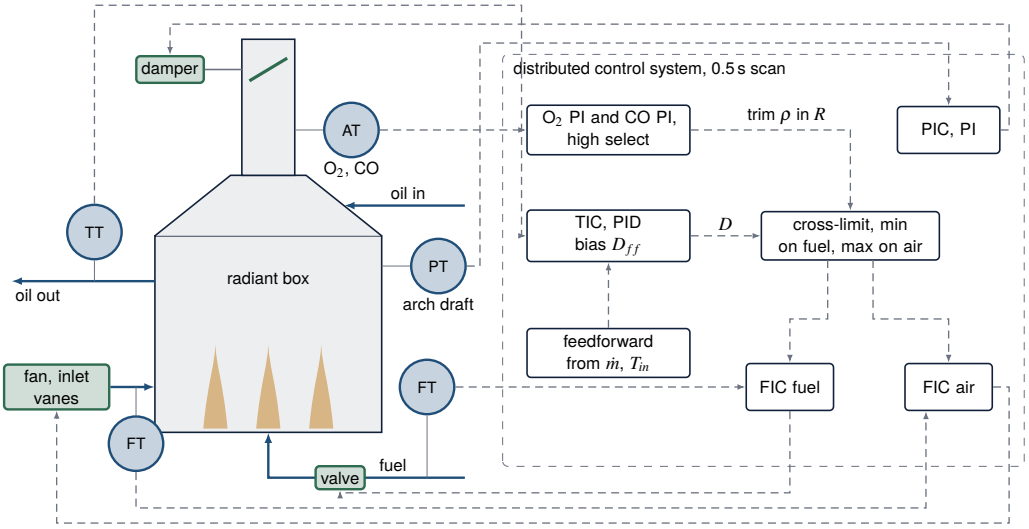


Figure 4.4 The combustion loops on real hardware. The thermocouple TT feeds TIC, whose output bias carries the feedforward from the process flow and inlet temperature. Its demand D enters the cross-limiting selectors, which also read the measured flows and set the fuel and air flow loops on the fuel gas valve and on the fan's inlet vanes. The analyzer AT drives the trim, a high select between the oxygen and carbon monoxide controllers, and PIC holds the arch draft with the stack damper. Solid arrows are process streams and dashed gray arrows are signals. Notice that the trim and the selectors only move setpoints of the two flow loops, so on a load increase the slow fan sets the pace.

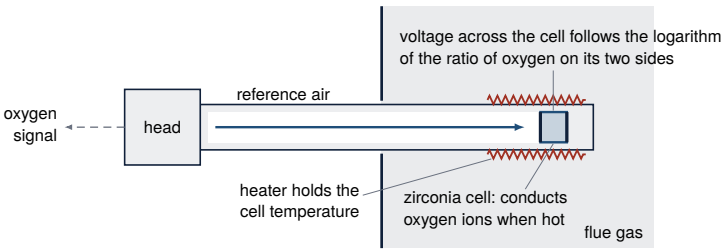


Figure 4.5 The zirconia oxygen cell. Notice that the voltage compares the oxygen at the probe with the reference air. Nothing in the cell knows whether that oxygen passed through a burner, so tramp air reads as excess air, the lesson of scenario 8.

−1.3 % at the end of the run, far inside the ±8 % limit. Only the override keeps carbon monoxide legal, with a peak of 173 ppm, and the clamp then never binds, so the suggested configuration of the last row runs identically to cross-limiting with the override alone. The run average efficiency is 80.70 to 80.71 % in every configuration, and the outlet stays within 0.22 °C over the last ten minutes in every configuration, so the challenge lies entirely in the firebox. Over five other noise seeds the run as loaded spends 3.3 to 3.7 s fuel-rich and 108 to 117 s over the permit, and the suggested configuration never goes rich or over the permit.

🕮 **Pause.** Try Exercise 4.3 before reading on.

4.6 Technology in Practice

Figure 4.4 draws the loops of Section 4.4 with their hardware.

What Measures It

Each flow is measured by its own transmitter, because a flow loop on one line and a transmitter on the other is what turned a mechanical ratio into one a controller can compute [7]. The model hides both meters inside the closed-loop lags of Equations (4.8) and (4.9). The outlet thermocouple TT reads after 15 s. The stack gas is read by an analyzer for oxygen and carbon monoxide, and the oxygen cell is drawn in Figure 4.5. The 12 s delay and 6 s lag of AT stand for the sample line and the cell, and its bias and drift stand for a cell between calibrations.

> **TECHNOLOGY** **Sensor: A Zirconia Cell at the End of a Sample Line**
>
> A common flue gas oxygen detector is a sandwich of platinum electrodes on a solid zirconium oxide electrolyte, held hot, with the flue gas on one side and reference air on the other. Its voltage follows the logarithm of the ratio of the two oxygen activities, and a microcomputer inverts that function [7]. An oxygen analyzer for flue gas spans zero to the oxygen of air and is calibrated against air at 20.9 % and against nitrogen at zero. An extractive system adds a probe and a heat-traced sample line, and analyzer systems need much more upkeep than other instruments [7]. A slow drift looks on a trend like a real change in excess air.

What Moves It

The firing demand reaches the burners through the fuel gas control valve, which the model treats as a 2 s lag. The air path is the slow one.

> **TECHNOLOGY** **Final Element: A Forced-Draft Fan and Its Inlet Vanes**
>
> A forced-draft fan pushes the combustion air into the furnace [13]. Its flow is commonly set by inlet vanes, which swirl the entering air in the direction of the impeller and so lower the fan's output, or by dampers, which add backpressure and move the fan along its curve. A fan throttled far back can run unstably, its airflow surging back and forth. Vanes and dampers move through mechanical linkages, and dirt built up on a linkage impairs its travel [15]. The model's 2 s dead time, 12 s lag and stroking limit stand for that actuator and the fan. Under cross-limiting, air that cannot rise holds the fuel back with it, so a sticking linkage shows as an outlet temperature that sags.

The stack damper holds the draft. A damper is a multi-element device that throttles large flows of air at low pressure, built with parallel or radial vanes, and furnace and boiler draft control is one of its common uses [7]. The model gives it the exponential characteristic of Equation (4.12).

Where the Controller Runs

Control loops in a plant are often nested or cascaded, with cycle times from milliseconds to minutes, and a distributed control system uses a supervisory loop to mediate a group of localized controllers [16]. In the app every loop runs on a 0.5 s scan, against a fuel lag of 2 s and a coil lag of 90 s. On such a platform TIC is a PID function block whose output bias carries the feedforward. The ratio is a multiplying block, cross-limiting is a low and a high selector on the two flow setpoints, and the trim is a high select over two PI blocks in external reset form. The burner management system is a separate safety system that watches for low or high fuel pressure and loss of flame and shuts off the fuel [7].

> **IN PRACTICE** **The Annual Tune-Up: Burner, Ratio, Then Carbon Monoxide**
>
> A United States emission rule requires an annual tune-up of each boiler or process heater of 10 million Btu per hour or more. The burner is inspected and cleaned, the flame pattern adjusted and the air-to-fuel ratio control checked for calibration. Carbon monoxide is then optimized, and carbon monoxide in parts per million and oxygen in percent are measured before and after the adjustments. The annual frequency does not apply to units with a continuous oxygen trim that holds an optimum air-to-fuel ratio [17].

What Is Modernizing It

The first development is the fan drive. Fan speed adjustment is in general the most efficient way to control airflow, and in dirty air an adjustable-speed drive avoids the fouled linkages of vanes and dampers [15]. The second is where the gas is read. A single-beam infrared analyzer of the gas filter correlation kind suits carbon monoxide among carbon dioxide and water vapor, and it can pass its beam across the diameter of the stack instead of through a sample cell [7]. A path across the stack has no sample line to wait for. The third is the fuel itself. A federal roadmap counts large fired heaters among the larger greenhouse gas sources of a refinery and judges that the most cost-effective capture would likely be a central reformer with carbon capture supplying hydrogen to the process fired heaters [18], the subject of the open challenge below.

4.7 Limitations

The single burner air ratio hides burner-to-burner variation, which on a real heater makes the carbon monoxide knee both earlier and less sharp.

The carbon monoxide correlation is not kinetic, and its knee has been placed so that the efficiency peak falls beyond the permit. That placement is typical but not universal, and on a heater with a sharper knee the optimum and the constraint can coincide.

Tramp air enters at one point with a square-root law. Real leaks are distributed and some enter above the burners, where they contribute to the convection section but not to combustion.

The fuel and air flow loops are lags, so valve nonlinearity, fan surge and air density changes with ambient temperature are absent [12].

There is no burner management system, no flame supervision and no startup or shutdown sequence. The fuel-rich flag marks the condition on which such a system would trip.

> ⚠ **OPEN CHALLENGE** **Firing Refinery Heaters on Hydrogen**
>
> Hydrogen burned in air gives a flame near 2100 °C in special burners. For refining, a federal roadmap calls for the capability to burn hydrogen for high-temperature process heat and lists the needs that remain: applicability at furnaces with higher hydrogen in blends, temperature control and management, embrittlement, unified standards for retrofits and moisture after combustion [18]. The ratio of this chapter is set for a methane-like fuel (Section 4.3), and the constraint its override holds already moves with burner condition, fuel and load. Progress would be heaters run on rising hydrogen fractions with their ratio, trim and override shown to hold.

4.8 Conclusion

A fired heater teaches that the safe structure and the economic structure are made of the same parts. In the simulator of this chapter, a 20 % load step with the air following the fuel takes the burners to $\lambda = 0.990$ for 3.7 s with carbon monoxide near 7300 ppm, and the same step cross-limited never goes below 1.136, at a cost of 5.8 °C of outlet dip against 4.6 °C. The efficiency optimum lies beyond the carbon monoxide permit, so the best operating point is the permit itself, reached by a logarithmic carbon monoxide override at 2.15 % oxygen and 82.09 % efficiency against 81.16 % at the conventional 3 %. The same override detects an analyzer drifting 0.1 % per minute that would otherwise make the firebox fuel-rich within half an hour, and it separates tramp air from a real shortage of air. The model uses one air ratio for the whole firebox, a correlated rather than kinetic carbon monoxide, and closed-loop lags for the flow loops. A fuller model would add multiple burners, hydrogen-rich fuel gas, air preheat and a burner management sequence. Within those limits the heater shows how control structures are built from selectors.

KEY POINTS

- On a burner, too little air is a hazard and too much is only a cost, and the fan is much slower than the fuel valve.
- With the air following the fuel, a 20 % load step takes the burners to $\lambda = 0.990$. Cross-limited, the same step never goes below 1.136, at a cost of 5.8 °C of outlet dip against 4.6 °C.
- The efficiency optimum lies beyond the carbon monoxide permit, so the best operating point is the permit itself, held by a logarithmic override.
- High oxygen, normal carbon monoxide and an active override are the combination that identifies a leak.

Where else this applies. Volume 4, Chapter 2 and Volume 4, Chapter 6 hold protective overrides on a limit with a hazardous side, Volume 9, Chapter 8 holds an optimum on a constraint that moves, and Volume 15, Chapter 11 cross-limits coal and air on a kiln.

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] F. G. Shinskey and B. G. Lipták, “Ratio Control,” in *Process Control*. Elsevier, 1995, pp. 102–105.

[3] G. F. J. Gilman, *Boiler Control Systems Engineering*. Wiley, 2026.

[4] D. Lindsley, *Power Plant Control and Instrumentation: the control of boilers and HRSG systems*. Institution of Engineering and Technology, 2000.

[5] D. Krishnamoorthy and S. Skogestad, “Systematic design of active constraint switching using selectors,” *Computers & Chemical Engineering*, vol. 143, p. 107106, 2020.

[6] A. Reyes-Lúa and S. Skogestad, “Systematic Design of Active Constraint Switching Using Classical Advanced Control Structures,” *Industrial & Engineering Chemistry Research*, vol. 59, no. 6, pp. 2229–2241, 2019.

[7] 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.

[8] J. Weissbart and R. Ruka, “Oxygen Gauge,” *Review of Scientific Instruments*, vol. 32, no. 5, pp. 593–595, 1961.

[9] J. J. Sweeney, “Boiler Flue Gas Analysis and Automatic Combustion Air Trim Control by Measurement of Carbon Monoxide, Unburned Hydrocarbons, and Opacity,” *Marine Technology and SNAME News*, vol. 21, no. 01, pp. 74–79, 1984.

[10] C. E. Baukal, Jr., *Heat Transfer in Industrial Combustion*. CRC Press, 2000.

[11] N. P. Lieberman, *Troubleshooting Process Plant Control*. Wiley, 2008.

[12] W. Bussman and C. Baukal, “Ambient condition effects on process heater efficiency,” *Energy*, vol. 34, no. 10, pp. 1624–1635, 2009.

[13] U.S. Department of Energy, Advanced Manufacturing Office, *Improving Process Heating System Performance: A Sourcebook for Industry*, 3rd ed., U.S. Department of Energy and Industrial Heating Equipment Association, Washington, DC, 2015.

[14] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[15] U.S. Department of Energy, *Improving Fan System Performance: A Sourcebook for Industry*, U.S. Department of Energy and Air Movement and Control Association International, Washington, DC, 2003, dOE/GO-102003-1294.

[16] 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.

[17] U.S. Environmental Protection Agency, “40 CFR 63.7540, how do I demonstrate continuous compliance with the emission limitations, fuel specifications and work practice standards?” Code of Federal Regulations, title 40, volume 15, 2025 edition, U.S. Government Publishing Office, 2025.

[18] U.S. Department of Energy, “Industrial decarbonization roadmap,” U.S. Department of Energy, Washington, DC, Tech. Rep. DOE/EE-2635, 2022.

4.9 Exercises

4.1 Concept check · 2 min

The fuel and air are cross-limited, and the process flow steps up by 20 % with the feedforward on. Which quantity sets the fuel flow setpoint in the seconds after the step?

(a) The firing demand of the TIC, as a fuel rate at the design heating value
(b) The measured air flow divided by the ratio, with the margin δ added
(c) The measured fuel flow multiplied by the ratio, with the margin δ taken off
(d) The stack oxygen reading, acting through the trim on the air-to-fuel ratio

4.2 Concept check · 2 min

Why does the carbon monoxide override act on $\ln(\mathrm{CO}_m/\mathrm{CO}_{SP})$ rather than on the difference $\mathrm{CO}_m - \mathrm{CO}_{SP}$?

(a) The logarithm filters the analyzer noise that the sample line adds
(b) The permit is written as a logarithmic limit on the stack concentration
(c) CO is exponential in λ_b, so an e-fold of CO is a fixed change in λ_b
(d) A logarithm turns negative below the setpoint and lets CO remove burner air

4.3 Concept check · 1 min

In the challenge, cross-limited without the override, a trim authority of ± 8 % leaves 101 s over the permit, the same as ± 25 % (Table 4.6). Why does the clamp change nothing?

(a) The clamp limits the fuel setpoint, so it cannot act on the air
(b) The margin δ of the selectors cancels any limit set on the trim
(c) The temperature loop restores the air that the clamp holds back
(d) The drift has carried the trim only to about -1.3 % by 30 min

4.4 Theory and calculation · 6 min

At the design point of Table 4.1 the heater burns $F_f = 0.785\,\mathrm{kg/s}$ with $F_a = 15.36\,\mathrm{kg/s}$, AFR $= 17.0$, and a tramp leak of 3 % at −25 Pa.

(a) Find the stoichiometric air, λ_b, the tramp air of Equation (4.2), λ_t and the wet stack oxygen of Equation (4.7). *Checkpoint:* 13.35 kg/s, 1.15, 0.45 kg/s, 1.185 and 3.0 %.

(b) The sight doors open and the leak rises to 11 % at −25 Pa. The trim holds the stack oxygen at 3.0 %. With F_f kept at 0.785 kg/s, find λ_b. *Checkpoint:* 1.060.

(c) Find the carbon monoxide at the burners and at the stack for that λ_b. *Checkpoint:* About 470 ppm and 425 ppm.

4.5 Theory and calculation · 6 min

The fuel gas heating value rises by 6 %, and AFR rises with it as in Section 4.3. The O_2 trim is off with $\rho = 0$, so the controller applies $R = \mathrm{AFR}_0\lambda_0$, and the TIC returns the outlet to its setpoint.

(a) Find the burner air ratio once the heater has settled. *Checkpoint:* 1.085.

(b) Find the trim ρ that restores $\lambda_b = 1.150$, and say whether the default authority allows it. *Checkpoint:* +6.0 %, inside the ± 8 % authority.

(c) With the trim off, estimate the carbon monoxide at the burners and at the stack and the stack oxygen. Take the tramp air as 0.45 kg/s, the stoichiometric air as 13.35 kg/s and the fuel as $0.785/1.06$ kg/s. *Checkpoint:* About 196 and 190 ppm, and 2.0 %.

4.6 Theory and calculation · 6 min

Scenario 3 starts with the firing demand at 73.8 %, the air at 15.36 kg/s and $R = 19.55\,\mathrm{kg/kg}$. At 60 s the process flow steps to 120 % with $T_{SP} = 360\,^{\circ}\mathrm{C}$ and $T_{in} = 253\,^{\circ}\mathrm{C}$.

(a) Find the change in D_{ff} from Equation (4.14), the demand just after the step, and the fuel rate F_{dem} it asks for. *Checkpoint:* 14.66 %, 88.5 % and 0.941 kg/s.

(b) Find the cross-limited fuel and air setpoints just after the step. *Checkpoint:* 0.801 kg/s and 18.4 kg/s.

(c) With fuel leading instead, suppose the fuel reached 0.941 kg/s before the air had moved at all. Find λ_b. *Checkpoint:* 0.960.

4.7 App challenge · 9 min

Load scenario 2, Fuel leads, and a load step goes rich. Change no setting, run to 3 min at the default seed, read the Time fuel-rich, Peak CO and Lowest burner λ tiles, and export the CSV file. Load scenario 3, The same step, cross-limited, run to 4 min, read the same tiles and export. Then set Fuel and air to Air leads, press Reset, run to 4 min and export again.

(a) Report the three tiles for scenarios 2 and 3. *Checkpoint:* Scenario 2: 3.7 s, 7268 ppm, 0.990. Scenario 3: 0.0 s, 50 ppm, 1.136. Peak CO within 20 ppm.

(b) In each of the three files find the lowest `tout_C` after 60 s and the dip below 360 °C. *Checkpoint:* Fuel leads 4.6, cross-limited 5.8 and air leads 6.3 °C, within 0.1 °C.

Explain the order of the three dips.

4.8 App challenge · 8 min

Load scenario 4, O_2 trim toward the efficiency peak, and use the default seed.

(a) Before Run set Stack O_2 setpoint to 2.0 %. Run to 40 min and export the file. Average `efficiency_pct` and `co_ppm` over the rows from 30 to 40 min. *Checkpoint:* 82.23 % within 0.03, and 210 ppm within 5 ppm.

(b) Press the reload button, set Stack O_2 setpoint to 1.0 % and CO override On before Run, and run to 40 min. Average `o2_true_pct`, `co_ppm`, `efficiency_pct` and `co_selected` over the same rows. *Checkpoint:* 2.15 %, 150 ppm, 82.09 % and 1.

Explain why the heater in run (b) does not go to the 1.0 % it was asked for.

4.9 App challenge · 8 min

Load the challenge, Half an hour of trouble, and use the default seed. For each run below go to 30 min and read the Time fuel-rich, Time CO over permit, Peak CO and Lowest burner λ tiles.

(a) Run the challenge as loaded. *Checkpoint:* 3.7 s, 118 s, 7566 ppm, 0.989.

(b) Set Fuel and air to Cross-limited, press Reset and run again. *Checkpoint:* 0.0 s, 101 s, 295 ppm, 1.073.

(c) Also set CO override On, press Reset and run again. Read the Efficiency, run average tile and the challenge verdict as well. *Checkpoint:* 0.0 s, 0 s, 173 ppm, 1.088, 80.70 %, passed.

Say which of the three scripted events each change answers.

4.10 Application area · 4 min

With the override on, the heater of Section 4.5.2 settles at 2.15 % oxygen and 82.09 %, burning about 32 kg/h less fuel than at 3 % oxygen and 81.16 %.

(a) Find the fuel saved in a year of continuous operation, in tonnes and in terajoules at 47.0 MJ/kg, and check the 32 kg/h against the two efficiencies and the design fuel rate of 2828 kg/h. *Checkpoint:* About 280 t and 13.2 TJ. A fuel ratio of 81.16/82.09 saves 1.13 %, or 32 kg/h.

(b) The Limitations section says that burner-to-burner variation makes the knee of a real heater earlier and less sharp. What does that mean for the saving, and for the choice between a fixed oxygen setpoint and the override?

4.11 Application area · 3 min

The arch draft setpoint trades tramp air against the risk of a firebox above atmospheric.

(a) With the design leak of 3 %, find the tramp air of Equation (4.2) at -25, -10 and -5 Pa. *Checkpoint:* 0.45, 0.28 and 0.20 kg/s.

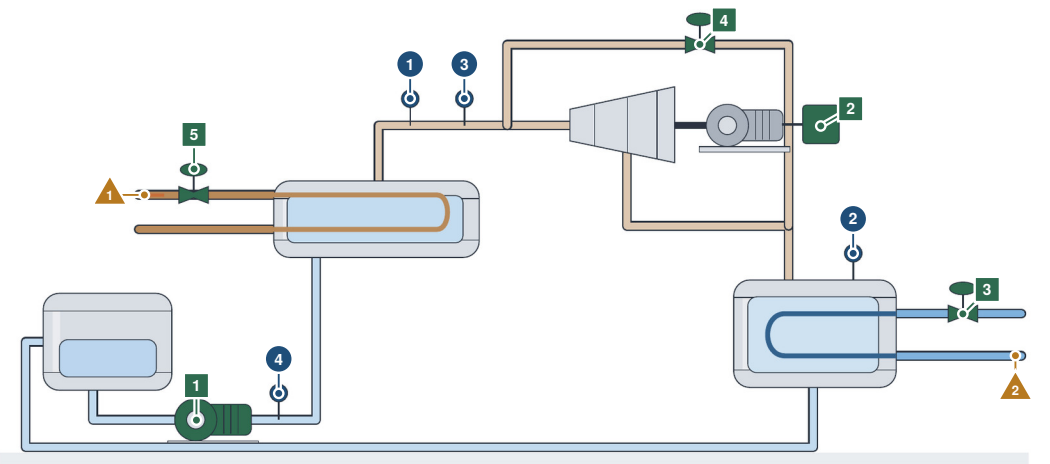
(b) In 30 Pa gusts the arch goes positive for 0, about 2 and about 86 s in fifteen minutes at these setpoints, and for 4 to 12 s and 71 to 97 s at the two higher setpoints over five other seeds. Which setpoint would you run in this wind, and why does the worst gust set it?

4.12 Review the agent · 5 min

An AI assistant reviews a run of scenario 8, Tramp air the burners never see, with the CO override off, and replies: "The stack O_2 is on its 3.0 % setpoint, so the burners have their 15 % excess air. Carbon monoxide near 420 ppm with correct air means the burners are fouled. Raise the O_2 setpoint to 4.0 % until CO is under the permit, and schedule a burner cleaning." Find the flaw and say how you would check it.

CHAPTER 5

Keeping an Organic Rankine Expander Dry When Its Waste Heat Falls



The controller moves 1 feed pump speed 2 expander speed 3 cooling-water flow 4 vapor bypass around the expander 5 source bypass

The controller reads 1 evaporating pressure 2 condensing pressure 3 expander inlet temperature, for the superheat 4 feed flow

What disturbs it 1 hot source water: its flow falls or its temperature changes 2 cooling water warms

Notice that the vapor leaving the evaporator must stay above its boiling point on its way to the expander, and that the heat arriving in the hot water can fall faster than the thermometer reports it. Illustration, not to scale.

IN PLAIN TERMS **Keep the Vapor Dry When the Heat Drops**

The main idea. Hot water from an engine boils a fluid that boils more easily than water. The vapor spins a machine that makes electricity. The vapor must stay a little hotter than its boiling point, or liquid drops can harm the machine.

The challenge. When the hot water suddenly slows, a thermometer in a metal sleeve reports the cooling vapor late. A control loop that waits for it acts too late, but measuring the heat as it arrives lets the pump slow down first.

Why it matters. Any plant whose heat supply can drop without warning faces the same race between a slow sensor and a fast loss.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history of organic-fluid engines, the fact on engine heat and the feedforward on a truck engine (Section 5.2), the try-it segments of scenarios 4 and 5 and the fact on fouling (Section 5.5) and the key points
Study	60 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	25 min	the first page, the results (Section 5.5), Technology in Practice (Section 5.6), Limitations (Section 5.7) and the open challenge

The control question When the waste heat behind an organic Rankine cycle falls by a third within seconds, can a well-tuned superheat loop keep the expander inlet dry, and which limits of the condenser and the heat source does no control structure remove?

Learning objectives

- Compute the expander's swept flow, fluid work, generator power and the cycle's net output from its model, and compare the efficiency with the Carnot bound.
- Derive the source-heat feedforward from a steady energy balance and convert a flow request into a pump speed.
- Explain why feedback through a thermowell, alone or with an override, loses the expander on a source loss that the feedforward rides through.
- Recognize a loop out of authority, such as the condensing loop on a hot day, and trade its setpoint against net power.
- Separate what a predictive controller gains from a forecast and what it gains from moving two handles together.

Prerequisites Sections 1.4.3 (static feedforward), 1.4.5 (inner loops), 1.4.2 (entry points), 1.6.2 (velocity form) and 1.6.3 (the MPC option). Chapter 2 for the structures on a simpler plant.

The app https://apmonitor.com/apps/organic-rankine/

See also Volume 4, Chapter 2 for a protective override that rests idle until its constraint nears, Volume 4, Chapter 9 for a power loop low-selected against an exhaust limit, and Volume 4, Chapter 10 for a closed power cycle with a predictive layer over its inner loops. Volume 17, Chapter 2 for the engines whose exhaust heat feeds a cycle like this one.



5.1 Introduction

An organic Rankine cycle (ORC) boils a refrigerant-like working fluid with heat that is too cool to raise useful steam and expands the vapor through a machine that drives a generator. In the simulated plant an intermediate water loop brings 959.95 kW from an on-site engine's exhaust at 150 °C to an R1233zd(E) evaporator at 12.90 bar (pressures are absolute throughout), a variable-speed positive-displacement expander delivers 91.74 kW at its generator, and after the feed pump, the cooling-water pumps and fans and the other auxiliaries the cycle nets 77.55 kW. That is 8.08 % of the accepted heat, against a Carnot bound of 29.54 % between the source and cooling-water supply temperatures, and the condenser rejects 863.96 kW, 9.4 times the generator output. The expander is specified for dry vapor, and its 10 K of superheat at the inlet is the quantity that the rest of the control system serves.

The simulator targets a specific misconception, the idea that a well-tuned superheat loop is enough to protect the expander. When the source-loop flow falls by 30 % in case 4, a cascade whose superheat controller runs at 0.08 kg/s per kelvin with a 20 s integral time lets the true superheat fall to 1.09 K, and the protection isolates the expander 17.3 s after the loss and keeps it out for 83.4 s. The controller sees the superheat through a 3 s thermowell after most of the evaporator's response has already happened. The same cascade with a feedforward computed from the source meters keeps the superheat between 9.39 and 11.08 K with the expander admitted throughout. A second belief, that a colder condenser always pays, fails on a hot day. With 35 °C cooling water, holding 1.95 bar drives the fans to their 30.0 kW limit and nets 8.69 kWh over eleven minutes, while letting the pressure rise to 2.6 bar nets 12.24 kWh.

The class of problem is general. An exchanger whose outlet state is held by the flow through it, while a second actuator sets the pressure that defines that state and an unmanaged source changes the heat, describes refrigeration evaporators and once-through boilers as well as every waste-heat ORC. A survey of the technology notes that both the pump speed and the expander speed must be controlled to hold the expander inlet conditions when the source varies [1], and the control of these cycles has been treated with multivariable designs [2] and with model-based feedforward on vehicle engines [3]. In process terms the superheat loop is a quality loop behind a measurement lag,

the inner flow loop is the secondary of a cascade (Section 1.4.5), the source-heat calculation is a measured-disturbance feedforward [4] (Section 1.4.3), and the overrides are selectors that hand an actuator to a constraint [5].

However, the published dynamic models of these cycles serve design and validation, with exchanger formulations compared against measured transients [6] and control strategies optimized on models [7] or validated on one engine [3]. A student who wants to watch the superheat fall through a thermowell, and then see a feedforward hold it, needs a plant whose loops, limits and protection are settings with consequences in kelvin, bar and kilowatts. This chapter separates what each element of a superheat control structure contributes to keeping the expander inlet dry on a waste-heat ORC, and locates the limits that the condenser and the heat source impose and no loop removes, on the simulator framework of Section 1.6.

The proportional-integral (PI) loops with back-calculation, the cascade, the feedforward and the minimum selector are textbook structures, and the chapter measures what each contributes in matched comparisons made on one plant with one random seed. Section 5.2 follows the domain from low-grade heat to the superheat loop. Section 5.3 builds a finite-volume cycle with one pressure per side, a two-zone receiver and audited property tables, whose working-fluid inventory closes to 3×10^{-12} kg and whose energy ledger closes to 8×10^{-5} of the accepted heat. Section 5.4 gives the loops, the selector, the protection layer and a predictive controller. Section 5.5 reports ten graded cases and the comparisons. Among them, the feedforward cuts the superheat error integrated over the source loss from 276.7 to 29.0 K s at the same gains, and doubling every exchanger's cell count moves the second figure to 64.8 K s without changing the ranking (Section 5.5.6). Section 5.6 describes the real instruments, drives and control platform, Section 5.7 states what the model leaves out, and Section 5.8 sums up.

5.2 From Low-Grade Heat to a Superheat Loop

Engines that run a Rankine cycle on an organic fluid have a long history, which a chapter-length account of the technology traces [8] and a review follows from the concept to current technology [9]. That review estimates the cumulative capacity of ORC systems that went into operation after 1995 at about 2000 MW of electricity, names geothermal reservoirs, waste heat from engines and industrial processes, biomass combustion and concentrated solar radiation as their sources, and notes that the same systems suit the cogeneration of heating or cooling [9]. Reviews presented the cycle for low-grade waste heat in 1997 [10] and catalogued its applications later [11], and manufacturers have sold ORC modules since the early 1980s [1].

> ⧖ HISTORY **From an Ether Engine to a Geothermal Plant, 1826 to 1952**
>
> An engine that runs a Rankine cycle on a fluid other than water is an old idea. In 1826 a patent described an engine with ether as its working fluid. Naphtha engines for launches followed, with the naphtha serving as fuel, working fluid and lubricant, and their low boiler pressure spared the owner the specialized operator a steam engine needed. In 1907 a flat solar collector boiled ether near 120 °C to drive a 2.6 kW engine. L. D'Amelio (1893 to 1967) worked out a solar plant on an organic fluid for irrigation in 1936, and in 1939 a 2.6 kW prototype for low-grade geothermal heat ran in a university laboratory. The first commercially operated geothermal plant of this kind ran for a short period in 1952. It used water at 91 °C and supplied 200 kW of electricity to a mining company [9].

Engine exhaust became a target early. A Rankine cycle on an internal combustion engine was first developed after the 1970 energy crisis, when a truck-engine prototype reported a 12.5 % reduction in fuel use over a 450 km road test, and present systems differ from that one mainly in their expansion machines and in the wider choice of working fluids [1]. The engines themselves are the plant of Volume 17, Chapter 2, where a module's power settles within seconds of a fuel step while its recovered heat needs minutes. The same survey treats the heat-rejection capacity as a constraint of its own,

because cooling fans can sharply reduce the net output [1], which is the trade of case 5. Data centers offer a further low-grade source, whose cooling technology and heat recovery opportunities have been reviewed [12], and case 8 measures how little a cycle makes from it.

> ✦ DID YOU KNOW? **At Least Half the Fuel Leaves as Heat**
>
> Even a modern engine cannot convert more than 50 % of its fuel energy into useful work. The rest is lost through the coolant and the exhaust gases, and the exhaust gases carry it at the highest temperatures [3].

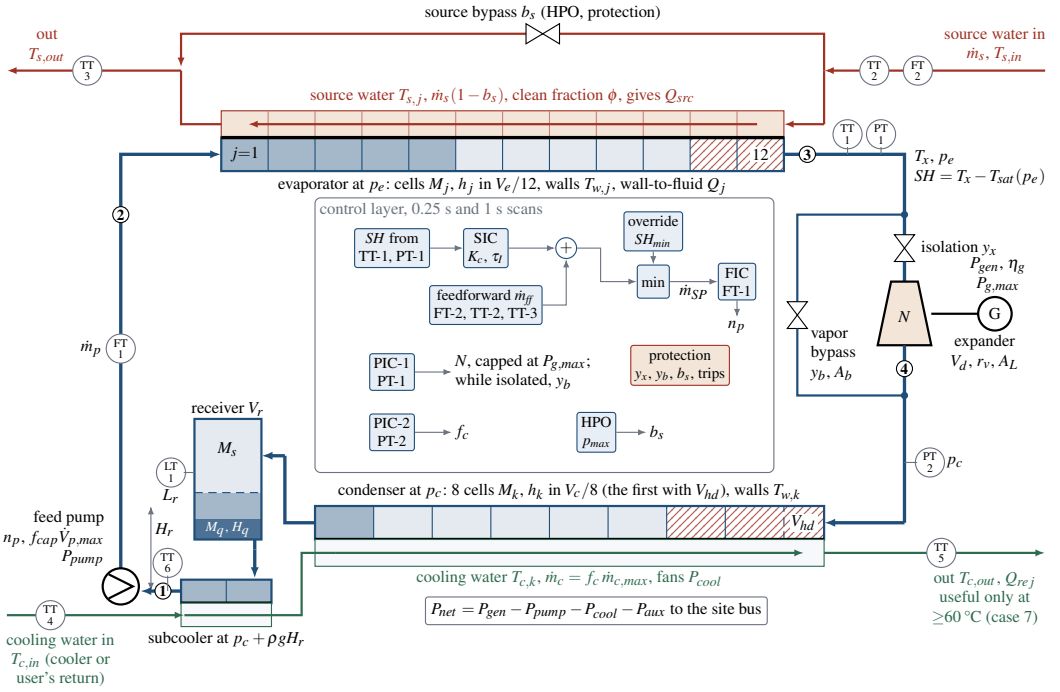
The working fluid and the machine decide where the superheat constraint comes from. A fluid whose saturated-vapor line has a positive slope stays superheated through an expansion, while a wet fluid such as water must be superheated at a turbine inlet to keep droplets from forming [1]. R1233zd(E) has been tested as a drop-in replacement for R245fa in an experimental ORC for low-temperature heat [13], and its reference equation of state, recommended as an international standard by the working group revising ISO 17584, holds from the triple point at 165.75 K to 450 K and 100 MPa [14]. Small units use positive-displacement expanders, often derived from compressors, which tolerate liquid during expansion and lose efficiency when the system's volume ratio departs from the machine's built-in volume ratio [1]. A semi-empirical scroll-expander model divides the expansion into a supply pressure drop, heat transfer, an isentropic expansion to the pressure the built-in volume ratio imposes and an expansion at constant machine volume, with the leakage mixed at the exhaust, and the machine it describes reached a maximum overall isentropic effectiveness of 68 % [15]. The expander simulated here keeps that structure in reduced form. Its dry-inlet requirement belongs to the chosen machine and is not a property of positive-displacement expanders in general.

Control came with the dynamic models. Dynamic ORC models were written for waste-heat recovery [16], the finite-volume and moving-boundary forms are the most widely used for exchangers with phase change, and the effect of the finite-volume form's number of control volumes has been examined against measurements [6]. Optimal control strategies for waste-heat cycles were built on such models [7], and a survey reports that re-optimizing the operating point continuously for a variable source raised the cycle's second-law efficiency by about 10 % [1]. Multivariable control has been applied to ORCs in waste-heat processes [2]. On a heavy-duty diesel engine a model-based feedforward on the pump speed held the superheat far closer to its setpoint than feedback alone, as the history segment below tells, and the superheat had to stay positive to prevent droplets that could damage the turbine [3]. The simulator reproduces that ordering with a steady energy-balance feedforward in place of a model inversion.

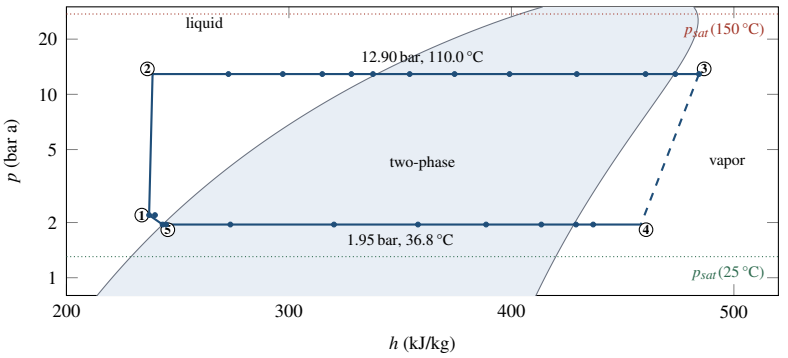
> ⧖ HISTORY **A Feedforward on a Truck Engine's Cycle, 2017**
>
> A vehicle's exhaust is a harder source than a geothermal well, because it changes strongly with the driving. A study published in 2017 fitted an organic Rankine cycle to a heavy-duty diesel engine in a test cell and held the superheat at the evaporator outlet with the pump speed. A gain-scheduled PID alone was compared with the same PID plus a dynamic feedforward, computed by inverting a reduced model of the evaporator. Under the same swings of engine torque and speed, the largest superheat error was about 1.9 K with the feedforward and about 10 K without it. A supervisor stepped the cycle through stop, cold idle, hot idle and power production, and over a road cycle of more than 50 min the superheat stayed tightly controlled [3].

5.3 A Closed R1233zd(E) Circuit Between Two Water Loops

The plant in Figure 5.1 was designed for teaching and copies no installation, and every size, efficiency and threshold in it is a synthetic value. Pressurized water from an engine's exhaust-heat exchanger arrives at 140 to 170 °C and passes a twelve-cell counterflow evaporator that preheats, boils and superheats the R1233zd(E). The vapor passes an isolation valve into the expander, or a vapor bypass around it, then an exhaust header and an eight-cell condenser cooled by water that has first passed a two-cell subcooler. The receiver between the condenser and the subcooler holds whatever inventory



(a) Process and control schematic



(b) Design point on the pressure–enthalpy plane

Figure 5.1 The simulated cycle as drawn on the application's Cycle view. (a) The source water ($\dot{m}_s$, $T_{s,j}$) gives Q_{src} to the evaporator's twelve cells at p_e, the expander at speed N behind the isolation valve y_x drives the generator, the vapor bypass y_b throttles around it, and the condenser's eight cells at p_c reject Q_{rej} to the cooling water at $\dot{m}_c$. The receiver holds a saturated zone M_s over a subcooled layer (M_q, H_q), and the subcooler and pump suction sit the static head H_r below it. Shading marks the design point's phases (dark liquid, light two-phase, hatched vapor). Circles are the instruments the controllers read, and the block diagram's outputs carry the actuator symbols. (b) The design point on the pressure–enthalpy plane from the application's property table, with the numbered states and every cell. Process model only, not to scale.

the rest of the circuit does not, and a positive-displacement feed pump returns the liquid to the evaporator. No other application runs, so the source is a declared boundary condition. Case 7 replaces the cooling water with a warm-water user's 50 °C return, and the condenser heat counts as useful only while it leaves at or above the user's 60 °C.

Table 5.1 sorts the variables by role. The integrated state has 84 components. Seventy-seven are physical, the two side pressures p_e and p_c, twelve evaporator cell masses, eleven low-side masses (eight condenser cells, the receiver and two subcooler cells), the subcooled layer's mass and enthalpy, 22 wall temperatures, 22 secondary-water temperatures and six actuator positions. Seven ledgers integrate the boundary fluxes so that conservation can be checked against independent states. Inside the model the units are SI, and the interface shows bar, °C and kW.

5.3.1 Cells at One Pressure per Side

Every working-fluid cell has a fixed volume V_j and shares its side's pressure p, so its balances read as shown in Equation (5.1).

$$\begin{aligned} \dot{M}_j &= \dot{m}_{j-1} - \dot{m}_j \\ M_j \dot{h}_j - V_j \dot{p} &= \dot{m}_{j-1}(h_{f,j-1} - h_j) - \dot{m}_j(h_{f,j} - h_j) + Q_j \end{aligned} \tag{5.1}$$

where $\dot{m}_j$ is the flow through the face that leaves cell j, $h_{f,j}$ the enthalpy that face carries and Q_j the heat from the wall. The second line is the energy balance $\dot{U}_j = \dot{m}_{j-1} h_{f,j-1} - \dot{m}_j h_{f,j} + Q_j$ for $U_j = M_j h_j - p V_j$ with h_j times the mass balance subtracted. The states are p and M_j. A cell's enthalpy, temperature and phase follow from p and $\rho_j = M_j / V_j$ through the property table, so its mass balance is $V_j[(\partial\rho/\partial p)_h \dot{p} + (\partial\rho/\partial h)_p \dot{h}_j] = \dot{m}_{j-1} - \dot{m}_j$. Eliminating $\dot{h}_j$ leaves one linear relation per cell between its two face flows and the pressure rate, as shown in Equation (5.2).

$$\begin{gathered} A_j \dot{m}_{j-1} - B_j \dot{m}_j = C_j \dot{p} + D_j \\ A_j = 1 - q_j(h_{f,j-1} - h_j), \quad B_j = 1 - q_j(h_{f,j} - h_j) \\ C_j = V_j\left[(\partial\rho/\partial p)_h + q_j\right], \quad D_j = q_j Q_j \end{gathered} \tag{5.2}$$

where $q_j = (\partial\rho/\partial h)_p / \rho_j$. Marching from the side's inlet flow makes every face flow affine in the pressure rate, $\dot{m}_j = \gamma_j + \delta_j \dot{p}$, and the outlet flow fixes it, $\dot{p} = (\dot{m}_{out} - \gamma_n)/\delta_n$. The outlet flow is the expander plus the bypass on the high side and the pump on the low side, each evaluated at the cell states of the moment, so the pressure is not relaxed toward anything and no cell carries a pressure of its own. A forward face carries the upstream cell's enthalpy, and the receiver drains its saturated liquid while it holds two phases. A reversed face carries the downstream cell's saturated vapor when that cell is two-phase, since vapor is the phase that can move back, and otherwise its enthalpy, except that a dense back-flow into a lighter cell with $B_j < 0.2$ carries the upstream enthalpy instead. The directions are iterated, at most $n+1$ times, until every flow agrees with the direction assumed for it. Each flux leaves one cell and enters the next, so the internal faces cancel exactly in both inventories.

This formulation departs from the per-cell (M, U) states with a pressure recovered in each cell that the development plan specified. That form was built first and, according to the build record, diverged within 1 s at the 0.05 s step from an exact steady start, growing about sixfold per step from the last condenser cell, which sits on the saturated-liquid line at the design point. A liquid cell's pressure is set by a change of about one part in a million in its density, and Runge-Kutta stages on either side of the dome saw compressibilities differing by orders of magnitude. That code is no longer in the source tree, so the divergence is quoted from the record and was not re-run for this chapter. With p and M_j as states, the inversion $h(p, \rho)$ is one-dimensional and monotone in every phase, because at fixed pressure the density falls steadily with enthalpy through liquid, dome and vapor. The (ρ, u) inversion remains implemented, and a check of the application confirms that every cell's density and internal energy return its side's pressure within 1×10^{-5} of it.

The price is a set of assumptions. There is no frictional pressure drop or momentum balance, pressure equalizes instantly along each side, and cells are homogeneous and in equilibrium. Where a subcooled stream meets a nearly liquid two-phase volume with no compressible vapor downstream, δ_n can pass through zero and Equation (5.2) lose its solution. The review stage found two such places, a receiver fed with subcooled condensate and an evaporator flooded by a pump held against a

Table 5.1 Variables Sorted by Role, Valued at the Design Point (Case 1, $t = 0$)

Variable	Role	Design	Range
SH	CV, expander-inlet superheat	10.0 K	SP 3–25
p_e	CV, evaporating pressure	12.90 bar	SP 6–18, trip 21
p_c	CV, condensing pressure	1.95 bar	SP 1.2–6
$\dot{m}_p$	CV (inner loop), feed flow	3.909 kg/s	0–6.25
n_p	MV, pump speed	62.71 %	0–100
N	MV, expander speed	2146 rpm	600–3600
f_c	MV, cooling-water flow	55.95 %	10–100
b_s	MV, source bypass (override)	0 %	0–100
y_x, y_b	protection, isolation, bypass	100, 0 %	0–100
$\dot{m}_s$	DV, source-loop flow	100 % of 5 kg/s	0–130
$T_{s,in}$	DV, source supply	150 °C	40–170
$T_{c,in}$	DV, cooling-water supply	25 °C	5–40, or 50
ϕ, f_{cap}	DV, clean source side, pump delivery	100, 100 %	30–100
σ	DV, instrument noise multiple	1	0–4
M_j	states, cell masses	Fig. 5.1b	charge 160 kg
M_r	state, receiver mass	93.72 kg	level 50.2 %
M_q, H_q	states, subcooled layer	0, 0	—
T_w, T_s, T_c	states, walls, water	per cell	—

lost source, resolved by the two-zone receiver below and by a flood switch that trips the plant first. Equation (5.1) holds to round-off while a cell keeps its phase, and the only energy residual comes from cells that cross the saturated-liquid line within a step, where the density partials jump between Runge-Kutta stages.

5.3.2 The Receiver in Two Zones and the Static Head

The receiver is a saturated zone, homogeneous at p_c with mass $M_s = M_r - M_q$ in the volume $V_r - V_q$, over a layer of subcooled condensate with its own mass M_q and enthalpy content H_q. Condensate more than 0.2 kJ/kg below saturation is routed to the layer, the layer supplies the outflow first and warms through an interface conductance, and a layer within 0.5 kJ/kg of saturation merges into the saturated zone, as shown in Equation (5.3).

$$\begin{aligned}
\dot{M}_q &= \alpha\dot{m}_{in} - \beta\dot{m}_{out} - \mu \\
\dot{H}_q &= \alpha\dot{m}_{in}h_{in} - \beta h_q\dot{m}_{out} + Q_i - \mu h_q + V_q\dot{p}_c \\
\alpha &= \mathrm{clip}_{[0,1]}(h_l - h_{in} - 0.2), \quad \beta = \frac{M_q}{\max(M_q, 0.5)} \\
Q_i &= UA_i \max(0, T_{sat} - T_q)\min(1, M_q/0.5) \\
\mu &= kM_q
\end{aligned} \tag{5.3}$$

where enthalpies are in kJ/kg and masses in kg inside the clips, $h_q = H_q/M_q$, α is the share of the arriving condensate routed to the layer, β the share of the outflow it supplies, $UA_i = 0.2\,\mathrm{kW/K}$, and $k = 1\,\mathrm{s}^{-1}$ times $\mathrm{clip}_{[0,1]}\big((h_q - h_l + 0.5)/0.5\big)$. The receiver's row of Equation (5.2) becomes the sum of the two zones' volume constraints, the layer's weighted by $\omega = \rho_s/\rho_q$, so that $A = (1-\alpha)A_s + \omega\alpha A_q$, $B = (1-\beta)B_s + \omega\beta B_q$ and $C = C_s + \omega C_q$, with the merge flow and the interface heat in D. The saturated zone must keep 3 % of the receiver's volume, or the run leaves the model. With saturated condensate the layer stays empty and the receiver is the homogeneous cell exactly. A check of the application confirms that manual pump steps to 40, 35, 33 and 0 %, which once ended outside the model, now stay inside it.

The receiver's liquid stands $H_r = 2.0\,\mathrm{m}$ above the pump inlet, so the subcooler and the pump suction sit at $p_S = p_c + \rho_{nom} g H_r$, a constant 0.245 bar above the condensing pressure with $\rho_{nom} = 1250\,\mathrm{kg/m^3}$. The potential energy, about 20 J/kg, is left out of the energy balance. The head is what lets a stopped pump restart. At standstill the circuit settles to the cooling-water temperature with the condensing

pressure at 1.303 bar, and the suction still has 5.11 K of subcooling, against 8.72 K at the design point.

5.3.3 Walls, Secondary Water and Conductances

Each heat-transfer cell has a wall node and a secondary-water node in counterflow, as shown in Equation (5.4).

$$\begin{aligned} C_w \dot{T}_{w,j} &= UA_{s,j}(T_{s,j} - T_{w,j}) - UA_{f,j}(T_{w,j} - T_j) \\ C_s \dot{T}_{s,j} &= \dot{m}_s(1-b_s)c_{p,s}(T_{s,j+1} - T_{s,j}) \\ &\quad - UA_{s,j}(T_{s,j} - T_{w,j}) \\ UA_{f,j} &= U(x_{h,j})\left(\dot{m}/\dot{m}_{ref}\right)^{0.6} \\ UA_{s,j} &= \frac{UA_s}{12}\,\phi\left(\frac{\dot{m}_s(1-b_s)}{\dot{m}_{s,0}}\right)^{0.8} \end{aligned} \tag{5.4}$$

where $Q_j = UA_{f,j}(T_{w,j} - T_j)$, $T_{s,13} = T_{s,in}$ and the source leaves at $T_{s,out} = T_{s,1}$. $U(x_h)$ blends the liquid, boiling and vapor conductances on the enthalpy quality x_h, liquid at $x_h \le 0$, a ramp to boiling up to 0.1, boiling to 0.85 and a dry-out ramp to vapor at 1, so it is continuous as a cell changes phase. The working-fluid flow $\dot{m}$ is the pump flow in the evaporator and subcooler and the expander plus bypass flow in the condenser, each flow ratio is floored at 0.05, and $\dot{m}_{ref} = 4.0\,\mathrm{kg/s}$. The low side is written the same way with the cooling water $\dot{m}_c = f_c \dot{m}_{c,max}$ entering the last subcooler cell and leaving the first condenser cell. The accepted heat is $Q_{src} = \dot{m}_s(1-b_s)c_{p,s}(T_{s,in} - T_{s,out})$ and the rejected heat $Q_{rej} = \dot{m}_c c_{p,c}(T_{c,out} - T_{c,in})$. Heat may cross any wall in either direction, because no check-valve or bypass regime prevents it. In case 7 the switch to the user's return moves 0.54 kWh from the water into the condenser metal and 0.12 kWh from the metal into the fluid, and in case 4 and the suggested challenge no heat flows against its normal direction.

5.3.4 Expander, Bypass, Pump and Auxiliaries

The expander swallows a fixed volume per revolution and expands along the isentrope to its built-in volume, after which the pressure jumps to the exhaust at constant volume, as shown in Equation (5.5).

$$\begin{aligned} \dot{m}_{sw} &= y_x \rho_{in} V_d N \\ \dot{m}_L &= y_x A_L \sqrt{\rho_{in}(p_e - p_c)} \\ w &= \eta_e \eta_{su}(h_{in} - h_{ad}) + (p_{ad} - p_c)/\rho_{ad} \\ \rho_{ad} &= \rho_{in}/r_v, \quad s_{ad} = s_{in} \\ \eta_{su} &= \max\left(0,\, 1 - c_{su}(N/N_{50})^2\right), \quad W_f = \dot{m}_{sw} w \\ P_{gen} &= \eta_g W_{sh}\ (W_{sh} \ge 0), \quad W_{sh}/\eta_g\ (W_{sh} < 0) \end{aligned} \tag{5.5}$$

where $W_{sh} = W_f - 2\pi\tau_f N$ is the shaft power, N is in revolutions per second, $N_{50} = 50\,\mathrm{s}^{-1}$ (3000 rpm), and the leakage does no work and mixes with the expanded flow, so $h_{out} = h_{in} - W_f/(\dot{m}_{sw} + \dot{m}_L)$. Under-expansion ($p_{ad} > p_c$) adds the blow-down term and over-expansion subtracts it. At the design point the swept flow is 3.567 kg/s with 0.341 kg/s of leakage (8.7 %), $w = 28.62\,\mathrm{kJ/kg}$ with $p_{ad} = 4.53$ bar, the fluid work is 102.09 kW, friction takes 4.49 kW, the shaft isentropic efficiency is 0.655 and the torque 434 N m. Rotor inertia is neglected, so the drive follows its command through a 0.5 s lag and a 300 rpm/s slew. The vapor bypass is an isenthalpic throttle, $\dot{m}_b = y_b A_b \sqrt{\rho_{in} \min(p_e - p_c, 0.5 p_e)}$, critical above a pressure ratio of two.

The feed pump is a positive-displacement machine drawing from the last subcooler cell, as shown in Equation (5.6), together with the remaining parasitics and the net output.

$$\begin{aligned} \dot{m}_p &= n_p \dot{V}_{p,max} f_{cap} \rho_1, \qquad h_2 = h_1 + \frac{p_e - p_S}{\rho_1 \eta_p} \\ P_{pump} &= \frac{\dot{m}_p (h_2 - h_1)}{\eta_m}, \qquad P_{cool} = P_{c,max} f_c^3 \\ P_{aux} &= P_{stby} + (P_a - P_{stby})\, r_a \\ r_a &= \mathrm{clip}_{[0,1]}\big(\max(n_p, y_x)/0.05\big) \\ P_{net} &= P_{gen} - P_{pump} - P_{cool} - P_{aux} \end{aligned} \tag{5.6}$$

At design the pump head is 10.705 bar, 6.10 kW enters the fluid and the motor draws 6.94 kW, the cooling water costs 5.26 kW and the other auxiliaries 2.00 kW. The actuator positions n_p, f_c, b_s, y_b and y_x respond to their commands as first-order lags with time constants of 1, 2, 3, 1 and 1 s. The electrical boundary is a site bus with nonnegative terms, generation plus import equal to the pump, cooling, auxiliary and generator-motoring demand plus export, with import and export never both nonzero and the residual logged at every step. No data-center load is modeled, so no service can be unmet or curtailed.

5.3.5 Property Tables for R1233zd(E)

The browser never evaluates an equation of state. A generator script evaluates the reference equation of state for R1233zd(E) [14] through the HEOS backend of CoolProp 8.0.0 [17], pinned in its own virtual environment, on four grids. The saturation table has 161 nodes uniform in $\zeta = \ln p$ from 0.2 to 30 bar, with both phases' temperature, densities, internal energies and entropies and their exact slopes along the saturation line, and a 57-node saturated-liquid table is uniform in temperature from 255.15 K to the saturation temperature at 30 bar. The vapor surface has 57×33 nodes in (ζ, ξ), where $T = T_{sat}(p) + \kappa(\xi)(T_{max} - T_{sat})$, $\kappa(\xi) = \xi(\xi + 0.25)/1.25$ and $T_{max} = 450$ K, the equation's limit. The liquid surface has 57×17 nodes in (T, ψ) with $p = p_{sat}(T) + \kappa(\psi)(32\,\mathrm{bar} - p_{sat})$, on a temperature axis because a liquid's density follows its temperature far more than its pressure. The vapor surface stores $\ln\rho$, u and s and the liquid surface ρ and u, each as a deviation from the saturated phase at the same pressure or temperature, so the dome edge is exact, and both are read by bicubic Hermite interpolation with fourth-order finite-difference slopes. The reference state is that of the International Institute of Refrigeration. The SHA-256 checksum of the stored arrays, recomputed by a check of the application from the arrays it reads, matches the one the generator wrote. R245fa was rejected because its equation of state in the same library ends at 440 K, below the 170 °C top of the source range.

The cell closure solves $\rho(p,h) = M_j/V_j$ at fixed p by Newton's method from the cell's previous accepted state, in one variable whatever the phase, and returns $(\partial\rho/\partial p)_h$ and $(\partial\rho/\partial h)_p$ from the same interpolants, so the coefficients of Equation (5.2) are the exact derivatives of the surface being inverted. Against 1500 CoolProp states drawn at random off the grids below 25 bar, the worst errors are 4.9×10^{-7} K in saturation temperature, 8.6×10^{-6} relative in vapor density, 4.2×10^{-8} in liquid density, and 0.68, 0.008 and 0.012 J/kg in the enthalpy the closure returns for vapor, liquid and two-phase cells. The inversions put every state in its phase, and the table's own audit records a Newton residual of 9.7×10^{-13}.

5.3.6 Instruments, Protection and Events

The controllers, the feedforward and the protection's interventions read only a measurement packet. PT-1 and PT-2 read the two pressures through 0.3 s lags with 0.005 and 0.003 bar of noise. TT-1 reads the expander-inlet temperature through a thermowell lag $\tau_{TT} = 3$ s with 0.05 K of noise, and the superheat transmitter computes $SH = T_x - T_{sat}(p_e)$ from the same table. FT-1 reads the feed flow

(0.5 s, 0.3 %), FT-2 the source flow (1 s, 0.5 %), TT-2 to TT-5 the water temperatures (5 s), TT-6 the pump inlet (3 s), giving the suction subcooling against $T_{sat}(p_c + \rho_{nom} g H_r)$, and LT-1 the receiver level (2 s, 0.5 %). Noise is added after the lag, scaled by the noise multiple, and every scan draws the same six deviates from a seeded generator. The CSV export marks each truth column with the prefix `truth_`, so that no replay can use it as a measurement.

Protection sits outside every controller, acts in Manual and changes state only at a step boundary, after the integration. Its interventions act on the instruments. A measured superheat below 3 K, a feed flow below 0.1 kg/s for 2 s, or generator motoring beyond 6 kW for 5 s isolates the expander and opens the vapor bypass. Readmission needs 20 s above 8 K with the feed flowing, a pressure ratio of at least 1.5 and a measured superheat that has fallen by no more than 0.5 K over the last 5 s. An evaporating pressure above 19 bar forces the source bypass open until it has stayed below 17.5 bar for 20 s. Latched trips are a measured superheat below 1 K while the expander still admits, a pressure above 21 bar, vapor or an empty receiver at the pump suction, liquid over half the evaporator's outlet cell for 5 s, and a state outside the table. A trip stops the pump, isolates the expander and opens both bypasses, and Reset trip requires 30 s, the pressure below 17.5 bar, 15 % of receiver level and 2 K of suction subcooling. Warnings at 6 and 25 K of superheat and 17.5 bar carry no action. All of these thresholds are teaching values. Actual ones would be taken from the expander and pressure-equipment manufacturers' data and the applicable codes.

5.3.7 Initialization and Integration

Each case starts at the steady state of its own disturbances and setpoints, solved from the 160 kg charge. Both sides are marched against their secondary water from the end whose state is known, the high side from the expander inlet back to the pump outlet for the flow that closes the march and the low side from saturated liquid at the condenser outlet back to the inlet for the cooling-water flow that does, with the pump work and the expander exhaust coupling them in a fixed-point loop. The receiver takes what the other cells leave of the charge, so its level, 50.2 % at design, is a result. At that state the largest scaled derivative is 1.6×10^{-8} and the steady first law, accepted heat plus pump work minus expander work and rejected heat, closes to 2.6×10^{-8} W. The integrator is classical fourth-order Runge-Kutta, inside the engine of Section 1.6.1. The fastest retained mode, the vapor in the first condenser cells, lies between 35 and 112 s^{-1} over seven operating points, outside the method's stability limit at the 0.05 s physics step, so each step takes three substeps of $1/60$ s, which put $|\lambda h|$ at or below 1.9. The slowest modes have time constants of 17 to 47 s. The controllers scan every 0.25 s and the historian logs one row per simulated second.

5.3.8 Assumptions

Parameters are listed in Table 5.2. Beyond the uniform pressure and homogeneous cells of Section 5.3.1, each wall is one node per cell, and with twelve evaporator cells the superheated region spans one or two cells, so the outlet temperature is resolved coarsely and the exchangers' effectiveness is not converged in the cell count, a sensitivity that Section 5.5.6 quantifies. The expander map is a teaching structure with no fitted machine behind it. The source and cooling water arrive at their slider temperatures whatever the plant does, because the engine and the cooler lie outside the boundary. Cold startup and filling, lubricant, non-condensable gas, pump cavitation and potential energy are outside the model. None of the values describes a particular plant.

5.4 Loops, a Selector and a Protection Layer in Code

One framework scan of $\Delta t_c = 0.25$ s runs the instruments, the flow loop, the evaporating-pressure loop and the high-pressure override. A resettable counter inside it runs the superheat loop, the low-superheat override and the condensing-pressure loop every fourth scan, each handed the time that actually elapsed. Table 5.3 lists the loops with their defaults. Every counter, filter and controller state is created again on a reset or a case reload.

Table 5.2 Model Parameters. Class T Is a Synthetic Teaching Value, L Comes From the Cited Equation of State Through CoolProp 8.0.0, and C Is Fixed in Code. Every Parameter Not Marked H Is Shown and Adjustable in the Parameters Panel

Symbol	Meaning	Value	Unit	Cl.
Source loop and evaporator				
$\dot{m}_{s,0}$, $c_{p,s}$	design flow, heat capacity	5.0, 4300	kg/s, J/(kg K)	T
M_s, UA_s	water held, conductance	55, 108	kg, kW/K	T
V_e, $C_{w,e}$	fluid volume, metal	60, 300	L, kJ/K	T
$UA_{l,e}$, $UA_{b,e}$, $UA_{v,e}$	liquid, boiling, vapor	54, 126, 21.6	kW/K	T
τ_{bs}	source bypass response	3	s	T
Expander, generator and bypass				
V_d, r_v	swept volume, built-in ratio	1.5, 3.0	L, —	T
A_L, η_e	leakage area, internal efficiency	40, 0.85	mm^2, —	T
c_{su}, τ_f	supply loss at 3000 rpm, friction	0.10, 20	—, N m	T
η_g, $P_{g,max}$	generator efficiency, rating	0.94, 120	—, kW	T
N_{min}, N_{max}	speed range	600, 3600	rpm	T
$\dot{N}_{max}$, τ_N	drive slew, response	300, 0.5	rpm/s, s	T
A_b, τ_v	bypass area, valve stroke	650, 1	mm^2, s	T
Condenser, receiver and subcooler				
V_c, V_{hd}, $C_{w,c}$	fluid, header, metal	120, 40, 200	L, L, kJ/K	T
$UA_{v,c}$, $UA_{c,c}$, $UA_{l,c}$	vapor, condensing, liquid	60, 300, 120	kW/K	T
UA_{cw}, M_{cw}	water side, water held	480, 80	kW/K, kg	T
$\dot{m}_{c,max}$, $\dot{m}_{c,0}$	full flow, conductance basis	40, 26	kg/s	T, V/H
$c_{p,c}$, $f_{c,min}$	heat capacity, least flow	4180, 10	J/(kg K), %	T, H/V
$P_{c,max}$, τ_c	pumps, fans at full flow, lag	30, 2	kW, s	T
V_r, H_r, UA_i	receiver, head, layer conductance	150, 2.0, 0.2	L, m, kW/K	T
V_{sc}, $C_{w,sc}$, $M_{sc,w}$	subcooler, metal, water	10, 20, 10	L, kJ/K, kg	T
UA_{sc}, $UA_{sc,w}$	subcooler fluid, water side	4, 16	kW/K	T
Pump, instruments and plant				
$\dot{V}_{p,max}$, τ_p	displacement flow, response	5.0, 1	L/s, s	T
η_p, η_m	hydraulic, motor efficiency	0.55, 0.88	—	T
τ_{TT}	thermowell lag (8 s in case 2)	3	s	T
M_{tot}	charge	160	kg	T
P_a, P_{stby}	auxiliaries, standby	2.0, 0.3	kW	T
$T_{u,ret}$, $T_{u,min}$	user return, least supply	50, 60	°C	T
Fixed in code				
$\dot{m}_{ref}$, exponents	flow basis, film exponents	4.0, 0.6, 0.8	kg/s, —	C
—	layer routing, ramp, merge	0.2, 1.0, 0.5	kJ/kg	C
ρ, u, s, T_{sat}	R1233zd(E) properties	tables	SI	L
—	cells: evaporator, condenser, sub.	12, 8, 2	—	C
ρ_{nom}, g	head density, gravity	1250, 9.80665	kg/m^3, m/s^2	C
—	layer mass, merge rate, zone	0.5, 1, 3	kg, s^{-1}, %	C
Δt, h	physics step, RK4 substep	0.05, 1/60	s	C

Table 5.3 Loops, Scans, Default and Suggested Tunings and Actions. Action −1 Is Direct Acting, +1 Reverse

Loop	PV → MV	Scan	K_c default (suggested)	τ_I	Act.
SIC, superheat	$SH \rightarrow \dot{m}_{SP}$	1 s	0.04 (0.08) kg/s/K	30 (20) s	−1
FIC, feed flow	$\dot{m}_p \rightarrow n_p$	0.25 s	8 %/(kg/s)	1.5 s	+1
override, low SH	$SH \rightarrow \dot{m}_{SP}$	1 s	0.08 kg/s/K, SP 5 K	15 s	−1
PIC-1, evaporating	$p_e \rightarrow N$	0.25 s	250 (800) rpm/bar	8 (5) s	−1
PIC-2, condensing	$p_c \rightarrow f_c$	1 s	40 %/bar	15 s	−1
HPO, high pressure	$p_e \rightarrow b_s$	0.25 s	30 %/bar, SP 17 bar	10 s	−1

5.4.1 The Shared PID Law

Every loop is an instance of the velocity-form PID of Section 1.6.2, written in this chapter's symbols in Equation (5.7).

$$\begin{aligned} e_k &= a(SP_k - PV_k), \quad P_k = K_c e_k \\ D_k &= -K_c \tau_D a \dot{PV}_{f,k} \\ z_k &= z_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ \hat{u}_k &= u_b + z_k + I_k, \qquad u_k = \mathrm{clip}_{[u_{min},u_{max}]}(\hat{u}_k) \\ I_{k+1} &= I_k + \Delta t_k \left[(K_c/\tau_I) e_k + (u_k - \hat{u}_k)/T_t \right] \end{aligned} \tag{5.7}$$

where $a = \pm 1$ is the action, u_b a bias, z_k the running proportional-plus-derivative total (S_k of Equation (1.10)), $\hat{u}_k$ the output before limiting (v_k of Equation (1.11)) and $T_t = \tau_I$ the back-calculation tracking time. An integral time of 3000 s or more removes integral action. All loops here run with $\tau_D = 0$ and without an output slew. A loop in Manual, or a controller that does not hold its actuator, tracks the applied value through its integral as Section 1.6.2 describes, and a PID is never both updated and tracked in the same scan.

5.4.2 Independent Loops and the Cascade

In the independent architecture the superheat controller's output, a flow in kg/s, becomes a pump speed through the pump's nominal capacity, $n_p = 100\,\dot{m}_{SP}/(\dot{V}_{p,max}\rho_{nom})$, so its gain means the same thing in both architectures. In the cascade it becomes the setpoint of the inner flow loop, which moves the pump until FT-1 reads it and so corrects a worn pump or a density change within seconds. The superheat controller's bias is the feedforward of Equation (5.8), the static compensator of Section 1.4.3 written as a steady energy balance.

$$\begin{aligned} \dot{m}_{ff} &= \frac{\hat{\dot{m}}_s c_{p,s} (\hat{T}_{s,in} - \hat{T}_{s,out})}{h(\hat{p}_e, T_{sat}(\hat{p}_e) + SH_{SP}) - \hat{h}_2} \\ \hat{h}_2 &= h\big(\hat{p}_S, \min(\hat{T}_1, T_{sat}(\hat{p}_S) - 0.05)\big) + \frac{\hat{p}_e - \hat{p}_S}{\rho_{nom}\eta_p} \end{aligned} \tag{5.8}$$

where hats are measurements from FT-2, TT-2, TT-3, PT-1, PT-2 and TT-6, and $\hat{p}_S = \hat{p}_c + \rho_{nom} g H_r$. It is zero when the enthalpy rise is below 10 kJ/kg, it passes a 2 s first-order filter, and switching it in or out moves its step into the controller's integral so the output does not jump. While the generator is at its rating the feedforward is capped at the flow the speed-capped expander can pass, because extra feed beyond it only floods the evaporator. It is a measured-heat feedforward, built from the source flow and both source temperatures, rather than a feedforward of the flow alone.

The low-superheat override joins through a minimum selector with external reset. Each supervisory scan both controllers' outputs are previewed from their states without changing them, $\hat{u} = u_b + K_c a(SP - PV) + I$, and the lower is selected, as shown in Equation (5.9).

$$\begin{aligned} I_o &\leftarrow \hat{u}_{SIC} - u_{b,o} \quad \text{while idle, so} \\ \hat{u}_o &= \hat{u}_{SIC} + K_{c,o}(SH - SH_{min}) \end{aligned} \tag{5.9}$$

The idle override therefore sits exactly its proportional action above the main request and takes the flow setpoint as soon as the superheat falls below $SH_{min} = 5$ K. While it is selected the main controller tracks its output, and when the pump runs at full speed the main controller's ceiling falls to the measured flow, so neither winds up while the pump cannot deliver more. A protective override that waits idle until its constraint nears is also the subject of Volume 4, Chapter 2.

5.4.3 Pressure Loops, the Overrides and Isolation

PIC-1 moves the expander speed between 600 and 3600 rpm. When the measured generator power exceeds 98 % of its 120 kW rating the drive caps the speed at $N(0.98P_{g,max})/P_{gen}$, and the loop's

upper limit follows that cap. While the protection has the expander isolated, PIC-1 keeps holding the evaporating pressure through the vapor bypass, whose opening is set to pass what the expander would swallow at the speed the loop asks, $y_b = (\hat{\rho} V_d N + A_L \sqrt{\hat{\rho}\Delta\hat{p}})/(A_b \sqrt{\hat{\rho}\,\min(\Delta\hat{p}, 0.5\hat{p}_e)})$ with $\hat{\rho}$ the vapor density at the measured inlet state, and the drive idles at that speed so that readmission hands the same flow back (a flying start). On a trip or an unload the bypass opens fully and the drive stops. PIC-2 moves the cooling-water flow between 10 and 100 %. The high-pressure override, available in the cascade, opens the source bypass when the evaporating pressure passes 17 bar and so removes heat at the source rather than fighting it with the expander.

The application keeps three models apart. The plant is everything in Section 5.3. No model is identified for the loop structures, and the gain a learner reads from case 1's step, about −24 K per kg/s at fixed expander speed, is the learner's. The loops' only internal models are the steady energy balance of Equation (5.8) and the bypass characteristic, both on the plant's own property table and parameters, so they meet mismatch only when a case changes the plant, as fouling does in case 6. The predictive controller of Section 5.4.4 is the exception, because its model is identified once at the design point and then held fixed.

5.4.4 Model Predictive Control

A third architecture replaces the superheat controller and PIC-1 with one model predictive controller (MPC), the option of Section 1.6.3. Every $T_s = 1$ s it plans the flow setpoint $\dot{m}_{SP}$ of the inner flow loop and the expander speed N over $N_p = 120$ steps in 12 move blocks, whose lengths grow from one step for each of the first four to 75 steps for the last, and it holds the first move until the next plan. Its flow request takes the superheat controller's place in the minimum selector of Equation (5.9), so the overrides, the drive's power cap, PIC-2, the inner loop and the protection act exactly as under the cascade. Industrial predictive controllers are usually arranged in this way, writing setpoints to regulatory loops [18], and Volume 4, Chapter 10 plans the bypass and both inner-loop setpoints of a closed carbon dioxide power cycle in the same way.

The model was identified by step tests on the simulator and is fixed in the source, so it reads neither the plant nor a run's parameters. From case 1 lined out with the declared parameters and no noise, with the inner loop in Auto under a constant setpoint, PIC-1 in Manual and PIC-2 in Auto, each input was stepped up and down by 0.06 kg/s of flow setpoint, 60 rpm, 5 % of source flow or 3 K of supply, and the mean response of the measured superheat and PT-1 over 300 s was fitted by linear least squares [19], as Equation (5.10) states.

$$\begin{aligned} \hat{y}(t) = \sum_{j=1}^{4}\sum_{i=1}^{3} \Big[& \alpha_{ij}\big(1 - \mathrm{e}^{-t/\tau_i}\big) \\ & + \beta_{ij}\Big(1 - \big(1 + \tfrac{t}{\tau_i}\big)\mathrm{e}^{-t/\tau_i}\Big)\Big]\Delta v_j \\ v = \big(\ln\dot{m}_{SP},\ \ln N,\ \ln\dot{m}_s,\ T_{s,in}\big), \quad & \tau_i = 2,\ 8,\ 30\,\mathrm{s} \end{aligned} \tag{5.10}$$

Here $\hat{y} = (SH, p_e)$, each α_{ij} and β_{ij} holds one coefficient per output, and the thermowell, the inner loop, the drive and PIC-2 lie inside the fit, whose RMS residuals are 0.09 to 0.45 % of each response. The steady gains include −25.43 K and 1.298 bar per kg/s of flow setpoint and 0.022 06 K and −0.003 538 bar per rpm. Identified again at 70 % source flow, the gains per physical unit change by factors of 1.07 to 1.74 but those per unit of the logarithm only by 0.87 to 1.22, because the superheat follows the heat that each kilogram receives. Each coefficient is therefore scaled by its input's design value, 3.909 kg/s, 2145 rpm or 5 kg/s. The 24 lag states and two random-walk disturbances on $\ln\dot{m}_{SP}$ and $\ln N$ are discretized with an exact hold at T_s. Against five unfitted steps from lined-out points at 100, 70 and 60 % source flow, the open-loop superheat error is 0.23, 0.60 and 1.17 K RMS.

Each plan minimizes Equation (5.11), the objective of Equation (1.14) with move blocks and soft output limits, over the block values of v_1 and v_2, taken as deviations from the commands at the

estimator's anchor.

$$\begin{aligned}\mathscr{J} = &\sum_{k=1}^{N_p}\Big[q_{SH}\big(\widehat{SH}_k - SH_{SP}\big)^2 + q_p\big(\hat{p}_{e,k} - p_{e,SP}\big)^2\Big] \\ &+ \lambda\sum_{j=1}^{12}\Big[\Big(\frac{\Delta v_{1,j}}{s_1}\Big)^2 + \Big(\frac{\Delta v_{2,j}}{s_2}\Big)^2\Big] + \sum_{o=1}^{2}\big(\rho\,\varepsilon_o^2 + \sigma\varepsilon_o\big) \qquad (5.11)\\ \text{s.t. } &\widehat{SH}_k \geq \underline{SH}_k - \varepsilon_1, \quad \hat{p}_{e,k} \leq \overline{p}_k + \varepsilon_2, \quad \varepsilon_o \geq 0 \\ &\dot{m}_{min} \leq \dot{m}_{SP,j} \leq \dot{m}_{max}, \quad N_{min} \leq N_j \leq N_{lim}\end{aligned}$$

The weights $q_{SH} = 1\,\mathrm{K}^{-2}$ and $q_p = 100\,\mathrm{bar}^{-2}$ are the inverse squares of the loops' 2 K and 0.2 bar trend bands and were fixed before tuning. The move scales $s_1 = 0.010$ and $s_2 = 0.021$, about 0.04 kg/s and 45 rpm, each shift the superheat by about 1 K, and $\lambda = 0.03$. The bounds $\underline{SH}_k$ and $\overline{p}_k$ are the setpoints of the two overrides, $SH_{min} = 5\,\mathrm{K}$ and $p_{max} = 17\,\mathrm{bar}$, read as the overrides read them, and both are soft, with one slack per output priced at $\rho = 10^4$ per unit squared and $\sigma = 10^3$ per unit. A limit that the estimate already violates is approached along a funnel, $\underline{SH}_k = SH_{min} + (\widehat{SH}_0 - SH_{min})\mathrm{e}^{-kT_s/\tau_f}$ with $\tau_f = 20\,\mathrm{s}$, and the ceiling mirrors it. The flow setpoint is bounded hard by $\dot{m}_{min} = 0.25\,\mathrm{kg/s}$ and the superheat controller's own ceiling, and the speed by $N_{min} = 600\,\mathrm{rpm}$ and $N_{lim} = \min(3600\,\mathrm{rpm}, N_{cap}, 0.98NP_{g,max}/P_{gen})$, the drive's power cap of Section 5.4.3 and its formula applied before the cap engages. Each decision is one receding-horizon quadratic program [20], which a dual active-set solver [21] completes in a mean of 1.6 ms in a headless engine.

A steady-state Kalman filter for the 26-state model makes the plan offset-free [22], with measurement variances of $(0.055\,\mathrm{K})^2$ and $(0.005\,\mathrm{bar})^2$ and random-walk intensities $\gamma r/G^2$, where G is the gain of the flow on the superheat or of the speed on the pressure, r the variance of that output, and $\gamma = 0.1$. It runs under every architecture on the mean commands of the last second and on FT-2 and TT-2, stops while the expander is out and anchors again on its return. The plan holds the supply at TT-2 and the source flow at FT-2, the feedforward's own signal, unless the preview adds the case's announced source-flow changes at the declared 5 kg/s per 100 %. With the preview on, the MPC is therefore told in advance what the cascade can only measure, and no shipped option gives the cascade the schedule. In case 4 and the challenge the scheduled change is the source loss itself.

The values of λ and γ came from one pass on scenario 9 over 11 values of λ and seven of γ. Among settings with no rejected plan, trip, intervention or extra travel on either handle, it scored the mean of the superheat and pressure IAEs relative to the cascade with feedforward and overrides, and it took the largest λ within 10 % of the best score, then the smallest γ. The rated-speed term of N_{lim} was added after this pass as a correction, and repeating the pass with it selects the same values. No case is retuned.

The MPC acts only with the superheat and pressure loops in Auto and the expander in service. Otherwise, or when a plan fails to reach its optimum, holds a value that is not finite or leaves its bounds, the superheat controller and PIC-1, which track the applied commands every scan, take over, the historian records an MPC FALLBACK mark, and control returns with the first valid plan. With the noise off, three forced failures at 104 s of scenario 9 move the flow setpoint by at most 0.000 65 kg/s and the speed by 0.29 rpm per scan, against 0.003 71 kg/s and 1.30 rpm just before. Every protective action remains shipped code. After an unannounced loss of source flow to 20 % at 130 s in case 4, both structures isolate the expander on the first measured superheat below 3 K, at 183.30 and 192.30 s, and the MPC solves no plan until it acts again 0.5 s after the readmission.

5.5 What Structure Recovers and What the Plant Withholds

Table 5.4 lists the ten cases with the requirement that credits each one, in the scenario form of Section 1.6.4. Every result below was measured by a script that runs the shipped engine headless at the default seed. It loads a case as the case menu does, applies a named setting ahead of the

Table 5.4 The Ten Cases, Their Events and the Requirement That Completes Each

#	Case, controller and events	Lesson	Required to complete
1	Manual, every loop held; learner steps the pump	which signal moves first	After 10 s raise the pump to at least 64.2 %, hold it for 150 s, and confirm when the pressure, superheat and net power moved.
2	Independent PI 0.02, 90 s; SP 10 to 7 to 10 to 7 K at 60, 360, 780 s; noise ×3 and τ_{TT} 3 to 8 s at 660 s	tuning for the measurement	Set $K_c \geq 0.05$, hold the true superheat within 0.5 K of setpoint for 30 s between 420 and 540 s, run past the events, and confirm the comparison of IAE and travel.
3	Cascade, slow inner loop, PIC-1 in Manual; p_e SP 12.9 to 13.4 bar at 300 s and back at 600 s	loop interaction	Put PIC-1 in Auto before 300 s, see the superheat dip below 8 K, raise the inner gain to at least 5, and confirm the coupling.
4	Cascade, feedforward off; source flow to 70 % at 120 s, back at 480 s	feedback against feedforward	See the expander isolated with the feedforward off, switch it on and Reset, and keep the expander admitted with no intervention past 300 s.
5	Suggested cascade; cooling water 25 to 35 °C over 60 to 180 s	condensing capacity	See PIC-2 LIMITED after 150 s, raise its setpoint to at least 2.59 bar, hold net power above 60 kW for 60 s, and confirm the trade.
6	Cascade with feedforward, 0.04, 30 s; source side fouls to 60 % over 60 to 360 s	changed plant or bad tuning	Run past the fouling, see accepted heat below 870 kW with the superheat within 3 K for 30 s after 420 s, and confirm the reading of the temperatures.
7	Suggested cascade; condenser water becomes the user's 50 °C return at 60 s	useful heat costs work	After the switch raise the condensing setpoint to at least 4.19 bar, deliver useful heat for 60 s, and confirm the penalty.
8	55 °C server water at 20 kg/s, p_e 2.7 bar, SH 7 K; cooling water 15 to 22 °C over 120 to 300 s	when parasitics win	See net power below zero for 30 s after 120 s, unload the cycle and hold net above −1 kW for 30 s, and confirm the Carnot comparison.
9	MPC (Section 5.4.4) over the suggested cascade, SH_{SP} 7 K, schedule preview; source flow to 60 % at 120 s, back at 480 s, both announced	predictive against feedback control	Run to 600 s under the MPC, Reset, switch to the cascade before 120 s, run to 600 s again, and confirm the comparison of lowest superheat, time below 5 K and net energy.
C	Challenge, independent PI 0.03, 60 s: source flow to 70 % at 120, cooling water 25 to 32 °C over 480 to 540 s, pump delivery 70 % at 840 s	the whole problem	Run 1200 s with no trip, the true superheat within 5 to 20 K from 60 to 1200 s, back within ±2 K of setpoint no later than 120 s after each event, $p_e \leq 19$ bar, and the receiver at 15 to 85 % at the end.

first integration step, which makes it an initial condition, applies timed actions at their instants, and advances the plant at its own 0.05 s step. Each run starts from a private copy of the case. Integrated absolute error (IAE) is taken on the true superheat at every physics step over the stated window, and recovery is the last instant in the window at which the true superheat lies outside the stated band about its setpoint, after which it stays inside to the window's end. Pump travel is the total variation of the commanded pump speed, and stored inventory is the receiver's true liquid level at the window's start and end.

5.5.1 The Dynamic Signature

Case 1 shows the coupling before any loop closes. Raising the pump by two points adds 0.122 kg/s, the measured evaporating pressure moves 0.80 s later because the expander passes a fixed volume per turn, the measured superheat starts to fall after 2.55 s, and the true superheat covers 90 % of its change in 48.5 s and settles at 7.05 K, a gain of −24.1 K per kg/s. Warming the cooling water from 25 to 30 °C then raises the condensing pressure from 1.969 to 2.290 bar and lowers net power from 78.86 to 74.04 kW, and a five-point pump step reaches the 3 K intervention after 34.05 s. In case 3 the evaporating-pressure setpoint rises by half a bar, which lifts the saturation temperature by 1.85 K at once, so with PIC-1 in Auto the true superheat falls to 6.44 K although the pressure loop never touches the pump. The IAE after the step is 549 K s with a slow inner flow loop and 113 K s with the inner loop tuned first, against 7 K s with PIC-1 left in Manual. Small steps confirmed every loop's sign, the pump's at 13 of 13 lined-out points and the expander's at all five points below the generator's rating.

🕮 **Pause.** Try Exercise 5.2 before reading on.

Table 5.5 Matched Comparisons on the Simulator's Truth. Recovery Is the Last Exit From the Band About the Setpoint Within the Window. Limited Is the Time the Expander Was Isolated (A) or PIC-2 Was LIMITED (B). Energies Are Over the Window, the Receiver Level at Its Start and End

Setting	IAE (K s)	Max dev. (K)	Recovery (s)	*SH* range (K)	Limited (s)	Pump travel (%)	Net (kWh)	Accepted (kWh)	Receiver (%)
A. Case 4, 120–480 s, ±2 K									
independent PI, 0.08, 20 s	276.7	8.17	50.0	1.83–10.25	54.9	42.9	4.755	72.25	50.2 → 50.0
cascade, no feedforward	437.2	8.91	76.0	1.09–10.14	83.4	126.7	4.187	72.47	50.2 → 50.0
cascade with feedforward	29.0	1.08	0	9.39–11.08	0	129.4	5.819	71.83	50.3 → 50.0
MPC, preview off	27.9	0.75	0	9.25–10.24	0	126.6	5.812	71.82	50.3 → 50.0
MPC, loss announced	24.6	0.68	0	9.49–10.68	0	129.0	5.812	71.81	50.0 → 53.2
A. Case 4, 480–900 s, ±2 K									
independent PI	261.4	9.07	54.3	9.87–19.07	0	53.2	8.939	111.92	50.0 → 50.2
cascade, no feedforward	254.8	9.30	48.1	9.91–19.30	0	181.6	8.934	111.93	50.0 → 50.2
cascade with feedforward	38.0	1.08	0	8.92–10.61	0	186.4	8.938	112.58	50.0 → 50.3
MPC, preview off	36.9	1.36	0	9.70–11.36	0	185.2	8.941	112.64	50.0 → 50.2
MPC, loss announced	37.0	1.31	0	9.63–11.31	0	184.2	8.944	112.64	53.2 → 50.2
B. Case 5, 240–900 s, ±2 K									
p_c SP held at 1.95 bar	22.9	0.16	0	9.89–10.16	660	280.9	8.689	171.04	50.7 → 50.7
p_c SP 2.6 bar from 240 s	22.7	0.16	0	9.89–10.16	0	281.4	12.235	169.99	50.7 → 51.1
p_c SP 2.8 bar from 240 s	223.0	0.79	0	9.46–10.79	0	285.3	12.596	169.19	50.7 → 51.6
C. Case 2, 360–660 s, ±0.5 K, before the sensor change									
0.02, 90 s (shipped)	286.0	2.02	>300	7.98–9.40	0	6.4	6.512	—	—
0.04, 30 s	125.5	3.17	85.0	6.83–10.06	0	12.9	6.465	—	—
0.08, 20 s	50.8	3.34	31.5	6.66–10.17	0	26.2	6.444	—	—
C. Case 2, 780–1080 s, after it									
0.02, 90 s (shipped)	416.0	2.54	>300	—	0	19.2	—	—	—
0.04, 30 s	116.7	3.00	71.0	—	0	38.1	—	—	—
0.08, 20 s	64.9	3.04	47.1	—	0	76.2	—	—	—
D. Case 6, 360–900 s									
0.08, 20 s (clean-plant tuning)	844.0	2.92	—	7.86–12.92	0	247.9	10.193	125.83	—
0.04, 30 s (shipped)	68.6	0.75	—	9.74–10.75	0	192.0	10.213	125.23	—

A uses the suggested gains in every row (SIC 0.08 kg/s per K and 20 s, PIC-1 800 rpm/bar and 5 s), and the second row is case 4 as shipped. The MPC rows (Section 5.4.4) run over the same cascade with scenario 9's settings, and with the preview on the case's scheduled loss is announced to the plan. B and D run the cascade with the feedforward. C runs the independent architecture, and its windows are the 300 s after each 3 K setpoint step. A dash marks a quantity not computed for that window, and >300 a run still outside ±0.5 K at the window's end. Halving the physics step to 0.025 s gives IAEs of 276.66 and 29.01 K s in A's first window and 261.48 and 38.02 K s in its second.

5.5.2 Three Matched Comparisons

Each group in Table 5.5 shares the initial state, the seed, the instruments, the limits and the events, and differs only in the controller settings named. No load is served by this plant, so unmet and curtailed service are not applicable, and the accepted heat takes the place of fuel.

Comparison A tests nominal disturbance rejection on case 4's loss of 30 % of the source flow. The feedforward reads the lost heat from FT-2, TT-2 and TT-3 and lowers the flow setpoint within seconds, while either feedback structure waits for the thermowell, as shown in Figure 5.2a. Independent PI at the suggested gains lets the true superheat fall to 1.83 K, and the protection isolates the expander from 138.05 to 193.00 s. The case's own cascade without the feedforward does worse, 437.2 against 276.7 K s, mostly because it is readmitted later, at 220.75 s. Its low-superheat override shares the lagging measurement, takes the flow setpoint about 15 s after the loss, holds it for 41 s, and cannot prevent the intervention. The cascade with the feedforward stays within 1.08 K of setpoint with no intervention and makes 5.819 kWh against 4.755 and 4.187 kWh over the same six minutes. After the restoration at 480 s feedback lets the superheat reach 19.07 K and the feedforward 10.61 K, while all three runs keep the expander and their energies agree to 0.005 kWh. The feedforward costs actuator travel, 129.4 against 42.9 % of pump speed, mostly because the inner flow loop answers FT-1's noise every 0.25 s, as the first 120 s show with no event at all (48.4 against 9.0 %). On the pressure–enthalpy

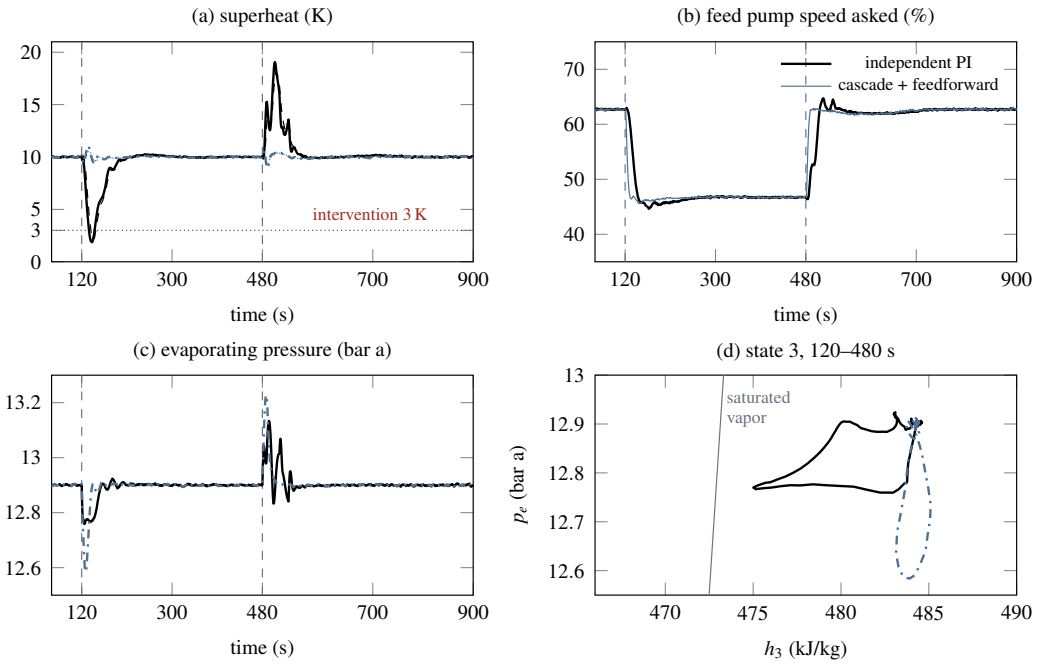


Figure 5.2 Comparison A, case 4, from the exported records: independent PI (black, solid) and the cascade with the source-heat feedforward (blue, dash-dotted) at the same gains, the source-loop flow falling to 70 % at 120 s and restored at 480 s (gray dashed verticals). (a) True superheat (thick) and the TT-1 and PT-1 measurement (thin, dashed for independent PI and dotted for the feedforward) against the 3 K intervention (dotted horizontal). (b) Commanded pump speed, independent PI thick and the feedforward thin. (c) True evaporating pressure. (d) The expander inlet, state 3, on the pressure–enthalpy plane from 120 to 480 s beside the saturated-vapor line (thin gray).

plane of Figure 5.2d the feedback run's inlet state slides to 475.0 kJ/kg, 1.9 kJ/kg from saturated vapor, and loops back at a lower pressure while the bypass holds the evaporator, and the feedforward run's stays within 1.2 kJ/kg of the design enthalpy. Figure 5.3 draws the first window of this comparison as bars.

> **TRY IT IN THE APP** **4 · Lose Source Flow: Feedback or Feedforward**
>
> Load the scenario and press Run. At 120 s the source flow falls to 70 %. Watch the superheat and the protection, which isolates the expander 17.3 s after the loss and keeps it out for 83.4 s. Then set Source-heat feedforward (cascade) to On, press Reset and run again. The true superheat stays between 9.39 and 11.08 K, and the expander never leaves the line.

Pause. Try Exercise 5.1 before reading on.

The second comparison is a capacity limit. In case 5 the cooling water warms from 25 to 35 °C over two minutes, PIC-2 reaches 100 % flow, and the condensing pressure floats to 2.383 bar, where the fans draw 30.0 kW and the cycle nets 47.47 kW, as shown in Figure 5.4a and b. No tuning of PIC-2 removes that limit, and the loop reads LIMITED for all 660 s of the window. Raising its setpoint to 2.6 bar at 240 s gives up generator energy, 15.303 against 15.807 kWh, to save cooling energy, 1.467 against 5.500 kWh, and nets 12.235 against 8.689 kWh. Lined out, the net power is 47.47 kW at the limit, 67.55 kW at 2.6 bar and 68.76 kW at 2.8 bar. The higher setpoint adds only 0.361 kWh over the window, and its larger move disturbs the superheat, an IAE of 223.0 K s against 22.7, because the

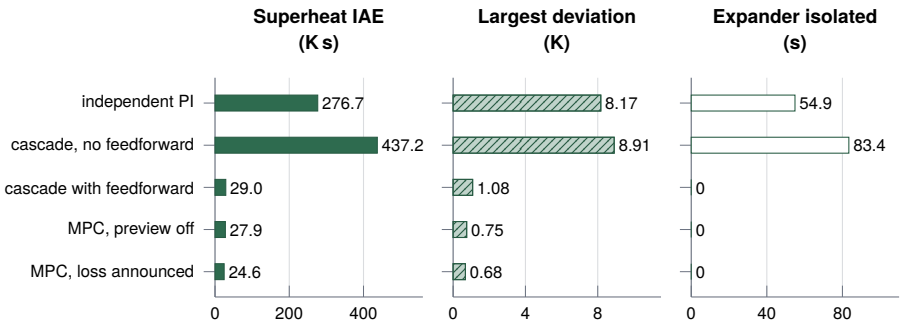


Figure 5.3 Comparison A of Table 5.5 redrawn: case 4 from 120 to 480 s, the source flow at 70 %, the PI and cascade rows at the suggested gains and the MPC rows with scenario 9's settings. Bars give the integrated absolute error of the true superheat, its largest deviation from setpoint and the time the protection kept the expander isolated. Notice that the two structures that wait for the thermowell lose the expander, while every row that acts on the source meters keeps it, with an error below 30 K s.

cooling power left to save shrinks with the cube of the flow. The superheat barely notices the sink in either of the first two runs (0.16 K at most), which is why the challenge's warm-water event costs power but not band time.

> **TRY IT IN THE APP** **5 · A Hot Day at the Condenser**
>
> Run the scenario as loaded. The cooling water warms from 25 to 35 °C, PIC-2 reads LIMITED with the cooling water at full flow, and the net power lines out near 47.47 kW. Run again and, at about 240 s, raise the Condensing pressure setpoint to 2.6 bar. PIC-2 leaves LIMITED, the cooling power falls, and the net power lines out near 67.55 kW.

Pause. Try Exercise 5.3 before reading on.

The third comparison is a sensor mismatch. Case 2 slows the thermowell from 3 to 8 s and triples the noise at 660 s, and the three tunings are compared over the 300 s after equal 3 K setpoint steps before and after the change. The fast tuning, 0.08 kg/s per K with 20 s, goes from 50.8 to 64.9 K s and from 26.2 to 76.2 % of pump travel, the moderate one from 125.5 to 116.7 K s with travel rising from 12.9 to 38.1 %, and the cautious shipped tuning never settles within 0.5 K in either window, as shown in Figure 5.4e and f. Over the longer window from 780 to 1200 s that the case's own texts use, the three IAEs are 493.5, 122.3 and 71.9 K s. The faster loop still tracks best, but its margin over the moderate one shrinks from 75 K s to 52 K s while its travel nearly triples. The likely cause is the phase margin that an 8 s lag removes from a loop whose plant answers in tens of seconds, while the tripled noise reaches the pump directly. A model mismatch gives the same lesson in case 6. When the source side fouls to 60 % of clean, accepted heat falls from 960.5 to 835.2 kW and the source leaves at 111.16 °C instead of 105.33 °C. The clean-plant tuning of 0.08 kg/s per K with 20 s then cycles between 7.86 and 12.92 K with a period of about 34 s, an IAE of 844.0 K s, while the shipped 0.04 kg/s per K with 30 s holds 9.74 to 10.75 K at 68.6 K s, and both net the same energy within 0.02 kWh. The fouling is a changed plant that the temperatures reveal.

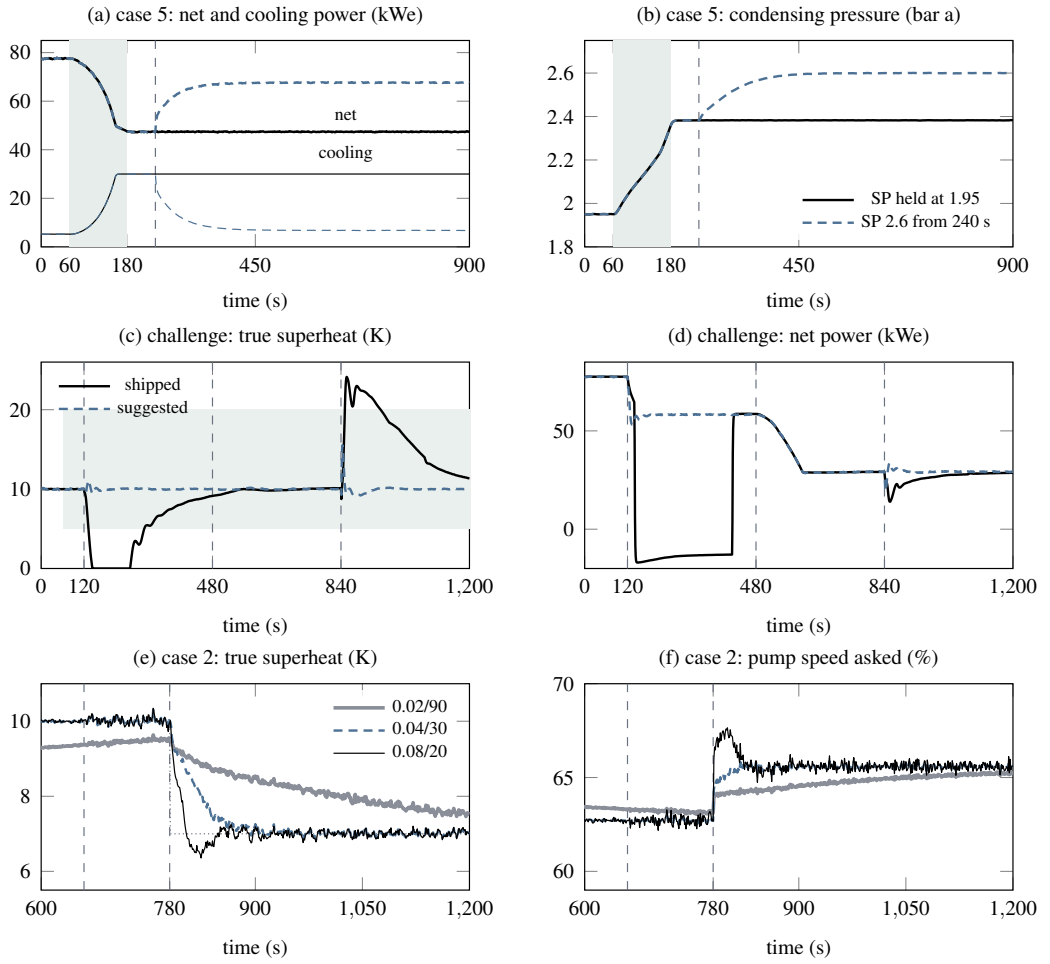


Figure 5.4 Comparisons B and C and the challenge, from the exported records. (a) Case 5, net power (thick) and cooling power (thin) with the condensing setpoint held at 1.95 bar (black, solid) and raised to 2.6 bar at 240 s (blue, dashed), the shaded band the cooling-water ramp from 25 to 35 °C. (b) The same runs' true condensing pressure. (c) The challenge's true superheat as shipped (black, solid) and with the suggested settings (blue, dashed) against the shaded 5 to 20 K band, the events at 120, 480 and 840 s as gray dashed verticals. (d) The challenge's net power, the same line styles. (e) Case 2, true superheat for the three tunings after the thermowell and noise change at 660 s and the setpoint step at 780 s: 0.02 and 90 s very thick gray, 0.04 and 30 s blue dashed, 0.08 and 20 s thin black, the setpoint dotted. (f) Their commanded pump speed.

> **+ DID YOU KNOW? The Calmer Tuning Wins on a Fouled Plant**
>
> From the simulator: with the source side fouled to 60 % of clean, the clean-plant tuning of 0.08 kg/s per K and 20 s cycles between 7.86 and 12.92 K, an integrated error of 844.0 K s. The calmer shipped tuning holds 9.74 to 10.75 K at 68.6 K s (Table 5.5).

5.5.3 Where Added Control Cannot Help

Case 8 runs the same machine on 55 °C server water at 20 kg/s with an evaporating pressure of 2.7 bar. At 100 s the cycle accepts 128.5 kW and its generator makes 3.29 kW, 2.56 % of the heat, at a

Table 5.6 The Challenge Over 1200 s. No Configuration Trips, Peaks Above 13.43 bar or Ends Outside 50.0 to 50.3 % of Receiver Level

Configuration	Out (s)	*SH* range (K)	Recovery (s)	Isol. (s)	IAE (K s)	Net (kWh)	Verdict
shipped, PI 0.03/60	256.7	0.00–24.14	266/0/305	274.2	4311	9.16	fails
PI 0.08/20	26.6	3.64–22.70	55/0/58	0	640	14.89	fails
PI 0.08/20, PIC-1 800/5	20.7	1.83–18.61	50/0/52	54.9	629	13.79	fails
cascade, no feedforward	39.6	1.09–15.57	76/0/9	83.4	521	13.27	fails
cascade + feedforward, shipped gains	0	6.77–18.89	42/0/16	0	395	14.92	passes
suggested	0	7.91–15.56	0/0/9	0	134	14.90	passes
MPC, loss announced	0	6.64–16.01	0/0/30	0	160	14.89	passes
MPC, preview off	0	6.64–16.01	0/0/30	0	161	14.89	passes

Out is the time outside 5 to 20 K from 60 to 1200 s. Recovery is the last exit from ±2 K after the events at 120, 480 and 840 s, allowed 120 s, and the shipped run is late twice. IAE is from 60 to 1200 s. The cascade rows use the suggested gains except the fifth. The MPC rows switch only the architecture of the challenge as loaded, and over the suggested settings the first is unchanged.

pressure ratio of 2.45, which leaves 1.02 kW after 2.28 kW of parasitics. As the cooling water warms from 15 to 22 °C, the condensing loop asks for more flow, net power turns negative at 204.3 s, and with the fans at full flow from 389 s it reaches −30.6 kW. The Carnot bound between the server water and the cooling water falls from 12.19 to 10.06 % over the ramp. A sweep of 192 steady settings at 20 °C cooling water, the evaporating pressure from 1.8 to 3.2 bar, the superheat from 3 to 12 K and the condensing pressure from 0.9 to 1.6 bar, solves every setting and finds no positive net output, the best being −0.59 kW. Unloading the cycle leaves −0.33 kW of standby load, so the correct operation is to stop, because the temperature head cannot pay for the pumps and fans. Routine server heat would not be sent to a power cycle, but the stylized case shows where parasitic power overtakes a vanishing temperature difference. The cooling of a data hall is itself the subject of Volume 9, Chapter 8.

The generator's rating sets a second limit at the other end of the source range. A step of the supply from 150 to 170 °C with the suggested settings caps the drive for 511 s, the generator ends at 117.7 kW, the evaporating pressure peaks at 17.12 bar, and the high-pressure override opens the source bypass by up to 4.4 %, with the superheat no lower than 3.57 K and no intervention. Without the feedforward the pressure peaks at 17.32 bar and the bypass opens to 9.5 %. Control keeps the plant inside its limits in both runs, but the heat the generator cannot convert is sent around the evaporator.

5.5.4 The Challenge

The challenge starts at the design point with the shipped independent PI at 0.03 kg/s per K and 60 s. It loses 30 % of the source flow at 120 s, warms the cooling water from 25 to 32 °C over 480 to 540 s, and wears the pump to 70 % of its rated delivery at 840 s, and Table 5.6 lists eight configurations. As shipped the expander is isolated from 139.55 to 413.75 s. The true superheat reaches 0 K at 144 s, while the closing isolation valve still admits 1.4 % and the bypass carries the wet vapor, and the wear later lifts it to 24.14 K. Each of the two scored events needs its own structural element. Faster tuning of the independent loops leaves 20.7 to 26.6 s outside the band, the cascade without the feedforward fixes the wear but not the source loss, and the cascade with the feedforward passes even at the shipped gains. The suggested settings keep the superheat between 7.91 and 15.56 K and recover within 8.7 s of the wear, as shown in Figure 5.4c and d. The warm water costs every configuration the same power, with net output at the end between 28.69 and 29.23 kW and the cooling water at its full flow. Over five further seeds the shipped run spends 256.0 to 256.8 s outside the band and fails each time, while the suggested run stays within 7.90 to 15.60 K and passes each time. Figure 5.5 shows the suggested run at 1200 s.

5.5.5 Predictive Control Through a Scheduled Turndown

Scenario 9 runs the superheat at 7 K, 2 K above the floor of the low-superheat override, and the on-site engine's announced schedule takes the source flow to 60 % at 120 s and back at 480 s. The thinner margin makes a little more power, and the turndown is deeper than case 4's. The case loads under the

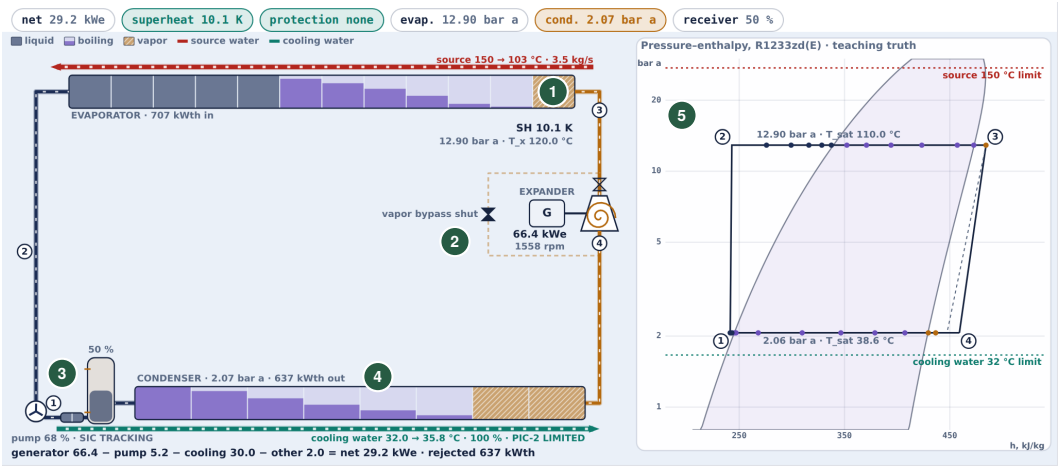


Figure 5.5 End of the suggested challenge run, $t = 1200$ s, on the Cycle view, cropped to the plant. (1) The hatched vapor cells at the evaporator outlet, where the superheat reads 10.1 K at 12.90 bar with the expander inlet at 120.0 °C. The evaporator accepts 707 kW from source water cooled from 150 to 103 °C at 3.5 kg/s. (2) The expander, its generator at 66.4 kW and the vapor bypass shut. (3) The receiver and the worn pump, running at 68 % under the flow loop. The flow setpoint of 2.93 kg/s that it delivers is read on the control panel, outside this crop. (4) The condenser, whose cooling water at 32.0 °C is at full flow, so PIC-2 is LIMITED and the condensing pressure reads 2.07 bar. (5) The cycle on the pressure–enthalpy plane. The bottom line gives the generator's 66.4 kW less 5.2 kW for the pump, 30.0 kW for cooling and 2.0 kW of auxiliaries, 29.2 kW net. The app's lead tile, also outside this crop, reports 0 s of 1140 s outside 5 to 20 K, the zero of Table 5.6.

Table 5.7 Scenario 9 Over 900 s, Default Seed. Upper Block as Scheduled, Lower Block With the Plant's Turndown Moved While the Announced Schedule Stays

	Cascade 7 K	Cascade 8.5 K	MPC, preview off	MPC, schedule
Lowest true *SH* (K)	2.43	5.46	5.11	5.80
at (s)	149	148	133	134
Time below 5 K (s)	21.0	0	0	0
SH IAE (K s)	275.0	174.2	135.9	122.2
p_e IAE (bar s)	16.15	14.17	6.35	5.88
Net energy (kWh)	16.848	16.727	16.857	16.857
Lowest true SH (K) / time below 5 K (s)				
10 s early	2.44/18.55	—	5.26/0	4.74/3.20
30 s early	2.45/20.90	—	5.10/0	4.91/4.00
30 s late	2.39/17.25	—	5.16/0	5.24/0
to 50 %, not 60	0.00/86.10	—	4.64/8.65	5.37/0
never comes	6.80/0	—	6.75/0	4.88/1.65

Both cascades run the suggested gains with feedforward and overrides, and 8.5 K is the lowest setpoint in steps of 0.5 K that keeps the floor. The cascade at 0.00 is isolated once.

MPC of Section 5.4.4, and Table 5.7 compares it over 900 s at the default seed with the strongest shipped structure, the cascade with feedforward and overrides at the suggested gains. Calmer cascade tunings fall further, to 3.69 K at 0.04 kg/s per K and 30 s and to 4.00 K at 0.02 and 60 s, and without the feedforward the expander is isolated and the plant trips. Each MPC run solves 900 plans and rejects none.

The MPC with the schedule keeps the true superheat at or above 5.80 K, reached at 134 s, while the cascade falls to 2.43 K at 149 s and spends 21.0 s below the floor, as shown in Figure 5.6a. The

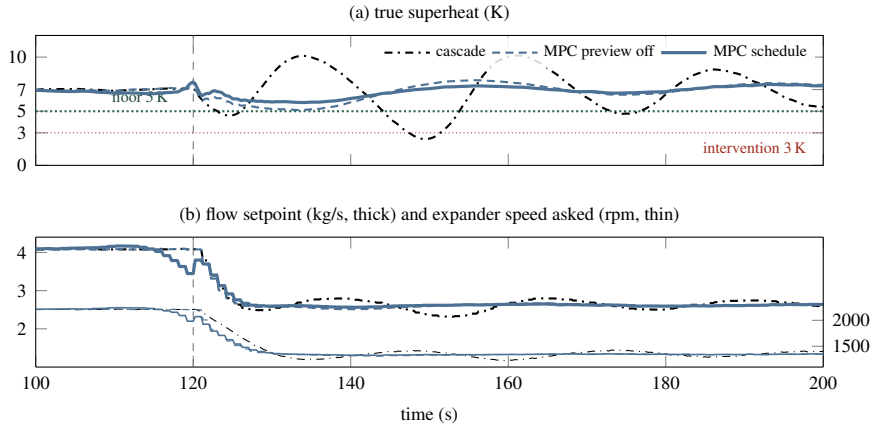


Figure 5.6 Scenario 9 from 100 to 200 s, default seed, the scheduled turndown at 120 s gray dashed: the cascade with feedforward and overrides (black, dash-dotted), the MPC with the schedule (blue, solid) and with the preview off (blue, dashed). (a) True superheat against the 5 K floor and the 3 K intervention (dotted horizontals). (b) Flow setpoint (thick, left axis) and expander speed asked (thin, right axis).

cascade's measured superheat stops at 3.20 K, 0.20 K above the intervention. Its first trough, 4.58 K at 124.5 s, follows the feedforward's filtered reading of FT-2, and the deeper one is its superheat loop swinging back while PIC-1 restores the pressure. The superheat IAE is 122.2 against 275.0 K s and the pressure IAE 5.88 against 16.15 bar s. Over ten noise seeds the lowest values span 5.73 to 5.86 K under the MPC and 2.23 to 2.43 K under the cascade, and the MPC's 5.80 K is unchanged when the run is extended to 3600 s. Part of the margin is information. With the preview off, given only the source flow that the cascade's feedforward also reads, the MPC still holds 5.11 K at an IAE of 135.9 K s, and that is the like-for-like comparison. The schedule is worth 0.69 K of the low point. At the same setpoint the two structures make the same energy, 16.857 against 16.848 kWh. The cascade keeps the floor from a setpoint of 8.5 K, where its lowest value is 5.46 K, and then nets 16.727 kWh, 0.130 kWh or 0.8 % less than the MPC at 7 K over the 15 min. Raised in steps of 0.5 K from 7 to 8.5 K, its setpoint lifts the cascade's lowest superheat through 2.43, 3.43, 4.44 and 5.46 K. Moving two handles together is therefore worth the gap from 2.43 K to the MPC's 5.11 K without a preview, and being told the future is worth the further 0.69 K.

What holds the superheat up is the early action and the two handles moving together, not the floor's price. Told of the turndown, the plan cuts the flow setpoint from 4.16 to 3.45 kg/s and the speed from 2239 to 1982 rpm between 110 and 120 s, before FT-2 moves, and by 130 s it asks for 2.60 kg/s and 1376 rpm, as shown in Figure 5.6b. Because the speed falls with the flow, the evaporating pressure stays within 12.78 to 13.09 bar after 100 s, against 12.45 to 13.39 bar under the cascade. The plan rests on the floor in 127 of its 900 plans, before each scheduled change. With the floor priced at zero, or moved to 3 K, the true low point is still 5.80 K, although the unpriced plan predicts −1.21 K. With the speed held by a move price a million times larger, or the flow held in the same way, the expander is isolated and the true superheat reaches 0 K. The result also rests on the output weights fixed before tuning, because with the pressure weight cut to a hundredth the low point falls to 0.79 K and the expander is isolated.

A wrong schedule misleads the plan, as the lower block of Table 5.7 shows for a plant whose turndown departs from the one announced. A turndown 10 s early takes the true superheat to 4.74 K for 3.20 s, 30 s early to 4.91 K, and a turndown announced but never made to 4.88 K at 482 s, after the plan has acted for a return that does not come. A late or deeper turndown stays above the floor, at 5.24 and 5.37 K. The floor constrains the measured superheat that the plan predicts, which is why in

the 10 s early run the measurement never falls below 5.36 K while the true value reaches 4.74 K.

The other cases load the MPC with scenario 9's settings and no retuning. In case 4 the preview announces the loss, and the MPC holds 9.49 to 10.68 K over it at an IAE of 24.6 K s, against 29.0 K s for the cascade with feedforward and 27.9 K s with the preview off, at the same energy within 0.007 kWh, as Table 5.5 shows, and it halves the pressure IAE over 900 s, 4.19 against 9.29 bar s. Its IAE is 151.6 against 1078.6 K s for case 2's cautious loop and 14.5 against 30.0 K s in case 8. It loses narrowly in case 5, 40.8 against 34.8 K s at equal energy, and in case 7, 163.1 against 147.3 K s. It loses clearly in two places. In case 6 it never settles, swinging between 8.0 and 13.1 K with a period of about 50 s to 3600 s, an IAE of 5583 against 335 K s. At 60 % fouling a small flow step takes 30 s to make 63 % of its superheat change, against 16 s in the model, with a gain per logarithmic unit 15 % higher, so a plan with $\lambda = 0.03$ overreaches, and with $\lambda = 3$, as a diagnostic, the superheat stays within 9.20 to 10.46 K. With case 4's events and the supply raised to 170 °C at 10 s, the superheat settles at 22.0 to 22.5 K from 600 to 1800 s against its 10 K setpoint, an IAE of 14 870 against 965 K s. The generator's rating holds the speed, and with a pressure weight of 100 per bar^2 against 1 per K^2 the plan cuts the flow to hold the pressure, which peaks at 15.35 bar. A weight of 10, as a diagnostic, lowers the IAE to 600 s from 2509 to 806 K s. That run nets 46.86 against 43.89 kWh with no intervention, where the cascade is isolated once, but its true superheat touches 3.34 K while the measurement stays above 5.43 K. The cascade is the better choice for fouling and for the hottest supply.

The challenge is met by the MPC alone, with only the architecture switched and with the preview on or off, as Table 5.6 shows. The superheat stays within 6.64 to 16.01 K and recovers 30.1 s after the pump wear, at an IAE of 160 against 134 K s for the suggested settings and 14.89 kWh of net energy. The challenge's rules are unchanged.

5.5.6 Verification

The numbers in this subsection check the implementation, not control performance. The working-fluid inventory stays at its charge within 3×10^{-12} kg in every run reported here, and the stored energy of fluid, walls and water matches the integrated boundary fluxes within 8×10^{-5} of the accepted heat, as does a separate ledger written from the face fluxes (2.8×10^{-5} in case 4, 5.8×10^{-5} in the suggested challenge). Nearly all of that residual sits in the last condenser cell, which crosses the saturated-liquid line 58 to 95 times per run, and in its neighbor, which crosses it up to four times, while every other cell closes within 0.022 kJ. Halving the physics step to 0.025 s changes comparison A's IAEs by at most 0.06 K s, and the application's own check holds the challenge's scored superheat and pressure within 1 % of their spans. The cell count is the larger numerical sensitivity. Doubling it to 24, 16 and 4 cells raises the design point's accepted heat by 6.60 %, from 959.95 to 1023.28 kW, and its net power by 6.84 %, from 77.55 to 82.86 kW. On case 4 the doubled mesh moves the IAE over the loss from 276.7 to 301.1 K s for independent PI, 437.2 to 445.5 K s for the cascade and 29.0 to 64.8 K s with the feedforward, deepens the feedback runs' dips to 0.41 and 0.00 K, and leaves every intervention count. In the challenge the shipped run's time outside the band becomes 303.3 s instead of 256.7 and the suggested run's range 6.78 to 14.60 K, so no verdict and no ranking changes. A simulated hour costs 23 s in a headless engine, and all 61 of the application's checks pass.

Agreement with literature is a narrower claim. The tables reproduce CoolProp's implementation of the equation of state to the errors of Section 5.3.5, a statement about interpolation only. The shaft isentropic efficiency of 0.655 lies below the 68 % maximum overall effectiveness measured on one scroll expander [15], and the feedforward's reduction of the superheat deviation, from 8.17 to 1.08 K, has the same direction as the 1.9 against 10 K measured on a vehicle cycle [3], but nothing in the model has been checked against data from an operating ORC. No learner study has been made, so this chapter claims nothing about learning gains.

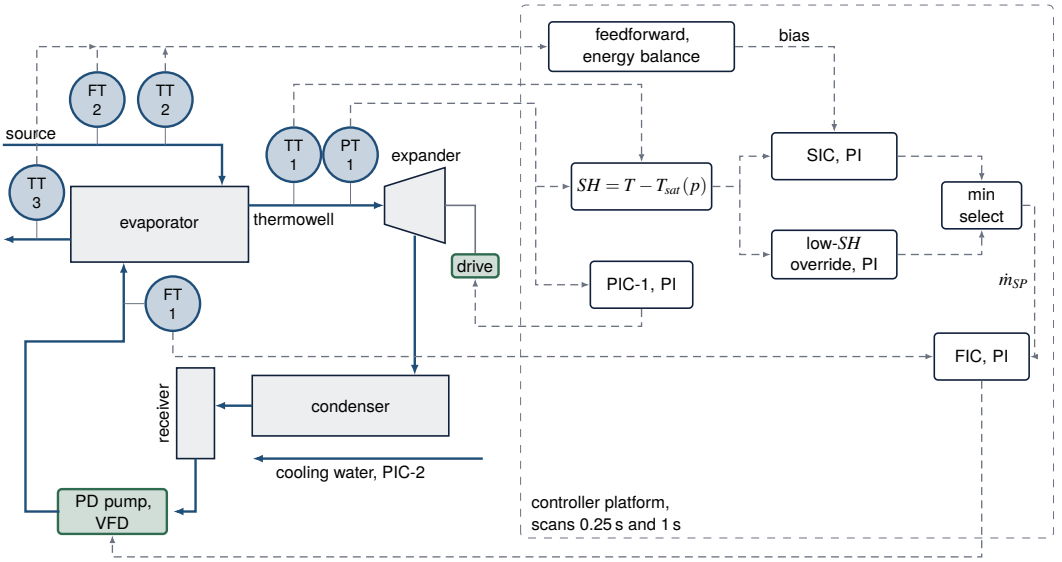


Figure 5.7 The superheat cascade on real hardware. TT-1 reads the expander inlet through its thermowell and PT-1 the evaporating pressure, and the platform computes the superheat from the two. SIC, biased by the feedforward from the source meters FT-2, TT-2 and TT-3, and the low-superheat override meet in the minimum selector, whose output is the flow setpoint of FIC. FIC reads FT-1 and sets the speed of the positive-displacement (PD) feed pump on its variable-frequency drive (VFD), and PIC-1 holds the pressure with the expander's drive. Solid arrows are process streams and dashed gray arrows are signals. Notice that the feedforward reaches the pump without passing the thermowell.

5.6 Technology in Practice

Figure 5.7 draws the superheat cascade of Section 5.4 and its feedforward with their hardware.

What Measures It

The controlled variable is computed rather than read. TT-1 measures the temperature at the expander inlet, PT-1 the evaporating pressure, and the transmitter subtracts the saturation temperature at that pressure. The 3 s lag of TT-1 stands for the thermowell and the 0.3 s lag of PT-1 for the pressure transmitter. A truck engine's cycle had the same data, pressure and temperature pairs between its components, and no sensor inside its heat exchangers [3]. The feedforward needs three instruments more, on the source loop: the flow meter FT-2 and the supply and return thermometers TT-2 and TT-3.

TECHNOLOGY **Sensor: A Resistance Thermometer in a Thermowell, Read With a Pressure**

Superheat is not measured directly. It is computed from a temperature and a pressure through the fluid's saturation properties, as the pressure and temperature pairs between the components of a vehicle cycle are used to estimate its superheat and subcooling [3]. The temperature element is commonly a resistance temperature detector, a fine platinum or copper wire whose resistance rises with temperature and is very linear. It sits in a thermowell, a pressure-tight sheath that lets the element be replaced without depressurizing the line. The sheath's metal adds a first-order lag, and an air gap left at the bottom of the well adds a second lag that can push a loop toward oscillation [23]. On a trend, a slow well shows as a measured superheat that trails every real change.

What Moves It

The superheat moves through the feed pump, whose speed sets the flow because the pump is of the positive-displacement kind. The model gives its drive a 1 s lag. The expander's speed is the second handle, held by PIC-1 through a drive that the model gives a 0.5 s lag and a 300 rpm/s slew, the kind of acceleration limit a motor drive carries [23]. The cooling flow is set by pumps and fans whose power the model takes as the cube of the flow, and the source bypass is the valve of the high-pressure override.

TECHNOLOGY **Final Element: A Positive-Displacement Pump on a Variable-Frequency Drive**

A positive-displacement pump discharges a fixed volume with every revolution of its shaft, so its speed sets the flow [3, 23]. Its motor is run by a variable-frequency drive, which sets an alternating-current motor's speed by the frequency of the power it supplies and its torque by the voltage and current. Such a drive needs no speed feedback to know roughly how fast the motor turns, although precise control at low speed may need a speed sensor. Drives limit current, torque, acceleration and deceleration, and a control system can change those limits over a network while the motor runs [23]. A worn pump delivers less per turn, which the chapter's inner flow loop answers by raising the speed.

Where the Controller Runs

Control loops in a plant are often nested or cascaded, with cycle times from milliseconds to minutes [24]. In the app the flow loop and PIC-1 scan every 0.25 s and the superheat loop every 1 s, against a thermowell lag of 3 s and slowest plant modes of 17 to 47 s. On a fieldbus platform a feedforward needs one more analog input block for the load transmitter, wired to the feedforward input of the PID block [23]. The energy balance of Equation (5.8) would sit in calculation blocks between the source meters and that input. The minimum selector is a control selector block whose back-calculation output tells a de-selected controller to halt its integral action [23], the external reset of Equation (5.9). On the truck engine the model-based feedforward was computed robustly enough for an embedded implementation, under a supervisor that moved the cycle through its stop, idle and production modes [3].

IN PRACTICE **Size the Feedforward, Then Time It**

A feedforward that keeps a mass or energy balance takes its size from that balance. It is applied to the dominant loads only, because compensating every load seldom repays the cost. Its timing is then tested. The engineer measures the lag with which the load moves the process variable and the lag with which the final element moves it, and compares the two. A quicker final element gets a delay or lag on the feedforward signal, so that it does not act too soon, and a slower one gets a lead. Even an imperfect feedforward is often far better than none [23]. This chapter's feedforward is static apart from a 2 s first-order filter.

What Is Modernizing It

Three developments bear on this loop. The first is a supervisory layer, a real-time optimization that hands optimal setpoints to the low-level superheat and pressure controllers. The truck study built

its controllers to serve such a layer [3], and the predictive controller of Section 5.4.4 is one form of it. The second is estimation. The same study added an extended Kalman filter that estimates the evaporator wall temperatures no sensor reads, so that a second controller could track a pressure setpoint with the evaporator bypass while the superheat stayed regulated, shown in simulation [3]. The third is virtual prototyping. A review expects simulation of the whole system, detailed enough for design, to overhaul the traditional sequential and iterative design process, with dynamic simulation bringing transient operation and control into design from its early stages [9].

5.7 Limitations

Every size, conductance, efficiency and threshold is a synthetic teaching value chosen so that the design point closes, not fitted to a plant, and passing a simulated case qualifies no equipment and estimates no savings. Uniform pressure per side omits the evaporator's frictional drop and every pressure wave. The exchangers are not converged in their cell count, so the operating point moves by about 7 % when the cells double, although no ranking does. The expander has the structure of a semi-empirical model without its fitted parameters and with no rotor inertia, and its dry-inlet specification is stricter than the liquid tolerance of many positive-displacement machines. The source loss of case 4 and the challenge falls by exactly 30 % in one step, which would not occur on a real engine loop, but the step isolates the thermowell's lag from every other cause. The protection acts on the measured superheat, which in the shipped challenge still reads 3 K when the true value is about 1.4 K. The engine behind the source loop and the cooler behind the cooling water do not respond to the cycle, and the loops' internal models share the plant's parameters, so mismatch enters only where a case changes the plant.

The predictive controller's model is linear in logarithms about one design point. At 60 % fouling the plant's flow response is about twice as slow and 15 % stronger than the model's, and the MPC never settles in case 6. It holds no model of the overrides, the protection or the condensing side, and its trades are set by output weights fixed before tuning, which at a 170 °C supply leave the superheat at 22 K. It takes the announced schedule as true, and its floor bounds the measured superheat rather than the true one, so a turndown 10 s early takes the true value to 4.74 K. Its gains come from the simulator, not from tests on an operating cycle.

> **OPEN CHALLENGE** **Staying in Production Through Transients**
>
> Recovery cycles on vehicle engines have reached prototypes and a few demonstration vehicles, yet it is still unclear how viable and efficient they are in real use. They often struggle to perform over a broad range of transient conditions, and a cycle that must shut down too often, or takes too long to reach power production, loses the gains it was built for [3]. A review counts automotive heat recovery, together with the conversion of concentrated sunlight, among the applications whose control aspects may be critical [9]. Progress would be a cycle that stays in production through a whole driving cycle, its superheat never below what the expander tolerates.

5.8 Conclusion

This chapter used a waste-heat ORC simulator with one pressure per side, a two-zone receiver and audited R1233zd(E) tables to weigh each part of a superheat control structure. On a 30 % loss of source flow the superheat loop alone, independent or in a cascade, lets the expander inlet fall to 1.83 or 1.09 K and loses the expander to protection, while a feedforward from the source meters holds it within 1.08 K of setpoint and cuts the IAE from 276.7 to 29.0 K s. On a hot day the condenser's fan limit, not the loop, sets the condensing pressure, and raising the setpoint to 2.6 bar nets 12.24 against 8.69 kWh. A slower, noisier thermowell erodes the advantage of fast tuning and nearly triples its pump travel. Server water at 55 °C cannot pay for the pumps and fans at any of 192 settings. A predictive controller told of a turndown holds the superheat at 5.80 K where the cascade falls to 2.43 K, or 5.11 K without the forecast, but its design-point model loses to the cascade under fouling.

The figures carry a 7 % cell-count sensitivity and rest on synthetic parameters. A fuller model would add frictional pressure drop, a converged mesh, a fitted expander map and a learner study. The simulator shows that the superheat a dry-inlet expander needs is protected by measuring the heat before it arrives.

KEY POINTS

- The expander needs dry vapor, 10 K above its boiling point, and the thermowell reports it after most of the evaporator's response.
- On a 30 % loss of source flow, the superheat loop alone loses the expander to protection, while a feedforward from the source meters holds the superheat within 1.08 K of setpoint.
- On a hot day the condenser's fan limit, not the loop, sets the condensing pressure, and raising the setpoint to 2.6 bar nets 12.24 against 8.69 kWh.
- A predictive controller told of a turndown holds 5.80 K where the cascade falls to 2.43 K, but its design-point model loses to the cascade under fouling.

Where else this applies. Volume 18, Chapter 4 feeds the steam flow forward to a drum level cascade, Chapter 6 feeds the raw water forward to a coagulant dose, and Chapter 4 biases a temperature loop with the measured process flow.

References

[1] S. Quoilin, M. Van Den Broek, S. Declaye, P. Dewallef, and V. Lemort, "Techno-economic survey of organic Rankine cycle (ORC) systems," *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 168–186, 2013.

[2] J. Zhang, W. Zhang, G. Hou, and F. Fang, "Dynamic modeling and multivariable control of organic Rankine cycles in waste heat utilizing processes," *Computers & Mathematics with Applications*, vol. 64, no. 5, pp. 908–921, 2012.

[3] J. Peralez, M. Nadri, P. Dufour, P. Tona, and A. Sciarretta, "Organic Rankine cycle for vehicles: Control design and experimental results," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 952–965, 2017.

[4] 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.

[5] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[6] A. Desideri, B. Dechesne, J. Wronski, M. Van den Broek, S. Gusev, V. Lemort, and S. Quoilin, "Comparison of moving boundary and finite-volume heat exchanger models in the Modelica language," *Energies*, vol. 9, no. 5, p. 339, 2016.

[7] S. Quoilin, R. Aumann, A. Grill, A. Schuster, V. Lemort, and H. Spliethoff, "Dynamic modeling and optimal control strategy of waste heat recovery organic Rankine cycles," *Applied Energy*, vol. 88, no. 6, pp. 2183–2190, 2011.

[8] L. Y. Bronicki, "History of organic Rankine cycle systems," in *Organic Rankine Cycle (ORC) Power Systems*, E. Macchi and M. Astolfi, Eds. Elsevier, 2017, pp. 25–66.

[9] P. Colonna, E. Casati, C. Trapp, T. Mathijssen, J. Larjola, T. Turunen-Saaresti, and A. Uusitalo, "Organic Rankine cycle power systems: From the concept to current technology, applications, and an outlook to the future," *Journal of Engineering for Gas Turbines and Power*, vol. 137, no. 10, p. 100801, 2015.

[10] T. C. Hung, T. Y. Shai, and S. K. Wang, "A review of organic Rankine cycles (ORCs) for the recovery of low-grade waste heat," *Energy*, vol. 22, no. 7, pp. 661–667, 1997.

[11] B. F. Tchanche, G. Lambrinos, A. Frangoudakis, and G. Papadakis, "Low-grade heat conversion into power using organic Rankine cycles – a review of various applications," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 8, pp. 3963–3979, 2011.

[12] K. Ebrahimi, G. F. Jones, and A. S. Fleischer, "A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 622–638, 2014.

[13] S. Eyerer, C. Wieland, A. Vandersickel, and H. Spliethoff, "Experimental study of an ORC (Organic Rankine Cycle) and analysis of R1233zd-E as a drop-in replacement for R245fa for low temperature heat utilization," *Energy*, vol. 103, pp. 660–671, 2016.

[14] R. Akasaka and E. W. Lemmon, "An international standard formulation for *trans*-1-chloro-3,3,3-trifluoroprop-1-ene [R1233zd(E)] covering temperatures from the triple-point temperature to 450 K and pressures up to 100 MPa," *Journal of Physical and Chemical Reference Data*, vol. 51, no. 2, p. 023101, 2022.

[15] V. Lemort, S. Quoilin, C. Cuevas, and J. Lebrun, "Testing and modeling a scroll expander integrated into an organic Rankine cycle," *Applied Thermal Engineering*, vol. 29, no. 14–15, pp. 3094–3102, 2009.

[16] D. Wei, X. Lu, Z. Lu, and J. Gu, "Dynamic modeling and simulation of an organic Rankine cycle (ORC) system for waste heat recovery," *Applied Thermal Engineering*, vol. 28, no. 10, pp. 1216–1224, 2008.

[17] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, "Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp," *Industrial & Engineering Chemistry Research*, vol. 53, no. 6, pp. 2498–2508, 2014.

[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] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[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] 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.

[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] 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] 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.

5.9 Exercises

5.1 ⌕ Concept check · 2 min

In case 4 the cascade without the feedforward loses the expander to the protection, although its low-superheat override takes the flow setpoint once the superheat falls below 5 K. Why does the override not prevent the intervention?

(a) It reads the same lagging superheat measurement, so it acts after most of the evaporator's response

(b) Its gain of 0.08 kg/s per K is too small to move the flow setpoint far enough

(c) The minimum selector keeps the main controller's request while the superheat is falling

(d) It opens the source bypass, which removes heat from the evaporator instead of feed

5.2 ⌕ Concept check · 1 min

In case 3 the evaporating-pressure setpoint rises from 12.9 to 13.4 bar with PIC-1 in Auto, and the true superheat falls to 6.44 K, although the pressure loop never touches the feed pump. What lowers the superheat first?

(a) PIC-1 raises the pump speed to lift the pressure and floods the evaporator

(b) The saturation temperature rises by about 1.85 K at once, and the vapor follows only slowly

(c) The thermowell lag grows with the pressure, so the transmitter reads low

(d) The expander speeds up, draws more vapor and cools the evaporator outlet

5.3 ⌕ Concept check · 2 min

In case 5 the cooling water warms to 35 °C, and PIC-2 reports LIMITED with the cooling water at 100 %. Which action recovers the most net power?

(a) Raise the condensing-pressure gain so that PIC-2 reaches its setpoint sooner

(b) Lower the condensing-pressure setpoint so that the loop asks for still more cooling

(c) Raise the condensing-pressure setpoint to about 2.6 bar, so that the flow and fan power fall

(d) Switch the superheat loop to the cascade with feedforward, so that the expander makes more work

5.4 Theory and calculation · 7 min

At the design point the expander of Equation (5.5) runs at 2146 rpm with a swept volume $V_d = 1.5\,\mathrm{L}$, friction torque $\tau_f = 20\,\mathrm{N\,m}$ and generator efficiency $\eta_g = 0.94$. It swallows 3.567 kg/s, leaks 0.341 kg/s, and the swept flow does $w = 28.62\,\mathrm{kJ/kg}$ of work.

(a) Compute the vapor density at the expander inlet and the leakage as a share of the total flow through the machine. *Checkpoint:* 66.5 kg/m^3, 8.7 %

(b) Compute the fluid work, the friction loss, the shaft power and the generator output. *Checkpoint:* 102.09, 4.49, 97.60 and 91.74 kW

(c) Compute the shaft torque, and say why the leakage adds to the flow the expander passes but not to its work. *Checkpoint:* 434 N m

5.5 Theory and calculation · 7 min

The feedforward of Equation (5.8) divides the measured source heat by the enthalpy rise the working fluid needs. At the design point the evaporator accepts 959.95 kW and the pump delivers 3.909 kg/s. The pump's displacement flow is 5.0 L/s at full speed and its nominal density 1250 kg/m^3.

(a) Compute the enthalpy rise from the pump outlet to the expander inlet at the design point. *Checkpoint:* 245.6 kJ/kg

(b) In case 4 the source flow falls to 70 % of 5 kg/s. Lined out, TT-2 reads 150 °C and TT-3 102.4 °C, with $c_{p,s} = 4300\,\mathrm{J/(kg\,K)}$. Take the enthalpy rise of part (a) as unchanged and compute the feedforward flow. *Checkpoint:* 716.4 kW, 2.917 kg/s

(c) Convert the flows of parts (a) and (b) into pump speeds through the nominal capacity, as the independent architecture does, and convert the suggested superheat gain of 0.08 kg/s per K into percent of pump speed per kelvin. *Checkpoint:* 62.5 %, 46.7 %, 1.28 % per K

5.6 Theory and calculation · 7 min

Use the design values of Section 5.3: generator 91.74 kW, pump motor 6.94 kW, cooling 5.26 kW at 55.95 % flow, other auxiliaries 2.00 kW, accepted heat 959.95 kW, source supply 150 °C and cooling water 25 °C.

(a) Compute the net power, the net efficiency and the Carnot bound between the two supply temperatures. *Checkpoint:* 77.54 kW, 8.08 %, 29.54 %

(b) The pump raises 3.909 kg/s of liquid at about 1247 kg/m^3 through 10.705 bar with a hydraulic efficiency of 0.55 and a motor efficiency of 0.88. Compute the enthalpy rise across the pump, the power that enters the fluid and the motor power. *Checkpoint:* 1.561 kJ/kg, 6.10 kW, 6.93 kW

(c) The cooling pumps and fans draw $P_{c,max} f_c^3$ with $P_{c,max} = 30\,\mathrm{kW}$. Compute the cooling power at the design flow, at full flow and at 61 % flow, the case 5 flow at a 2.6 bar setpoint. *Checkpoint:* 5.25, 30.0 and 6.81 kW

5.7 App challenge · 8 min

Load scenario 4, Lose source flow: feedback or feedforward, at the default seed. Run each case to 480 s, export the CSV and read the column `truth_sh_K` from 120 to 480 s, the column `expander_isolated`, and the Trips · interventions and Net energy tiles.

(a) Run the scenario as shipped, with Source-heat feedforward (cascade) Off. Give the lowest true superheat and when it occurs, and how long the expander is isolated. *Checkpoint:* 1.10 K at 141 s, isolated from 137.3 to 220.75 s, about 83 s

(b) Set Source-heat feedforward (cascade) to On, press Reset and run to 480 s again. Give the true superheat range and the two tiles. *Checkpoint:* 9.41 to 11.07 K, 0 · 0

(c) Compare the Net energy tiles of the two runs and explain where the difference comes from. *Checkpoint:* 6.77 against 8.40 kWh

Accept the superheat values within 0.05 K and the energies within 0.02 kWh.

5.8 App challenge · 7 min

Load scenario 5, A hot day at the condenser, at the default seed. The cooling water warms from 25 to 35 °C between 60 and 180 s.

(a) Run to 900 s as shipped. Read the status of PIC-2, the cooling-water flow, the true condensing pressure, the cooling power and the net power from the last row of the CSV, and the Net energy tile. *Checkpoint:* LIMITED, 100 %, 2.38 bar, 30.0 kW, 47.3 kW, 12.94 kWh

(b) Reload the scenario, run to 240 s, raise the Condensing pressure setpoint to 2.60 bar and run to 900 s. Read the same quantities. *Checkpoint:* 61 %, 2.60 bar, 6.8 kW, 67.6 kW, 16.49 kWh

(c) Explain why no tuning of PIC-2 recovers this power, and why the generator makes less in run (b) although the net power is higher.

Accept net power within 0.4 kW, which covers the noise on the last row, and energies within 0.02 kWh.

5.9 App challenge · 8 min

Load scenario 9, MPC or cascade through a scheduled turndown, at the default seed. It ships with Architecture MPC, MPC preview (load schedule) on Schedule and a superheat setpoint of 7 K. Run each case to 600 s and read the column `truth_sh_K` of the exported CSV.

(a) Run as shipped. Give the lowest true superheat and its time. *Checkpoint:* 5.80 K at 134 s

(b) Press Reset, set Architecture to Cascade before pressing Run, and run again. Give the lowest true superheat, its time and the time spent below the 5 K floor. *Checkpoint:* 2.45 K at 149 s, about 21 s

(c) Press Reset, set Architecture back to MPC and MPC preview (load schedule) to Off, and run again. Give the lowest true superheat, then split the gap between parts (a) and (b) into the part the forecast is worth and the part that moving two handles together is worth. *Checkpoint:* 5.11 K, 0.69 K and 2.66 K

Accept the superheat values within 0.05 K and the time below the floor within 2 s.

5.10 Application area · 4 min

In the shipped challenge the protection still reads 3 K of superheat when the true value is about 1.4 K.

(a) TT-1 is a first-order lag with $\tau_{TT} = 3$ s. For a superheat falling at a steady rate, a first-order lag reads low by the rate times its time constant. Estimate the rate at which the true superheat was falling. *Checkpoint:* about 0.53 K/s

(b) The chapter calls every threshold of its protection a teaching value. Where would the thresholds of an operating unit come from, and name two changes to the protection that the gap of part (a) suggests.

5.11 Application area · 3 min

A data-center operator asks for an ORC on 55 °C server return water with cooling water at 15 to 22 °C. Use case 8 to answer.

(a) Compute the Carnot bound at both cooling-water temperatures, and the share of the accepted 128.5 kW that the generator converts at 3.29 kW. *Checkpoint:* 12.19 and 10.06 %, 2.56 %

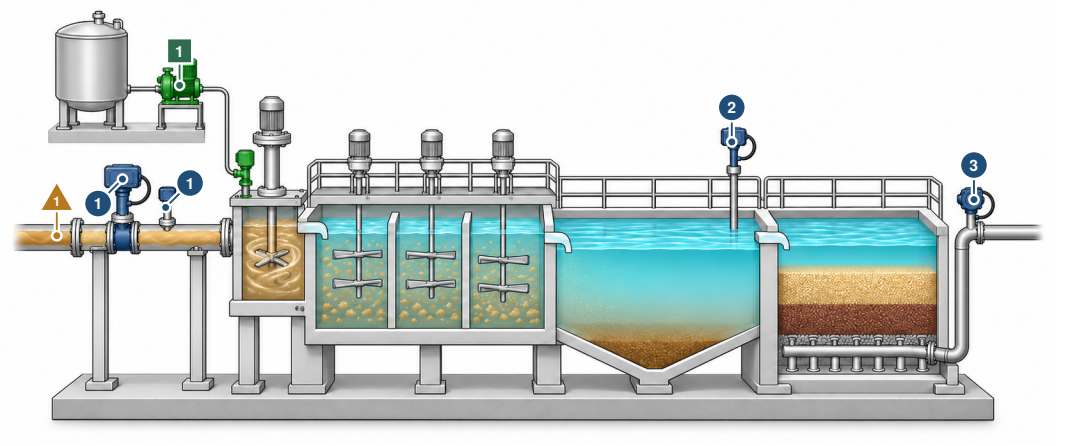
(b) The parasitics take 2.28 kW. What does the chapter's sweep of 192 settings say about operating the cycle, and what is the advice?

5.12 Review the agent · 5 min

An AI assistant reviews scenario 9 and writes: "The MPC holds the superheat at 5.80 K where the cascade falls to 2.43 K. Predictive control is therefore worth 3.4 K of superheat margin over the cascade on any loss of source flow." Find the flaw and say how you would check it.

CHAPTER 6

Municipal Water Treatment Simulator for Feedforward Control of Coagulant Dosing



The controller moves 1 alum metering pump

The controller reads 1 raw-water turbidity and temperature 2 settled-water turbidity 3 filtered-water turbidity

What disturbs it 1 storm turbidity and organic carbon in the raw water

The raw-water meters read a storm before the dosing point, while the settled-water meter waits behind the whole basin. Illustration, not to scale.

IN PLAIN TERMS Dose the Water Before the Storm Arrives

The main idea. A water plant adds a chemical that makes fine mud clump together so it can settle out. The right amount depends on how muddy the river water is as it comes in.

The challenge. The meter that judges the dose sits about three hours downstream, but a storm can muddy the river in half an hour. A controller that waits for that meter acts far too late, and the extra chemical it adds then makes the water worse.

Why it matters. Measuring the river water as it enters lets the plant set the dose in time, and the slow meter only fine-tunes it.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the history of the filtered-water rules (Section 6.1), the fact on the by-product limit (Section 6.3), the try-it segments of scenarios 4 and 8 (Section 6.5), the fact on alum use (Section 6.8) 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 6.5), Technology in Practice (Section 6.6), Limitations (Section 6.7) and the open challenge

The control question The settled-water turbidimeter that judges the coagulant dose reads three hours after the dosing pump, and a storm raises the raw turbidity fivefold in half an hour. Where should the dose be computed, and what is left for the slow measurement, the pH loop and the filter sequence to do?

Learning objectives

- Compute the coagulant demand, the dose ratio and the destabilized fraction from the model for a given raw water and dose.
- Explain why a feedback loop on the settled water crosses the dose optimum, and read the sign change of the process gain off the dose curve.
- Compare feedback with feedforward plus a multiplying trim on a storm, and say what the trim corrects that the feedforward cannot see.
- Judge the worth of a fast pH loop and a filter-to-waste timer from measured runs, and the cost of an integrating trim left on a fault.
- Locate the window between the disinfection credit and the by-product limit, and name the variable that moves it.

Prerequisites Sections 1.4.2 (where a disturbance enters), 1.4.3 (static feedforward), 1.4.4 (ratio with a trim), 1.6.2 (velocity form) and 1.6.4 (scenarios and exported files).

The app https://apmonitor.com/apps/water-treatment/

See also Volume 5, Chapter 4 and Volume 5, Chapter 3 for control ahead of a slow analyzer, and Volume 5, Chapter 7 for diverting material judged on a delayed measurement. Volume 8, Chapter 13 for a diversion valve under supervisory logic, and Volume 8, Chapter 12 for discrete units that fall into step on one shared header.

6.1 Introduction

The coagulant dose is the most important decision a conventional surface water plant makes, and the instrument that reports whether the decision was right sits at the far end of a flocculation basin and two sedimentation stages.

The simulator measures the separation. Stepping the alum from 31.3 to 38 mg/L in manual with everything else fixed moves the settled water from 1.200 to 2.435 NTU. Two percent of that change appears after 32.25 min, sixty-three percent after 193.6 min and ninety-five percent after 402.7 min. That is the plant the coagulant controller is asked to close a loop around.

The disturbance is faster by two orders of magnitude. A storm on the intake raises the raw turbidity from 12 to 60 NTU in half an hour, and a plant at 1200 m^3/h produces 600 m^3 of water in that half hour. By the time the settled turbidimeter has reported the first two percent of the problem, the water that caused it is already in the clearwell.

No tuning changes that arithmetic, and Section 6.5.2 shows that tightening the loop makes matters actively worse rather than merely insufficient. What changes it is that the disturbance is *measured*, entry point 2 of Section 1.4.2. A raw-water turbidimeter reads in seconds, costs a few thousand dollars and is installed on essentially every plant of this size. The coagulant demand is a known function of what it reports. Computing the dose from the disturbance and demoting the settled turbidimeter to a slow ratio trim takes the peak settled turbidity through the identical storm from 20.918 NTU to 3.600 NTU and the time above the filtered water specification from 141.9 min to none. Industrial practice has converged on the same structure, with the dose computed from online raw water measurements and corrected slowly [1].

Two further properties make this plant richer than a pure dead-time example. The coagulant loop's process gain reverses sign at the dose optimum, and a quarter of the plant is a discrete sequence rather than a feedback loop.

The physical fundamentals behind the model are the ordinary ones. Each unit is a material balance on a well-mixed volume, so its time constant is the hydraulic residence time and the train is a chain of first-order lags in series, the structure a first course develops as higher-order and apparent-dead-time behavior [2, 3]. The driving force for clarification is the gravitational settling velocity of the floc

against the surface overflow rate, and the driving force for disinfection is the product of free chlorine residual and contact time. The separation of time scales is what organizes the control problem, because the pH loop closes in minutes, the chlorine loop in tens of minutes and the coagulant loop in hours. The experiments built on that structure cover dead time and high-order lag, feedforward from a measured disturbance, ratio control, sign-changing process gain, cascade and trim structure, integral windup against a fault outside the loop's authority, sampled and filtered measurements, and the interaction between discrete sequencing and continuous loops [2, 4]. The behavior matters industrially because turbidity is the regulated surrogate for pathogen removal, so a coagulation excursion is a reportable compliance event rather than a quality inconvenience, and because coagulant is the largest chemical line on a conventional plant's operating budget.

> ⧖ HISTORY **From 0.5 to 0.3 NTU, 1989 to 2002**
>
> A federal rule of 1989 required surface-water systems in the United States that could not qualify to avoid filtration to filter their water as well as disinfect it, by 29 June 1993 at the earliest. For conventional filtration it held the filtered water at or below 0.5 NTU in 95 % of each month's measurements and never above 5 NTU [5, 6]. In 1993 a single waterborne outbreak of cryptosporidiosis in one city caused more than 400,000 cases [6]. A rule published in December 1998 tightened the standard for systems serving 10,000 people or more, from January 2002, to 0.3 NTU in 95 % of the measurements and never above 1 NTU. It added continuous turbidity monitoring of every individual filter [5, 6]. Those two numbers are the filtered-water specification this chapter grades against.

This chapter works with a simulated municipal water treatment plant whose coagulant measurement is three hours downstream of its actuator, so that the case for feedforward is measured rather than asserted, and it separates the three kinds of control problem the same plant carries. Section 6.2 presents the coagulation chemistry that gives the dose curve its interior optimum. Section 6.3 gives the model of the train, from the coagulant demand and the achievable target to the filters and the chlorine contact tanks. Section 6.4 describes the coagulant feedforward with its ratio trim and the pH, chlorine and level loops around it. Section 6.5 lists the scenarios and measures the storm, the unmeasured organic carbon, the soft water, the backwash sequence, the disinfection window and the graded shift. Section 6.6 describes the instruments, pumps and control platforms of a real plant, and Sections 6.7 to 6.9 state the limitations, discuss the three problems and summarize the results.

6.2 Background

Conventional surface water treatment has had the same unit sequence for more than a century, with coagulation and rapid mix, flocculation, sedimentation, granular media filtration and disinfection in series, and the design and operating basis for each unit is standard material [7]. The dose has traditionally been set from jar tests run once a shift or once a day, which is an open-loop procedure calibrated on a sample rather than on the water in the basin. Online raw water instrumentation changed what is possible, and dose control computed from continuous raw water measurements and corrected by a slow feedback trim is now the recommended structure [1].

Coagulation is not monotone in dose. Below the demand the colloid is still charged and does not aggregate. Above it the coagulant reverses the charge and restabilizes the particles, and the mechanism that dominates depends on both dose and pH in a way that has been mapped for aluminum since the nineteen eighties [8, 9]. The simulator represents this as a log-normal in the dose ratio multiplied by a Gaussian in pH,

$$\eta = \eta_{\max} \exp\left[-\frac{\ln^2(C_a/C_a^*)}{2\sigma_a^2}\right] \exp\left[-\frac{(\mathrm{pH}-\mathrm{pH}^*)^2}{2\sigma_p^2}\right] \tag{6.1}$$

with $\sigma_a = 0.45$ on the natural log of the ratio and $\sigma_p = 0.60$ pH units. Table 6.1 is the resulting curve at the design raw water, and Figure 6.1 plots its settled-water column against the dose.

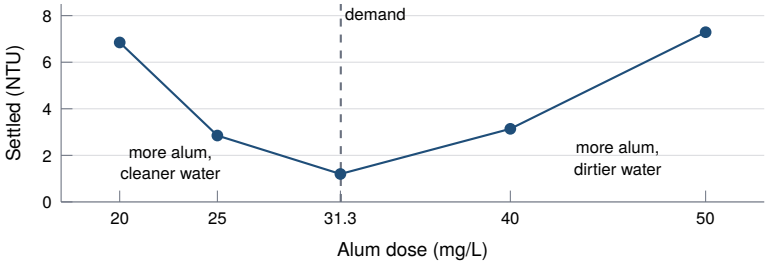


Figure 6.1 The settled-water column of Table 6.1 against the alum dose at the design raw water, with the demand marked by the dashed line. Notice that the slope changes sign at the demand: left of it more alum clears the water, right of it more alum makes it worse, and a settled-water reading alone does not say which side the plant is on.

Table 6.1 Settled and filtered water against alum dose at the design raw water of 12 NTU, 3 mg/L organic carbon and 18 °C. The demand is 31.32 mg/L.

Alum (mg/L)	η	pH	Settled (NTU)	Filtered (NTU)
20	0.598	6.947	6.848	0.5712
25	0.868	6.927	2.851	0.0773
31.3	**0.985**	6.900	**1.200**	**0.0200**
40	0.848	6.862	3.139	0.0925
50	0.568	6.814	7.289	0.6875

A feedback controller responding to high settled turbidity by adding coagulant is correct on the left of Table 6.1 and destructive on the right, and nothing in its measurement distinguishes the two. The same controller is an integrator with a three-hour plant in front of it, so once it crosses the optimum the error and the dose grow together rather than closing. The process gain has changed sign and the loop has become a positive feedback path. Section 6.5.2 measures that behavior and Figure 6.4 shows it in progress at a dose ratio of 3.85 and a destabilization of 1.1 %.

The pH term in (6.1) carries a second nonlinearity. Alum consumes alkalinity, 0.5 mg/L of it as calcium carbonate per mg/L of alum, so the coagulant moves its own pH. On a well-buffered water the effect is small. On a soft water it dominates, and Section 6.5.4 measures a plant that is dosed correctly and produces 9.7 NTU settled water because of it.

6.3 Process Model

Figure 6.2 shows the train. Every unit is a well-mixed volume whose time constant is its hydraulic residence time, and the twenty-seven states span three orders of magnitude in time constant. The dosing pumps and the rapid mix respond in 0.5 min, each flocculator holds 12 min, each clarifier stage holds 80 min, each chlorine contact stage holds 15 min, and a filter run at design lasts about 900 min.

6.3.1 Working Units and the Design Point

The clock is in minutes, flows in m^3/h, turbidity in NTU, doses in mg/L and filter headloss in meters. The design raw water is 12 NTU at 18 °C with 3 mg/L of organic carbon and 80 mg/L of alkalinity. At 1200 m^3/h the plant requires 31.32 mg/L of alum at pH 6.900, which destabilizes 98.5 % of the particles, settles to 1.200 NTU at a surface overflow rate of 1.5 m/h and filters to 0.0200 NTU at 7.5 m/h. A chlorine feed of 2.173 mg/L holds 1.000 mg/L of free residual and earns 15.0 mg · min/L of disinfection credit.

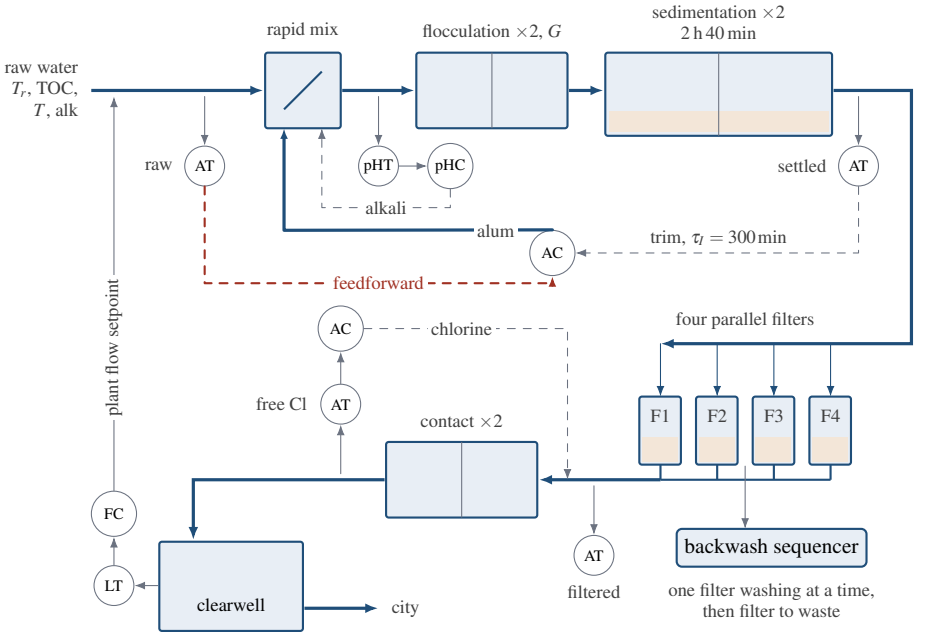


Figure 6.2 The simulated plant. The clarification train runs left to right along the top and the filters, contact tanks and clearwell run right to left below it. The coagulant dose is computed from the raw turbidimeter through the thick dashed path labeled feedforward and only trimmed by the settled turbidimeter, through the thin dashed path labeled trim, which is three hours downstream of the dosing point. The pH controller acts on the same rapid mix through the alkali pump and settles in about two minutes. The chlorine residual controller and the clearwell level controller close on the two ends of the plant. The filters are not a loop at all.

6.3.2 Coagulant Demand and the Achievable Target

The demand rises sub-linearly with turbidity, linearly with organic carbon and sharply in the cold,

$$C_a^* = \left(d_0 + d_1\sqrt{T_r} + d_{\mathrm{TOC}}\,\mathrm{TOC}\right)\left(1 + k_T\frac{(T_c - T)_+}{T_c}\right) \tag{6.2}$$

where C_a^* is the demand in mg/L of alum, T_r is the raw turbidity in NTU, TOC is the raw organic carbon in mg/L, T is the raw water temperature in °C and the subscripted plus sign denotes the positive part. The coefficients are $d_0 = 8.0\,\mathrm{mg/L}$, $d_1 = 5.0\,\mathrm{mg/L}$ per $\mathrm{NTU}^{0.5}$, $d_{\mathrm{TOC}} = 2.0$, $k_T = 0.35$ and $T_c = 15\,^{\circ}\mathrm{C}$, and the cold-water term is an empirical ratio of Celsius temperatures rather than a rate expression. Equation (6.2) gives 31.32 mg/L at design, 36.44 mg/L for the same water at 8 °C and 60.01 mg/L for 55 NTU runoff at 12 °C carrying 5.5 mg/L of organic carbon.

A second scaling governs what the clarifier can reach. No dose makes 60 NTU river water settle to 1.2 NTU, because a clarifier removes a fraction rather than an amount. The achievable target therefore moves with the load,

$$T_s^{\mathrm{SP}} = T_{s,0}^{\mathrm{SP}}\left(\frac{T_r}{T_{\mathrm{ref}}}\right)^{m}\left(\frac{\mu(T)}{\mu(18\,^{\circ}\mathrm{C})}\right)^{1.213}\left(\frac{Q}{Q_d}\right)^{1.85} \tag{6.3}$$

where $T_{s,0}^{\mathrm{SP}}$ is the operator-set target at the reference raw water $T_{\mathrm{ref}} = 12\,\mathrm{NTU}$, $\mu(T)$ is the water viscosity relative to 20 °C, $Q_d = 1200\,\mathrm{m^3/h}$ is the design flow and $m = 0.72$. The operator's 1.2 NTU becomes 2.32 NTU at 30 NTU raw, 3.59 NTU at 55 NTU and 4.27 NTU at 55 NTU and 12 °C. A trim loop held to a fixed setpoint through a storm is given a target the plant cannot reach, and it

integrates across the coagulation optimum in trying to reach it. The simulator shapes (6.3) through two lags of 87.5 min each before presenting it to the trim, so that the setpoint moves at approximately the speed of the plant.

6.3.3 Flocculation and Sedimentation

Each flocculator carries a floc size index whose target has an interior maximum in the mixing intensity,

$$S^{*} = \eta \, \frac{G}{G_{\mathrm{opt}}} \, \mathrm{e}^{\,1 - G/G_{\mathrm{opt}}} \min\!\left(1.35, \left(\frac{T_r}{T_{\mathrm{ref}}}\right)^{0.12}\right) \tag{6.4}$$

where G is the velocity gradient and $G_{\mathrm{opt}} = 45/\mathrm{s}$ is the value at which the floc grows fastest, because too little energy does not bring the particles together and too much tears the floc apart faster than it forms. Each stage relaxes toward S^{*} at $k_g = 0.12/\mathrm{min}$. Settling velocity follows Stokes through the square of the size index and inversely with viscosity [10], $v_s = v_{s0} S^{2} / \mu_r(T)$ with $v_{s0} = 11.24\,\mathrm{m/h}$, which gives 7.21 m/h at design. The basin removes a fraction per stage,

$$\mathrm{pass} = \left(1 + \frac{v_s A_s / n}{Q}\right)^{-n}, \qquad n = 2. \tag{6.5}$$

where $A_s = 800\,\mathrm{m}^2$ is the plan area of the basin. Turbidity is transported through the basin unchanged and (6.5) is applied algebraically at the outlet rather than as a sink inside the tanks. That choice sets the lag the coagulant controller sees. A well-mixed tank containing a removal sink responds at $V/(Q + v_s A_s)$ rather than at V/Q, which would give the train a much shorter lag than the three-hour transport response, and the lag is the subject of this chapter.

🕮 **Pause.** Try Exercise 6.3 before reading on.

6.3.4 Filtration and Backwash Sequencing

A filter is a log-removal exponent with three modifiers,

$$\mathrm{pass} = \exp\left[-\frac{k_f S}{v/v_0} \underbrace{\left(1 - 0.95\,\mathrm{e}^{-t_r/2}\right)}_{\text{ripening}} \underbrace{\left(1 - 0.35\,\hat{H}^{4}\right)}_{\text{breakthrough}}\right] \tag{6.6}$$

where k_f is the removal coefficient, S the floc size index, v the filtration rate against its design value v_0, t_r the minutes since the filter returned to service and $\hat{H}$ the fraction of available head consumed. The ripening term matters because the maintenance activity that relieves breakthrough creates it. A freshly washed filter passes almost everything for the first minute, which has been documented on operating plants since 1980 [11], and the resulting spike appears directly in the regulated measurement.

Headloss builds with the product of rate, applied turbidity and time, which gives runs of about 900 min at design and much shorter runs in a storm. The sequencer is discrete. A filter is queued when its headloss reaches 2.20 m or its run reaches 1200 min. Only one filter may be washing at a time: it draws wash water for 10 min and remains isolated for 14 min. An optional 7 min filter-to-waste period follows. A second filter may start washing while the first sends filtrate to waste. Only filters in service contribute to product quality, while the filtration rate includes filters sending water to waste.

6.3.5 Disinfection and By-Product Formation

Two chlorine contact stages carry a first-order decay driven by organic carbon and particles. The credit and the by-products are computed from the same residual and the same contact time,

$$\mathrm{CT} = 60\,C\,\frac{\mathrm{BF}\cdot V_c}{Q}, \qquad \mathrm{THM} = k_{\mathrm{THM}}\,\mathrm{TOC}\,C\,\frac{V_c}{Q} \tag{6.7}$$

where C is the free chlorine residual leaving the second contact stage, V_c is the total contact volume of 600 m^3, Q is the plant flow in m^3/h, and the factor 60 converts hours to minutes for CT. BF is a baffle factor of 0.50 relating the effective contact time T_{10} to the nominal residence time [12]. Both quantities scale with residual and with contact time, so raising the residual to earn the credit raises the by-products in exact proportion, and the only variable that separates them is organic carbon. Enhanced coagulation can remove it in a real plant [13], but the simulated TOC is an unremoved raw-water disturbance.

> ✦ DID YOU KNOW? **The By-Product Limit Is a Federal Number**
>
> The 80 µg/L by-product limit of this chapter is the federal maximum contaminant level for total trihalomethanes, 0.080 mg/L. The same rule names enhanced coagulation, enhanced softening or granular activated carbon, with chlorine as the disinfectant, as the best available technology for meeting it [5].

6.3.6 Initialization and Assumptions

Every stage is a first-order tank, so the steady state marches forward analytically from the raw water rather than iterating, and `init` returns an exact equilibrium. The manual helpers set alum to the true jar-test demand together with its pH correction, calculate chlorine for the selected residual, or request a backwash of the filter with the greatest headloss.

Turbidity is a single lumped particle population split into a settleable and a non-settleable fraction. The coagulation surface (6.1) is an empirical shape rather than a mechanism. Backwash water is scaled by 0.12 because the filter run clock is compressed by about ten times, so that the plant water balance remains correct at the compressed rate. Integration is fixed-step fourth-order Runge-Kutta at 0.05 min, and the controllers execute every 0.2 min, independent of the integration step.

6.4 Controller Design

6.4.1 Velocity Form PID Algorithm

Every continuous loop uses the framework's velocity-form PID of Section 1.6.2, Equations (1.9) to (1.12), with derivative on the measurement [14] and the ISA action convention. The scan is $\Delta t_c = 0.2\,\mathrm{min}$. Back-calculation acts only on the integral, with the tracking time equal to τ_I, and the integral update is disabled when the integral gain is zero. All default settings have $\tau_D = 0$, and the level loop also exposes derivative time and filtering. The nominal settings below follow SIMC rules [15] applied to measured gains and lags.

6.4.2 Feedforward With a Ratio Trim

The coagulant path has two forms selectable at run time. Pure feedback puts the settled turbidimeter directly on the dosing pump with $K_c = 4\,\mathrm{mg/L/NTU}$ and $\tau_I = 120\,\mathrm{min}$. The recommended structure is the static feedforward of Section 1.4.3 with the multiplying trim of Equation (1.6). It computes

$$C_a = r\,\hat{d}\left(T_r, \widehat{\mathrm{TOC}}, T\right), \quad r = \mathrm{PID}\left(T_s^{\mathrm{SP}}, T_s\right) \in [0.80,\ 1.35] \tag{6.8}$$

where $\hat{d}$ is (6.2) evaluated on the information available to the control system. Two inputs represent online raw-water information, turbidity and temperature. In the implementation the dose model reads the disturbance values directly, and the displayed raw turbidimeter lag and noise affect only the shaped target. The other two inputs are not online, because the organic carbon is a laboratory number reported once a shift and the demand coefficients come from a jar test of unknown age. The trim r is dimensionless, is direct acting with $K_c = 0.06/\mathrm{NTU}$ and $\tau_I = 300\,\mathrm{min}$, and exists to correct those two stale inputs rather than to reject the storm. Trimming a ratio rather than adding an offset matters,

because a correction identified at 12 NTU raw water is proportionally correct at 60 NTU while an additive correction is not. The trim on the air-to-fuel ratio of Chapter 4 plays the same part for an unmeasured heating value.

🕮 **Pause.** Try Exercise 6.2 before reading on.

6.4.3 pH Control of the Rapid Mix

The pH controller moves an alkali pump against the coagulation pH with $K_c = 30\,\text{mg/L}$ per pH unit and $\tau_I = 1.5\,\text{min}$. Its actuator spans −15 mg/L to 45 mg/L, and the loop is reverse acting because adding alkali raises the pH. The signed equivalent dose B adds alkalinity when positive and consumes it when negative, with $A = A_r - 0.5C_a + 1.25B$ and $\text{pH} = 4.610 + 0.55\ln(\max(2, A))$. The pump retains the sign through its 0.5 min lag, and reagent consumption uses $|B|$ for both acid and alkali. The loop is faster than the coagulant loop by two orders of magnitude and it is what gives the coagulant loop a stationary target, because (6.1) multiplies the dose term by the pH term rather than adding them. A dose exactly equal to the demand destabilizes 98.5 % of the particles at pH 6.90 and 4.3 % at pH 5.40. Here pH enters through the logarithm of the alkalinity left after the alum, a much milder curve than the titration of a strong acid, where Volume 2, Chapter 6 needs a gain schedule.

6.4.4 Chlorine, Clearwell Level and the Backwash Sequencer

Free chlorine goes to the chlorinator with $K_c = 3$, $\tau_I = 20\,\text{min}$ and 2 min of sample line delay ahead of a 1 min analyzer lag. The clearwell level sets the plant flow with $K_c = 15\,\text{m}^3/\text{h}$ per percent of level and $\tau_I = 180\,\text{min}$, detuned deliberately so that the clearwell absorbs the city demand rather than passing it to the coagulation basins. The backwash sequencer is a state machine with a single interlock limiting simultaneous washes. With one of four filters washing the remaining three carry the whole flow at 10 m/h instead of 7.5 m/h, which degrades (6.6) for all of them, so a plant whose filters foul together forms a queue that never clears. Discrete units that fall into step on one shared resource are also the problem of the thermostats of Volume 8, Chapter 12.

6.5 Results

6.5.1 Scenarios of the Application

The simulator carries eleven scenarios, and each is credited by a completion rule that checks the specified actions and measured responses rather than the selection of a case (Section 1.6.4). Table 6.2 lists them. All eleven scripted walkthroughs of these rules pass. Runs use seed 20260928, zero default instrument noise and the shipped model parameters and disturbances, including each scenario's scheduled changes.

Two elements are stylized on purpose. The filter run clock is compressed by about ten times, which brings a full wash cycle inside a single scenario so that the ripening spike and the wash queue can be studied together. The storm arrives as two steps rather than as a ramp, because a step is the disturbance whose response can be read directly off a trend.

6.5.2 Feedback Against a Storm

Scenario 4 raises the raw water from 12 to 35 to 60 NTU between 30 and 60 min with the coagulant loop in pure feedback and everything else in service. Scenario 5 runs the same storm with the coagulant on feedforward plus trim.

Table 6.3 is the central measurement of this chapter, and the mechanism is better followed in time than read as a single ratio. Ten minutes after the raw water reaches 60 NTU the feedforward plant has already moved the alum from 31 to 53 mg/L and its settled water is 1.28 NTU. The feedback plant is at 32 mg/L, because its measurement has barely moved, and its settled water is 1.57 NTU. At 140 min the feedback plant is at 55 mg/L and 6.86 NTU, at 200 min it is at 76 mg/L and 9.64 NTU, and at 300 min it is at its 140 mg/L stop and 20.92 NTU. Figure 6.3 draws the table as bars.

Table 6.2 Scenarios and Completion Requirements

#	Scenario	Kind	Completion requirements
1	Train lag	Experiment	Start the run, then set manual alum to 35.5–37 mg/L after at least 0.2 min and observe a settled-water change of at least 0.45 NTU.
2	Manual storm	Demonstration	Observe a fixed-dose violation after 180 min. Reset, raise the dose before 110 min and compare at 180 min.
3	Dose curve	Experiment	Hold 22, 47 and 63 mg/L manual doses (within 1 mg/L) for 240 min each, with an overdose alarm in the last case.
4	Feedback	Demonstration	Run feedback for 420 min. Reset, change gain or integral time before the storm and compare a 330 min run by 360 min.
5	Feedforward	Experiment	Hold a near-demand dose ratio for 60 min after the second storm step, then run to 420 min.
6	Unmeasured	Experiment	Observe the demand jump and settled offset, then the enabled trim's slow correction through 480 min.
7	Soft water	Experiment	Observe pH loss and restoration. Reset with pH control enabled and compare a 230 min run.
8	Sequencing	Experiment	Compare 355 min runs with Auto and Auto + FTW, enabling FTW within the first 20 min after reset.
9	Disinfection	Demonstration	Observe low CT after the flow increase, the bloom dip, and the CT/THM tradeoff after the residual target rises.
10	Commission	Experiment	Step pH, chlorine and level targets and wait for their response tolerances. Step the turbidity target and observe 180 min.
11	Challenge	Challenge	At 600 min, maximum logged settled error at most 1 NTU, off-specification time at most 60 min, zero turbidity-violation and CT-shortfall times, and maximum logged THM strictly below 80 micrograms/L.

Table 6.3 The same storm under two coagulant structures, measured over 300 min.

	Feedback only	Feedforward plus trim
Peak settled (NTU)	20.918	3.600
Peak filtered (NTU)	2.821	0.126
Above 0.30 NTU (min)	141.9	0
Above 1.00 NTU (min)	47.15	0
Peak alum (mg/L)	140.0	53.5
Alum consumed (kg)	415	306
Settled turbidity IAE (NTU min)	1555.6	28.3

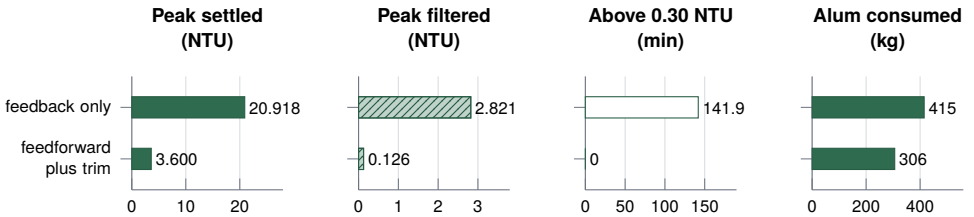


Figure 6.3 Table 6.3 redrawn: the storm of scenarios 4 and 5 under the two coagulant structures, measured over 300 min. Bars give the peak settled and filtered turbidity, the minutes the filtered water spent above 0.30 NTU and the alum consumed. Notice that the structure that doses earlier keeps the filtered water in specification throughout and uses less alum, not more.

The last part of that trajectory is not a slow controller. It is a controller that has crossed the optimum of Table 6.1. When the storm passes at 300 min the raw water falls to 20 NTU and the

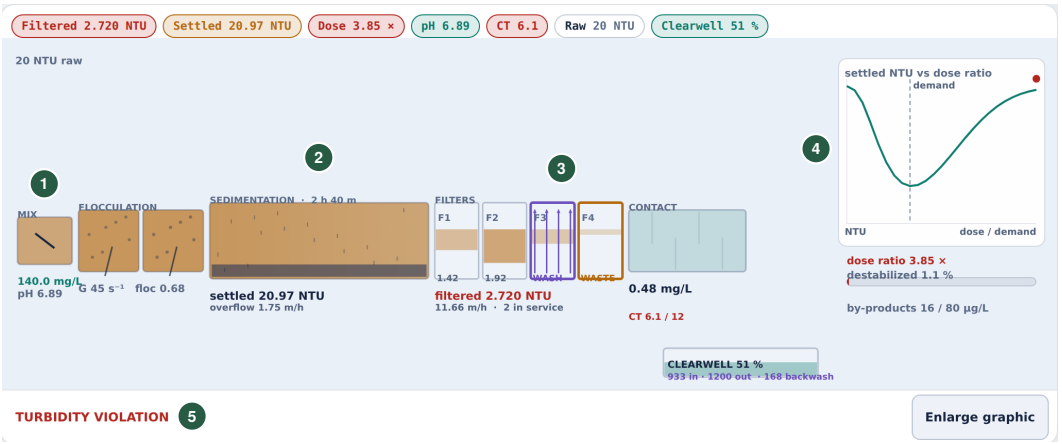


Figure 6.4 The plant view of scenario 4 at 300.2 min, just after the storm passes, cropped from the app window. The crop is a composite: an empty band of the view between the filters and the clearwell is cut out, and the run time is not in the crop. (1) The coagulant controller has integrated to its 140 mg/L stop, read under the rapid mix. (2) The sedimentation basin is drawn with its actual solids loading, and the settled water is 20.97 NTU against a shaped target of 3.61 NTU, which a trend outside the crop shows. (3) Two filters remain in service: one is washing and one is sending filtrate to waste, and the filtered water is 2.720 NTU. (4) The dose ratio is 3.85 and the operating point is drawn on the far side of the coagulation minimum, where destabilization has fallen to 1.1 %. (5) The plant is in turbidity violation.

demand with it to 36.4 mg/L, which leaves the dose ratio at 3.85 and destabilization at 1.1 %. Every additional milligram makes the settled water worse, which the controller registers as a larger error, which produces more coagulant. Figure 6.4 shows that state directly, with the operating point plotted on the far branch of the coagulation curve.

🕮 **Pause.** Try Exercise 6.1 before reading on.

> **TRY IT IN THE APP** **4 · Feedback Against a Storm**
>
> Load the scenario and press Run. The storm takes the raw water to 60 NTU by 60 min. Watch the coagulant trend: the dose climbs slowly, and by 300 min it sits at its 140 mg/L stop with the settled water near 20.92 NTU. Then load 5 · Feedforward against the same storm and run it. The dose moves with the storm, and the settled water peaks at 3.600 NTU.

6.5.3 Unmeasured Disturbances and the Ratio Trim

Scenario 6 runs the same plant with feedforward only and moves a variable the feedforward cannot see. At 30 min, an algal bloom takes the organic carbon from 3 to 6.5 mg/L, which raises the true demand from 31.32 to 38.32 mg/L. The modeled demand does not move, because the dose model is using the laboratory organic carbon number reported that morning. The dose ratio falls to 0.817, destabilization to 89.1 %, and the settled water comes to rest at 2.13 NTU against a 1.21 NTU target.

That offset is what the trim exists to remove, and switching it in at 240 min starts moving the dose upward. The correction is slow. By 600 min the alum has reached 33.4 mg/L and the settled water 2.04 NTU, still short of the 1.23 NTU target. A trim with a 300 min integral time on a 190 min plant does not converge within six hours and is not intended to. Feedforward rejects the fast measured disturbance and the trim corrects the slow model error, and neither substitutes for the other. Volume 5, Chapter 4 trims a residence-time prediction from a late kappa analyzer on the same division of labor.

Table 6.4 Snowmelt with and without the pH controller in service, measured over 300 min. The dose is computed correctly in both columns.

	pH loop out	pH loop in
Minimum pH	5.353	5.918
Time below pH 6.5 (min)	170.35	0.85
Minimum destabilization	2.8 %	25.8 %
Peak settled (NTU)	9.739	1.257
Peak filtered (NTU)	3.831	0.253
Above 0.30 NTU (min)	169.55	0
Above 1.00 NTU (min)	127.35	0

6.5.4 Effect of the pH Loop on Coagulation

Scenario 7 drops the raw alkalinity from 80 to 25 mg/L on snowmelt, with the coagulant feedforward computing the correct dose throughout. The scenario starts with the pH loop gain at zero and puts the loop in service at 200 min. The right-hand column of Table 6.4 is the same run with the pH loop in service from the start.

The two columns of Table 6.4 differ by a factor of 7.7 on the settled water peak, bought by a loop that settles in two minutes and whose reagent costs $0.58/kg. The coagulant controller computes the same dose in both columns and produces unusable water in one of them, because the variable it does not measure has moved. The two minima in the right-hand column are transients of about a minute at the instant the alkalinity steps, before the alkali pump has responded, while the left-hand column remains below pH 6.5 for 170.35 min. The peak filtered turbidity of 0.253 NTU in the right-hand column is a filter ripening spike rather than a coagulation effect, because scenario 7 runs the filters without a filter-to-waste period.

The recovery is far slower than the pH response that starts it. When the pH loop is put in service at 200 min the pH is back at 6.86 within eleven minutes, and the plant does not follow. Seven hundred minutes later the settled water is still 4.09 NTU. Two mechanisms hold it there and neither is the sedimentation lag. The alkali pump is saturated at 45 mg/L because the alum has to be neutralized as well as the raw water, and the coagulant trim wound its ratio to its 1.35 limit, or 42.3 mg/L of alum, during the four hours it spent correcting a fault outside its authority, which places the plant on the over-dosed side of Table 6.1 at 78.7 % destabilization. An integrating trim left running against a disturbance it cannot reach leaves a legacy that outlasts the disturbance. In the same run with the pH loop in service from the beginning, destabilization does not fall below about 90 % once the alkali pump has responded.

6.5.5 Backwash Sequencing and Filter Ripening

Filter runs at design are about 900 min and the four filters start staggered, at 225, 450, 675 and 880 min of run age, so that the wash queue clears. Scenario 8 loads them with a raw water step to 70 NTU at 100 min.

The measurement that matters is the effect of the wash itself. Over 600 min and eight backwashes with immediate return to service, the peak filtered turbidity is 0.804 NTU and the filtered water is above the 0.30 NTU specification for 2.9 min. Adding a 7 min filter-to-waste period after each wash takes the peak to 0.169 NTU and the time off specification to zero, at a cost of 518 m^3 of total water loss against 12 172 m^3 produced. The loss includes 224 m^3 of wash water. No continuous loop is involved in the spike. It is created by the maintenance action, it is predictable to the minute, and it is removed by a timer. Sending material in doubt to waste is also the job of the flow diversion valve of Volume 8, Chapter 13.

Table 6.5 The graded ten-hour shift, measured over 600 min. Each row after the second changes one element of the recommended configuration.

Configuration	Peak filt. (NTU)	Off spec (min)	Violation (min)	Min CT (mg · min/L)	Alum (kg)
As loaded	7.971	493.90	456.00	0.35	379
Suggested	0.279	0	0	13.97	660
feedback only	23.341	454.55	332.60	0.21	1549
no filter-to-waste	0.951	5.55	0	13.97	634

> **TRY IT IN THE APP** **8 · Four Filters and a Queue**
>
> Run the scenario with Filter sequencing on Auto and watch the filtered-water trend: it spikes each time a washed filter returns to service. Press Reset, switch Filter sequencing to Auto + FTW within the first 20 min and run again. Over 600 min the chapter measures a peak of 0.804 NTU without the timer and 0.169 NTU with it.

6.5.6 Disinfection Credit and By-Product Limits

Scenario 9 raises the city demand to 1650 m^3/h, which the clearwell level controller answers by running the plant at its 1700 m^3/h limit. The chlorine residual controller holds its setpoint to within 0.08 mg/L through the flow ramp, and the disinfection credit fails anyway. Contact time is volume divided by flow, so T_{10} falls from 14.88 min to 10.59 min and the credit from 14.60 to 10.59 mg · min/L against a requirement of 12 mg · min/L. The compliance failure is produced by a variable that is not on the chlorine loop at all. An algal bloom at 150 min doubles the organic carbon and the chlorine demand with it, which takes the residual to a transient minimum of 0.745 mg/L and the credit to 8.48 mg · min/L.

Raising the residual target to 1.6 mg/L restores the credit to 17.94 mg · min/L and takes the by-products from 59.0 to 101.1 µg/L against an 80 µg/L limit. Both quantities in (6.7) are proportional to residual and to contact time, so no setting of the chlorine loop widens the window. The only variable that does is the organic carbon. Removing it would require an upstream process change that this model does not implement.

6.5.7 Graded Ten-Hour Shift

The challenge is ten hours against a storm that takes the raw water to 55 NTU, an organic carbon rise, a runoff dilution of the alkalinity to 40 mg/L and a morning demand peak, with the target requiring the settled water to track its shaped setpoint, the filtered water to stay under 0.30 NTU for all but an hour and never to exceed 1.00 NTU, the disinfection credit above 12 mg · min/L and the by-products under 80 µg/L.

The current suggested configuration uses FF + trim and Auto + FTW, with the nominal gains given above except for the stronger trim $K_c = 0.12/\text{NTU}$ and $\tau_I = 150\,\text{min}$. Its residual target is 1.15 mg/L and its level target is 65 %. It passes every clause: maximum logged settled-water error is 0.845 NTU, no CT shortfall is accumulated and the maximum by-products are 70.15 µg/L. Replacing feedforward with feedback fails quality and disinfection, drives the coagulant to its stop and uses 1549 kg of alum against 660 kg. Removing filter-to-waste raises the filtered-water peak to 0.951 NTU and accumulates 5.55 min off specification. It still meets every grading clause on the shipped seed. Filter-to-waste is therefore a recommended quality safeguard rather than a mandatory control setting in the challenge verdict.

The settled turbidity integral against the shaped target is 9549 NTU min as loaded and 341 NTU min recommended.

The quality timers integrate the physical outputs at each three-second step, and the settled-error and by-product clauses read the exported historian. Changing model parameters or disturbance sliders,

or firing additional disturbances, voids challenge credit. Changing controller settings remains part of the task.

6.6 Technology in Practice

Figure 6.5 draws the coagulant feedforward of Section 6.4.2 with its ratio trim, and the pH loop beside it, with their hardware.

What Measures It

Three turbidimeters carry the coagulant structure. The raw-water meter at the intake feeds the dose model together with a raw-water thermometer, the settled-water meter at the basin outlet drives the trim, and the filtered-water meter reads the number the plant is graded on. Guidance for operators names this use of the raw meter: turbidity readings of a surface source serve to adjust the coagulant and disinfectant doses [16]. In the model the dose calculation reads the raw turbidity and temperature directly, and the raw meter's displayed lag and noise act only on the shaped target. The filtered water must be measured at least every four hours, or continuously if the plant validates the online meter on a regular basis [5].

> **TECHNOLOGY Sensor: A Nephelometric Turbidimeter**
>
> A nephelometric turbidimeter shines a beam through the sample and reads, with a photodetector, the light scattered at right angles to it. More scattered light means higher turbidity, and the reading is calibrated against a formazin suspension. The standard method covers 0 to 40 NTU, higher values after dilution, and asks for a sensitivity of 0.02 NTU or better below 1 NTU. Fine air bubbles read high, while coarse sediment that settles out and color in the water read low [17]. Online meters can measure below 0.1 NTU reliably if they are carefully calibrated and maintained [16]. In comparison studies of samples between 0.35 and 0.72 NTU, laboratories using different meters agreed to a relative standard deviation just under 20 % [6].

What Moves It

Alum and alkali enter the rapid mix through two dosing pumps, which the model gives a 0.5 min lag. The alum pump stops at 140 mg/L, and the alkali pump spans −15 mg/L to 45 mg/L, a negative value standing for acid. A chlorinator sets the chlorine feed, and the clearwell level loop sets the plant flow. The model works with the dose itself. On a real plant a dose in milligrams per liter becomes a pump rate only in proportion to the measured water flow, which is the job a chemical metering pump is set up for [18].

> **TECHNOLOGY Final Element: A Chemical Metering Pump**
>
> A metering pump packs the pump mechanism, its motor and the drive electronics in one unit, and water plants use such pumps to inject small flows of coagulant, disinfectant and pH chemicals. A standard 4 to 20 mA signal sets its output from zero to full flow, and knobs on the pump set what full flow is. Many use a plunger driven by a solenoid at variable intervals, so the chemical arrives in pulses rather than as a steady stream. Each chemical's flow must stay in proportion to the flow of water treated, which is why the pump is set by the control system and not by hand [18]. In this chapter the pump follows its dose after a 0.5 min lag and stops at 140 mg/L.

Where the Controller Runs

A distributed control system runs production within one site in industries that include water and wastewater treatment, with feedback or feedforward loops holding the key conditions [19]. Control cycle times range from milliseconds to minutes [19]. In the app every controller executes each 0.2 min, against a coagulant plant that takes 193.6 min to reach sixty-three percent of a step, so the scan never limits the loop. The structure of Equation (6.8) maps onto standard blocks. A calculation

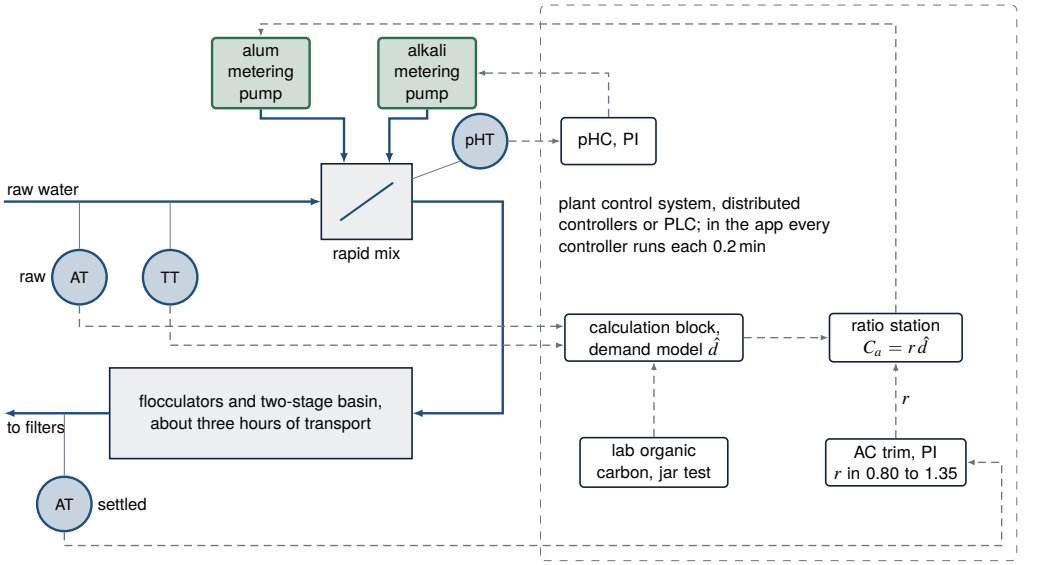


Figure 6.5 The coagulant feedforward with its ratio trim, and the pH loop, on real hardware. The nephelometric turbidimeter AT and the thermometer TT at the intake feed a calculation block that evaluates the demand model of Equation (6.2), with the organic carbon and the jar-test coefficients entered by hand. A ratio station multiplies the demand by the factor r of the trim controller, whose measurement is the settled turbidimeter at the basin outlet, and sends the dose to the alum metering pump. The pH controller moves the alkali metering pump on the same rapid mix. Solid arrows are process streams and dashed gray arrows are signals. Notice that the only signal that waits on the three-hour basin is the slow trim.

block evaluates the demand model, and a signal multiplying relay, or ratio station, multiplies its result by a factor that an operator would otherwise set by hand [18]. Here the trim's PI output sets that factor, between its limits of 0.80 and 1.35.

> **IN PRACTICE** **Every Filter, Every Fifteen Minutes**
>
> Large surface-water plants in the United States must monitor the turbidity of each individual filter continuously and record it every 15 min. A filter above 1.0 NTU in two consecutive readings 15 min apart is reported, and the plant produces a filter profile or names an obvious cause. A filter still above 0.5 NTU in two consecutive readings at the end of its first four hours after a backwash is reported in the same way [5]. Guidance for operators names start-up, filter ripening and the end of a run as the times when spikes occur, asks that they be minimized and ideally eliminated, and lists sending filtered water to waste among the immediate responses to an exceedance [16].

What Is Modernizing It

Three developments bear on this plant. The first is continuous monitoring tied to alarms. Turbidity can be read continuously, with the results linked to automatic alarm systems that allow a rapid response and remedial action when performance deviates [16]. The second is the industrial internet of things: sensors and instruments networked together, with computing at the edge that processes and acts on data where it is collected [19]. The same connectivity raises the need for resilience, safety and security [19]. The third is model-based control and estimation, which the Discussion (Section 6.8) proposes for this train: a predictive controller that evaluates the coagulation surface instead of inferring a gain from the error, and an estimator that turns the trim into a soft sensor for the organic carbon and the jar-test bias.

6.7 Limitations

Turbidity is a single lumped particle population with a settleable and a non-settleable fraction and one floc size index per stage. Real raw water carries a particle size distribution, and the reason a plant that meets its turbidity limit can still fail a pathogen limit is that the organisms of concern sit at the small end of that distribution. Turbidity is a surrogate for pathogen removal rather than a measure of it, and the strength of that surrogacy is itself a research question [20]. The simulator is not a substitute for that literature.

The coagulation surface (6.1) is an empirical shape fitted to give recognizable behavior, not a mechanism. It reproduces the peak, the pH window and the cold-water penalty, and it cannot distinguish charge neutralization from sweep floc, cannot represent polymer aids, and has no memory of the coagulant already in the basin. Data-driven models fitted to individual plants capture considerably more of the real surface [21], and none of them transfer between plants either.

Organic carbon is a single number that affects coagulant demand, chlorine decay and by-product formation, and that is removed by nothing. A real plant's enhanced coagulation removes a large fraction of it, which changes both the disinfection window and the coagulant demand downstream, so the model omits the upstream removal effect discussed in Section 6.5.6.

The filters share one settled water quality and differ only in headloss and run age. Real filters differ in media, in condition and in how well they were washed, and the diagnosis of which filter is the problem is a routine operating task the simulator cannot stage. Backwash water is scaled by 0.12 to keep the plant water balance right against a compressed run clock, so the wash water fraction of production is representative and any single wash is not. Sludge is not modeled, and sludge handling is where a real plant's storm response is often actually limited [22].

Finally, the plant has one train. Real plants have several and can take one out of service, which is the response an operator would reach for in scenario 9 before touching the chlorine setpoint.

> **OPEN CHALLENGE** **Turbidity Stands In for Pathogens**
>
> Low turbidity is a proven indicator that pathogens have been removed, yet no direct proportional relation between removing turbidity and removing pathogens has been shown, and studies that link turbidity to illness in communities have given mixed results. A durable, low-cost meter sensitive below 1 NTU is still missing [16]. Progress would be a measurement closer to the organisms themselves, cheap enough for every filter, so that the trim and the filter sequence could close on the quantity that matters.

6.8 Discussion

The plant carries three control problems with different answers, and separating them is the purpose of this chapter.

The first is a measured fast disturbance in front of a slow plant, and it has a structural answer rather than a tuning answer. The dose belongs on the disturbance measurement and the slow instrument belongs on the model. Table 6.3 gives a factor of 5.8 on the peak settled turbidity and a factor of 55 on the integral, and no tuning of the feedback structure approaches either. The general result is that when a disturbance is measured earlier than its effect, feedback on the effect is the wrong place to spend effort [2].

> **DID YOU KNOW?** **The Earlier Dose Uses Less Alum**
>
> From the simulator: through the same storm, feedforward plus trim peaks at 53.5 mg/L of alum and uses 306 kg, while feedback alone climbs to its 140.0 mg/L stop and uses 415 kg (Table 6.3). Acting early takes less coagulant.

The second is a fast inexpensive loop whose value is indirect. Table 6.4 shows the pH controller holding a variable that carries no specification of its own and deciding whether the plant meets the

one that does. Multiplicative interactions of the form of (6.1) are common in process chemistry and uncommon in introductory block diagrams, and they carry a consistent implication for commissioning order, because the inexpensive fast loop is placed in service first and removed last. The legacy the coagulant trim leaves behind in the same scenario is a second result, which is that an integrating controller acting on a fault outside its authority does not simply fail to correct it, it accumulates an error that outlasts the fault.

The third problem is not a loop. Four filters, an interlock and a timer produce a predictable turbidity spike at a predictable moment, and the answer is a sequence rather than a controller. Continuous control theory does not address it and the regulated measurement records the result. Discrete sequencing interacting with continuous control accounts for a large part of what a plant control system does, and it is underrepresented in control education [4, 23].

The same equations also suit methods the simulator does not implement. A model predictive controller on the twenty-seven state train could hold the filtered water and the disinfection credit as explicit constraints and coordinate the alum, alkali and chlorine doses across the three-hour lag [24]. Because it would evaluate (6.1) rather than infer a gain from the error, it would handle the sign reversal of Table 6.1 directly. An estimator on the raw, settled and filtered turbidimeters could track the organic carbon and the jar-test bias, which would turn the trim of (6.8) into a soft sensor for the two inputs the dose model does not measure.

6.9 Conclusion

This chapter studied a municipal water treatment plant that separates a coagulant decision from its evidence by three hours, and the simulator measures the separation as 32.25 min of apparent dead time and 193.6 min to sixty-three percent against a storm that arrives in thirty. Feedback on that plant does not merely respond slowly. It integrates past the coagulation optimum, reaches its actuator stop at 140 mg/L, destabilizes 1.1 % of the particles and puts the plant in turbidity violation for 47.15 min, while a feedforward computed from a raw turbidimeter and trimmed by the same slow instrument holds the same storm with no time off specification and 306 kg of alum against 415 kg. Three further results follow from the same model, a pH loop worth a factor of 7.7 on the settled water peak and carrying no specification of its own, a filter-to-waste timer that removes a violation no continuous loop reaches, and a disinfection window whose two limits move together and cannot be widened by the chlorine residual setting alone. The model also carries the constraints and the estimation problem that model predictive control and moving horizon estimation would address. The plant measures what feedforward is worth where feedback cannot arrive in time.

KEY POINTS

- The meter that judges the dose reads about three hours after the pump, while a storm arrives in half an hour.
- Feedback on the settled water crosses the coagulation optimum to its 140 mg/L stop, while feedforward with a slow ratio trim holds the same storm with no time off specification.
- A pH loop worth a factor of 7.7 on the settled-water peak carries no specification of its own.
- A filter-to-waste timer removes a violation that no continuous loop reaches.

Where else this applies. Chapter 4 and Volume 5, Chapter 4 trim a ratio or a prediction from a late measurement.

References

[1] H. Ratnaweera and J. Fettig, “State of the art of online monitoring and control of the coagulation process,” *Water*, vol. 7, no. 11, pp. 6574–6597, 2015.

[2] 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.

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

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

[5] U.S. Environmental Protection Agency, "National primary drinking water regulations, 40 CFR part 141, sections 141.64, 141.73, 141.74 and 141.173 to 141.175," Code of Federal Regulations, Title 40, edition of July 1, 2025, govinfo.gov, 2025.

[6] ——, "National primary drinking water regulations: Interim enhanced surface water treatment; final rule," *Federal Register*, vol. 63, no. 241, pp. 69 478–69 521, 1998, december 16, 1998.

[7] J. C. Crittenden, R. R. Trussell, D. W. Hand, K. J. Howe, and G. Tchobanoglous, *MWH's Water Treatment: Principles and Design*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2012.

[8] A. Amirtharajah and K. M. Mills, "Rapid-mix design for mechanisms of alum coagulation," *Journal AWWA*, vol. 74, no. 4, pp. 210–216, 1982.

[9] J. K. Edzwald, "Coagulation in drinking water treatment: Particles, organics and coagulants," *Water Science and Technology*, vol. 27, no. 11, pp. 21–35, 1993.

[10] Y. Adachi and Y. Tanaka, "Settling velocity of an aluminium-kaolinite floc," *Water Research*, vol. 31, no. 3, pp. 449–454, 1997.

[11] A. Amirtharajah and D. P. Wetstein, "Initial degradation of effluent quality during filtration," *Journal AWWA*, vol. 72, no. 9, pp. 518–524, 1980.

[12] J. D. Paccione, D. M. Follansbee, P. L. Young, and D. M. Dziewulski, "Analytical determination of the baffle factor for disinfection contact systems based on hydraulic analysis, disinfection kinetics, and Ct tables," *Journal of Environmental Engineering*, vol. 142, no. 7, p. 04016026, 2016.

[13] G. Hua and D. A. Reckhow, "Comparison of disinfection byproduct formation from chlorine and alternative disinfectants," *Water Research*, vol. 41, no. 8, pp. 1667–1678, 2007.

[14] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[15] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *Journal of Process Control*, vol. 13, no. 4, pp. 291–309, 2003.

[16] World Health Organization, "Water quality and health: Review of turbidity. information for regulators and water suppliers," World Health Organization, Geneva, Technical brief WHO/FWC/WSH/17.01, 2017.

[17] U.S. Environmental Protection Agency, "Method 180.1: Determination of turbidity by nephelometry, revision 2.0," Environmental Monitoring Systems Laboratory, Office of Research and Development, Cincinnati, OH, Tech. Rep., 1993.

[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] R. K. X. Bastos, D. B. Viana, and P. D. Bevilacqua, "Turbidity as a surrogate for Cryptosporidium removal by filtration in drinking-water QMRA models," *Water Supply*, vol. 13, no. 5, pp. 1209–1219, 2013.

[21] C. W. Baxter, S. J. Stanley, and Q. Zhang, "Development of a full-scale artificial neural network model for the removal of natural organic matter by enhanced coagulation," *Journal of Water Supply: Research and Technology—AQUA*, vol. 48, no. 4, pp. 129–136, 1999.

[22] N. N. Reza, S. Peldszus, and P. M. Huck, "Quantifying drinking water treatment plant robustness with respect to turbidity: Examining raw water and treatment," *Water Research*, vol. 295, p. 125550, 2026.

[23] B. W. Bequette, "Process control practice and education: Past, present and future," *Computers & Chemical Engineering*, vol. 128, pp. 538–556, 2019.

[24] 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.

6.10 Exercises

6.1 Concept check · 2 min

In Scenario 4 the storm passes at 300 min, the raw water falls to 20 NTU and the feedback controller holds the alum at its 140 mg/L stop while the pH loop holds pH 6.90. Why does the settled water get worse, not better, with every further milligram of alum?

(a) The pH loop cannot hold its target against so much alum, so the pH leaves its window
(b) The flocculators shear the larger floc apart once the dose reaches the pump limit
(c) Filters are washing, so the remaining beds carry more of the settled turbidity
(d) The dose is about four times the demand, so the extra coagulant restabilizes the particles

6.2 Concept check · 1 min

The trim r of Equation (6.8) multiplies the computed dose instead of adding a correction to it. Why?

(a) A correction found at 12 NTU raw water stays proportionally right at 60 NTU
(b) A multiplying trim follows the storm faster than an additive trim of equal tuning
(c) A ratio held between 0.80 and 1.35 cannot wind up against a fault
(d) A ratio needs no settled turbidimeter, so the three-hour lag drops out of it

6.3 Concept check · 2 min

The model applies the basin removal of Equation (6.5) algebraically at the outlet instead of as a sink inside each well-mixed tank. What would a sink inside the tanks change?

(a) The settled turbidity at steady state would be lower for every dose
(b) Each basin stage would respond at $V/(Q+v_sA_s)$, so the lag would be far shorter
(c) The coagulation optimum would move to a higher dose ratio
(d) Headloss would stop building on the filters between washes

6.4 Theory and calculation · 6 min

Scenario 4 starts at the design raw water of 12 NTU, 3 mg/L organic carbon and 18 °C, takes the turbidity to 60 NTU, and later lets it fall to 20 NTU. The feedback controller ends at its 140 mg/L stop, and the pH loop holds pH 6.90.

(a) Find the coagulant demand of Equation (6.2) at 12 NTU and at 60 NTU. *Checkpoint:* 31.32 mg/L and 52.73 mg/L.

(b) Find the dose ratio of the 140 mg/L dose at 60 NTU and at 20 NTU. *Checkpoint:* 2.66 and 3.85.

(c) Find the destabilized fraction η of Equation (6.1) for each ratio, and compare it with a dose equal to the demand. *Checkpoint:* 9.4 % and 1.1 %, against 98.5 %.

6.5 Theory and calculation · 6 min

The step of the Introduction moves the alum from 31.3 to 38 mg/L and the settled water from 1.200 to 2.435 NTU, with 2 % of the change at 32.25 min and 63 % at 193.6 min.

(a) Add the time constants of the train from Section 6.3 (dosing 0.5 min, two flocculators, two clarifier stages). Then read a first-order-plus-dead-time model off the step, taking the 2 % time as the dead time. *Checkpoint:* 184.5 min. $\theta \approx 32.25$ min, $\tau \approx 161.35$ min.

(b) Find the steady-state gain of the step, and the gain between 25 and 31.3 mg/L from Table 6.1. *Checkpoint:* $+0.184$ and -0.262 NTU per mg/L.

(c) The feedback loop raises the dose by $K_c = 4\,\mathrm{mg/L}$ per NTU of settled error. Find K_cK_p on each side of the optimum and say on which side the loop is positive feedback. *Checkpoint:* -1.05 below the optimum (negative feedback), $+0.74$ above it (positive feedback).

6.6 Theory and calculation · 6 min

Take Equation (6.7) with $V_c = 600\,\text{m}^3$, BF $= 0.50$ and the by-product coefficient $k_{\text{THM}} = 26$ (µg/L per mg/L of TOC, per mg/L of residual, per hour of contact), the value in Show Model.

(a) Find T_{10}, the CT credit and the by-products at a residual of 1.00 mg/L with 3 mg/L of organic carbon, at 1200 m^3/h and at 1700 m^3/h. *Checkpoint:* 15.0 min, 15.0 mg · min/L, 39 µg/L. At 1700 m^3/h, 10.59 min, 10.59 mg · min/L, 27.5 µg/L.

(b) Show that THM/CT does not depend on the residual or the flow. With 6.5 mg/L of organic carbon, find the by-products when the credit just meets 12 mg · min/L, and the credit when the by-products reach 80 µg/L. *Checkpoint:* 67.6 µg/L and 14.2 mg · min/L.

(c) Find the organic carbon above which no residual meets both limits. *Checkpoint:* 7.69 mg/L.

6.7 App challenge · 8 min

Run Scenario 4, *Feedback against a storm*, with its default settings and the default seed to 300 min, and export the CSV file. Then do the same with Scenario 5, *Feedforward against the same storm*.

(a) For Scenario 4, read `metric_offspec_min` and `turb_set_ntu` in the row at 300 min, and `alum_mgL` and `alum_need_mgL` at 70 min. *Checkpoint:* 141.9 min and 20.92 NTU. 32.2 mg/L dosed against 52.73 mg/L needed.

(b) Read the same four values for Scenario 5. *Checkpoint:* 0.0 min and 3.60 NTU. 52.6 mg/L against 52.73 mg/L.

(c) Read `metric_alum_kg` at 300 min in both files, and explain why the structure with the worse water uses more coagulant. *Checkpoint:* 414.5 kg and 306.0 kg.

Tolerances: 0.1 min, 0.02 NTU, 0.2 mg/L and 1 kg.

6.8 App challenge · 8 min

Run Scenario 7, *The loop that makes the other loop possible*, as loaded, with the default seed, to 300 min. The scenario starts with pH loop gain 0 and sets it to 30 at 200 min.

(a) Read the Off-spec time tile and the largest `turb_set_ntu` in the exported file. *Checkpoint:* About 169.6 min and 9.74 NTU.

(b) Press Reset, set pH loop gain to 30 before Run, run to 300 min and read the same two values. Find the ratio of the settled peaks. *Checkpoint:* 0.0 min and 1.26 NTU, a ratio of about 7.7.

(c) Reload the scenario as shipped, run to 900 min, and read `turb_set_ntu`, `base_mgL` and `ratio_r` in the last row. Explain why the settled water has not recovered. *Checkpoint:* 4.09 NTU, 45 mg/L and 1.35.

Tolerances: 0.2 min, 0.02 NTU and 0.1 mg/L.

6.9 App challenge · 8 min

Run Scenario 8, *Four filters and a queue*, as loaded (Filter sequencing Auto), with the default seed, to 600 min, and export the CSV file.

(a) Read the Off-spec time and Backwashes tiles. Then read the columns `metric_wasted_m3` and `metric_produced_m3` in the last row of the file. *Checkpoint:* 2.9 min, 8, 215 m^3 and 12 193 m^3.

(b) Press Reset, set Filter sequencing to Auto + FTW before Run, and repeat. Find the share of production sent to waste in each run. *Checkpoint:* 0.0 min, 8, 518 m^3 and 12 172 m^3. 1.8 % against 4.3 %.

(c) Find the largest `turb_filt_ntu` in each file. The first run logs a lower peak than the 0.804 NTU of Section 6.5.5. Explain why. *Checkpoint:* About 0.59 NTU and 0.17 NTU.

Tolerances: 0.1 min, 2 m^3 and 0.01 NTU.

6.10 Application area · 3 min

An operator sees the Filtered water tile stay below 0.30 NTU through the storm of Scenario 5 and concludes that the plant met its pathogen-removal obligation.

(a) What does turbidity measure, and why can a plant meet its turbidity limit and still fail a pathogen limit?

(b) Name two features of real filters that the simulator leaves out, and what each would add to an operator's work.

6.11 Application area · 4 min

A manager asks whether a raw-water turbidimeter pays for itself. Use Table 6.3 and the coagulant price of \$0.32/kg in Show Model.

(a) Find the alum cost of the storm under each structure and the saving per storm. *Checkpoint:* \$132.8 and \$97.9, a saving of about \$35.

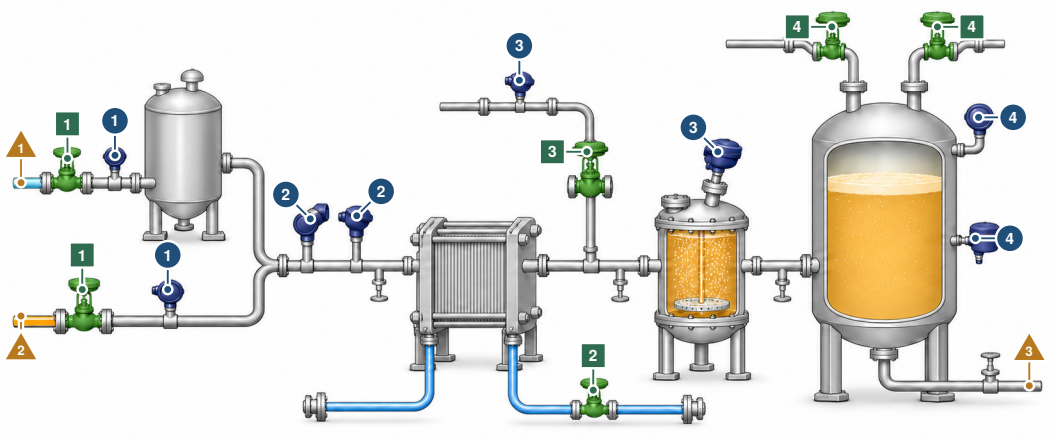
(b) Taking \$3000 as an example price for the instrument, how many such storms does the chemical saving alone need to repay it? What does the case for the instrument rest on instead? *Checkpoint:* About 86 storms. Compliance: 141.9 min off specification and 47.15 min in violation avoided.

6.12 Review the agent · 5 min

An AI assistant reviews Scenario 4 and writes: "The turbidity loop is too sluggish for a three-hour plant. Double the gain to 8 mg/L per NTU and halve the integral time to 60 min. The dose will then follow the storm, and the off-spec time over 300 min should fall by about half, to near 70 min." Find the flaw, and say how you would check the claim in the app.

CHAPTER 7

Soft Drink Carbonation Simulator Whose Foam Limit Moves With Temperature



The controller moves 1 water and syrup valves 2 glycol valve of the chiller 3 carbon dioxide dosing valve 4 gas make-up and vent valves on the tank

The controller reads 1 water and syrup flow meters 2 product temperature and sugar strength 3 gas flow and dissolved gas 4 tank pressure and level

What disturbs it 1 warmer water feed 2 weaker syrup batch 3 filler slowing or stopping

Notice the faint foam on the tank's liquid: the drink can carry the right amount of gas and still foam when it warms, unless the tank's pressure is kept above the gas's own push. Illustration, not to scale.

IN PLAIN TERMS The Right Fizz Is Not Enough

The main idea. A soft drink line mixes syrup and water in a fixed proportion, chills the mix and dissolves carbon dioxide gas in it under pressure. The controllers add syrup and gas in step with the measured flow, the way a cook scales a recipe to the size of the pot.

The challenge. The gas meter reports how much gas is dissolved. Foaming depends on how hard that gas pushes to escape, and a warmer drink pushes harder. So the meter can read right while the bottles foam over.

Why it matters. Only the tank's pressure, more cooling or a slower line keep the drink from foaming at the filling machine.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on the first sparkling water (Section 7.1), the history of the first apparatus and of Henry's law (Section 7.2), the try-it segments of cases 2 and 4 and the fact on the fooled analyzer (Section 7.6) 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 7.6), Technology in Practice (Section 7.7), Limitations (Section 7.8) and the two open challenges

The control question The CO_2 analyzer measures a content, but the filler needs a margin between the tank's top pressure and the product's own gas pressure. How can a line whose analyzer reads its setpoint fill foaming product, and which of cooling, top pressure and throughput buys the margin back?

Learning objectives

- Compute the CO_2 pressure, breakout margin and foam index of the product from Henry's law.
- Explain how a trimmed ratio blend and a dose in ratio to the flow hold the recipe.
- Predict what a warm feed, a weak CO_2 supply or a filler stop costs.
- Tell an analyzer fooled by air or by its thermometer from a real CO_2 error.
- Judge when to cut the rate, hold the filler or raise the top pressure.

Prerequisites Sections 1.4.4 (ratio with a trim), 1.4.3 (static feedforward), 1.6.2 (velocity form) and 1.6.4 (scenarios and exported files).

The app https://apmonitor.com/apps/beverage-carbonation/

See also Volume 4, Chapter 4 and Volume 4, Chapter 6 for a line speed bounded by a constraint, Volume 4, Chapter 5 for ratio loops under a throughput supervisor. Volume 6, Chapter 7 and Volume 6, Chapter 8 for where the production rate is set.

7.1 Introduction

A carbonated soft drink is made from a concentrated syrup and treated water, and the two meet on the way to the filler rather than in a batch. A syrup proportioner delivers the prescribed amount of syrup to the water, the blend is mixed in line and carbonated, and the product goes on to the filler, an arrangement that has always been the main one in the carbonated soft drinks industry [1]. Carbonation and filling are engineered together because the carbon dioxide (CO_2) has to be dissolved, held and filled under pressure [2], and the control of food lines is judged by the out-of-specification product it avoids as much as by its operating cost [3]. Three checks are made on what the filler takes, its sweetness in degrees Brix, its CO_2 in volumes, and whether it fills without foaming.

DID YOU KNOW? Sparkling Water in Two Minutes, 1772

Living near a public brewery, Joseph Priestley poured water back and forth between two vessels held close to the yeast of the fermenting liquor, and in two or three minutes had a glass of pleasant sparkling water. Shaking water hard under the gas, he wrote, gave it more than its own bulk of gas in a few minutes [4].

The first two checks are compositions and the third is a pressure margin. CO_2 stays in solution only while the pressure over the liquid exceeds the pressure the dissolved gas would exert, and at a fixed content that equilibrium pressure rises with temperature [5, 6]. In the line of this chapter the product in the buffer tank at 3.05 °C carries 3.79 vol in equilibrium with 2.70 bar of CO_2 under a headspace at 4.01 bar absolute, a margin of 1.12 bar once its dissolved air and water vapor are counted. When the water feed warms from 12 to 26 °C and the chiller runs out of refrigeration, the CO_2 analyzer stays on 3.80 vol while the tank warms to 10.5 °C, the margin falls to 0.48 bar, and 4.37 m^3 of the 6.67 m^3 filled in 20 min would foam. Bottled carbonated waters show the same dependence, since at a given CO_2 content the number of bubbles likely to form in a glass grows with the water temperature [7].

The simulator is built against one belief, that a carbonation line whose analyzer reads its setpoint is making good product. The analyzer measures a content, the filler needs a margin, and the temperature, the dissolved air and the top pressure move the margin without touching the content. The class of problem is general. Every carbonated product shares it, from champagne to bottled water [7, 8], and so does any liquid in which a gas is held by an imposed pressure, because a supersaturated liquid releases its excess as bubbles grown from gas cavities [9]. The line has a second general shape. Where a plant's production rate is set decides how every inventory in it must be controlled [10], as the recycle loop of Volume 6, Chapter 7 shows for a reactor and its column. Here the rate is set at the

outlet by a filler that slows, stops and speeds up, so the buffer tank's level is held by the flow entering it, and a headspace that shares its volume with that level breathes CO_2 through the vent whenever the level swings. A transfer function from dose to analyzer reading carries neither the margin nor the breathing.

The beverage literature treats carbonation and filling as unit operations [2], and the control literature treats ratio control, feedforward and throughput manipulation as structure [10, 11]. This chapter puts the two on one plant, on which the analyzer can be watched on its setpoint while the product foams. It turns the margin against foaming into something a student can measure, on a soft drink blending and carbonation line that runs on the simulator framework of Section 1.6. Dissolved and free CO_2 are separate states everywhere, the tank's level and headspace are one gas inventory, and what the filler receives is logged apart from what the instruments report. Henry's law, ratio control with a quality trim, analyzer feedback, split pressure control and the coordination logic are standard practice. What the model adds is a 185-state line with finite-rate dissolution through 30 fixed-mass cells, water, sugar, CO_2 and air ledgers that close to round-off, a pressure and temperature analyzer that dissolved air or its own thermometer can fool, and a coordinator that holds the filler when the recipe cannot be made. The model is written out fully enough to rebuild it from this text. Eight cases and one graded shift separate a content from a margin, a ratio from its trim, and an instrument from the product.

Section 7.2 traces the domain from fixed air to the modern filler. Section 7.3 gives the model of the line, from Henry's law through the cells, the chiller, the dose and the tank to what the filler receives and what the analyzer reports. Section 7.4 measures the margin against foaming at the design point and names the three levers that act on it. Section 7.5 describes the eight loops and the line coordinator, and Section 7.6 measures the eight cases and the shift. Section 7.7 sets the two ratio structures on real hardware, Section 7.8 states the limitations, Section 7.9 concludes, and the exercises of Section 7.10 close the chapter.

7.2 Fixed Air, Proportioners and Fillers

Water charged with CO_2 is older than the name of the gas, and the first apparatus and the law of solubility came between 1772 and 1803, as the history below tells.

> ⧖ **HISTORY** **From a Glass Apparatus to Henry's Law, 1775 to 1803**
>
> In 1775 John Mervin Nooth described to the Royal Society a glass apparatus of three vessels for impregnating water with fixed air. He built it because the bladder in Priestley's arrangement tainted the water and the process asked for more skill than most users had [12]. In 1803 William Henry recorded that Nooth's apparatus had in turn been superseded by condensing many times the water's bulk of gas into it, and credited Priestley with first suggesting the influence of pressure. His experiments gave the general law that water under two or three atmospheres takes up twice or three times the gas it takes up at one, and the law now carries his name [5, 13]. The solubility of CO_2 in water has since been correlated from about 100 experimental studies below 1 MPa [6] and compiled with its temperature dependence among the constants of many other gases [13].

Rates came after equilibria. The two-film theory of gas absorption appeared in 1924 [14], and the lumped coefficient $k_L a$ with which the simulator dissolves its gas descends from it. Production moved from the bottle to the line. Syrup proportioners have ranged from a simple pump with a manually adjustable piston stroke to computerized flow rate regulators, while the principle of delivering the prescribed syrup to the treated water stayed the same [1], and the supply and storage of CO_2, carbonation systems and the principles of filling are now treated as one subject [2]. Measurement followed the product. An automatic method for determining the CO_2 in bottled beer was published in 1969 [15], and a sample's equilibrium pressure and temperature give its CO_2 through Henry's law, which is the method the simulated analyzer uses.

Table 7.1 Variables of the Line by Role, With Their Values at the Reset of Case 1

Variable	Role	Design value	Range
B_{103}	CV, AT-103	10.80 °Bx SP	9–13
R	CV, FFC-102	0.2195 SP	0.15–0.30
Q_w	CV, FT-101	17.08 m³/h	—
T_{104}	CV, TT-104	3.0 °C SP	1–12
V_{105}	CV, AT-105	3.80 vol SP	2.5–5
P_t	CV, PT-201	3.0 bar(g) and 3.3 SP	1–5.5
L	CV, LT-201	50 % SP	20–80
B_f, V_f, FI	truth at the filler	10.80, 3.79, 30	see text
x_w, x_s, x_g	MVs, valves	59.3, 52.1, 56.3 %	0–100
$\dot{n}$	MV, CO_2 dose	147 kg/h	0–300
x_{mu}, x_v	MVs, make-up, vent	0.61, 0 %	0–100
Q_{SP}	MV of LIC-201	20 m³/h	0–25
ΔR, ΔV	MVs of the trims	0, −0.054 vol	±0.02, ±0.6
filler	DV, demand	100 %	0–120
T_w, P_w	DVs, water	12 °C, 100 %	4–32, 10–120
B_{batch}	DV, unmeasured	60.0 °Bx	54–66
P_{sup}	DV, CO_2 supply	9.0 bar(g)	0–12
P_{vac}	DV, deaerator	40 mbar	20–600
β	DV, unmeasured	0 °C	±3
w_k, T_k, c_k, g_k, a_k	states, 30 cells	see text	—
A_1, A_2	states, deaerator	0.083, 0.029 mol/m³	—
P_v, B_{syr}	states	40 mbar, 60.0 °Bx	—
T_g, $\dot{n}$, F	states	1.75 °C, 147, 20	—
M_L, S_L	states, tank	2604, 281 kg	—
$H_L = M_L T_L$	state, tank	$T_L = 3.05$ °C	—
N_c, N_a, N_b	states, tank	422.4, 0.11, 0 mol	—
n_c, n_a	states, headspace	435.5, 0.54 mol	—

Foam is the defect the arrangement exists to prevent. Bubble nucleation from gas cavities in supersaturated liquids has been reviewed [9], and in champagne and carbonated beverages bubbles were mostly found to form from tiny gas pockets trapped in cellulose fibers, at a frequency set by the dissolved CO_2, the temperature, the viscosity and the ambient pressure [8]. For bottled carbonated waters holding about 3 to 7 g/L of CO_2, the partial pressure of the gas depends strongly on temperature [7]. The recipe simulated here holds 7.46 g/L. Control of the line grew around these facts. The proportioner holds the recipe as a ratio of two flows, which is ratio control [11] as in Section 1.4.4, a refractometer after the blend corrects the ratio for syrup that is not what the recipe assumes, and the filler sets the rate at which everything upstream must run. The simulator puts these pieces in one line and adds a record that no plant keeps, the true Brix, CO_2 and foam index of every liter the filler takes.

7.3 The Line as Modeled

The simulated line makes 20 m³/h of one soft drink, 10.80 °Bx from a 60.0 °Bx syrup with 3.80 vol of CO_2, chilled to 3.0 °C and held at 3.0 bar(g) for a filler whose demand is imposed. The deaerator, the water and syrup valves, the chiller and its glycol circuit, the CO_2 header, the dissolution section, the buffer tank with its make-up, vent and purge, the filler and all of their instruments appear in Figure 7.1.

Table 7.1 sorts the variables by the part each plays. The state vector has 185 entries, 20 lumped states, 150 cell states and 15 cumulative ledgers of the water, sugar, CO_2, air and cooling that enter and leave. The ledgers are integrated by the same step as the plant, so every balance can be closed from the state itself. Units are SI inside the model, pressures absolute, and the panels show m³/h, kg/h, bar(g), mbar, °Bx and volumes.

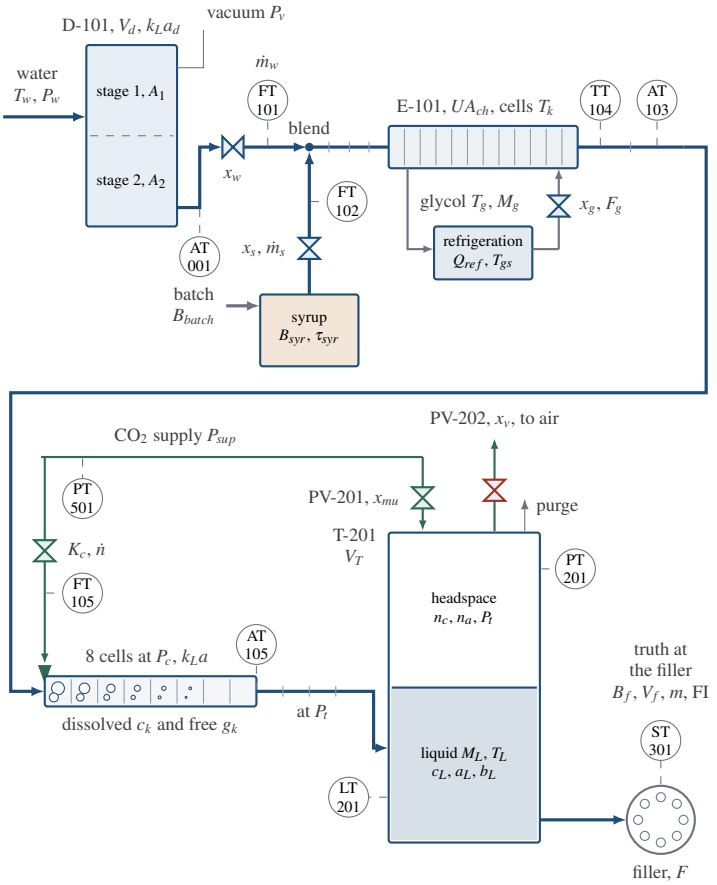


Figure 7.1 Blending and carbonation line as modeled, drawn in two rows. Water at T_w is stripped of air in the two stages of deaerator D-101 (A_1, A_2) at vacuum P_v, passes the water valve x_w and FT-101, and meets syrup of strength B_{syr} from the syrup valve x_s and FT-102 at the blend point. The product crosses the twelve cells of chiller E-101 against glycol at T_g, fed through valve x_g by a refrigeration plant limited to Q_{ref}, passes TT-104 and the refractometer AT-103, and continues in the lower row. There it takes its CO_2 dose $\dot{n}$ from the supply header through the dosing valve and FT-105 at the injector, and dissolves it in eight cells at P_c ahead of the analyzer AT-105. Buffer tank T-201 holds the liquid under a headspace of CO_2 and air at P_t, with make-up PV-201, vent PV-202, a fixed purge, LT-201 and PT-201, and the filler draws F. Dashed lines are measurements, and the open circles in the dissolution section stand for free gas.

7.3.1 Solubility

Every cell, the tank and the analyzer use one solubility law, Henry's law with a van 't Hoff temperature dependence, as shown in Equation (7.1).

$$\begin{aligned} H_i(T) &= H_i^\circ \exp\left[\theta_i\left(\frac{1}{T}-\frac{1}{T_i^\circ}\right)\right], \qquad i \in \{C,A\} \\ p_C &= \frac{\rho\, c}{H_C(T)}, \qquad p_A = \frac{\rho\, a}{H_A(T)}, \qquad V = \frac{\rho\, c}{\nu_1} \end{aligned} \tag{7.1}$$

where c and a are the dissolved CO_2 and air in mol/kg of solution, T is the absolute temperature, $\rho = \rho_B(B)$ is the density of a sucrose solution at Brix B, $H_C^\circ = 3.3 \times 10^{-4}\,\mathrm{mol/(m^3\,Pa)}$ at $T_C^\circ =$

298.15 K with $\theta_C = 2400\,\mathrm{K}$, $H_A^\circ = 9.3 \times 10^{-6}\,\mathrm{mol/(m^3\,Pa)}$ at $T_A^\circ = 288.15\,\mathrm{K}$ with $\theta_A = 1500\,\mathrm{K}$, and $\nu_1 = 44.615\,\mathrm{mol/m^3}$ is one volume, a liter of CO_2 at 0 °C and one atmosphere per liter of beverage, or 1.9635 g/L. The CO_2 constants are the compilation value [13], 1.47 g/L at 25 °C and one atmosphere. The air constants are fitted to water saturated with air, which holds 27.3 mg/L at 15 °C. The density is a cubic fitted to the sucrose table at 20 °C, $\rho_B = 998.282 + B(3.83315 + B(0.0143612 + 2.98508 \times 10^{-5} B))$ in $\mathrm{kg/m^3}$, the heat capacity is $c_p(w) = 4190 - 2350\,w$ in J/(kg K) with $w = B/100$, and the vapor pressure of water p_w follows the Magnus form $610.94 \exp[17.625\,T/(T + 243.04)]$ in Pa with T in °C. At 3 °C the recipe's 3.80 vol are in equilibrium with 2.70 bar of CO_2, and at 10.5 °C with 3.40 bar.

A bubble in contact with the liquid holds water vapor and the dissolved air's own pressure besides CO_2, so the dissolved CO_2 that a bubble can stand at a local pressure P is given by Equation (7.2).

$$c^*(T,P,a,B) = \frac{H_C(T)}{\rho}\,\max\left(0,\; P - p_w(T) - p_A\right) \tag{7.2}$$

7.3.2 Transport, Dissolution and Breakout

The line from the blend point to the tank is a chain of 30 cells of fixed mass $M_c = 5.2\,\mathrm{kg}$, 5 L at $1040\,\mathrm{kg/m^3}$, carried by the blend's mass flow $\dot{m}$ with upwind advection. Four cells lead to the chiller, twelve form the chiller, two lead to the injector, eight form the dissolution section and four lead to the tank. Each cell k holds its sugar mass fraction w_k, its temperature T_k, its dissolved CO_2 c_k, the free CO_2 g_k that moves with the liquid as bubbles, and its dissolved air a_k, all per kilogram of solution, as shown in Equation (7.3).

$$\begin{aligned}
\frac{\mathrm{d}w_k}{\mathrm{d}t} &= q\,(w_{k-1} - w_k), \qquad \frac{\mathrm{d}a_k}{\mathrm{d}t} = q\,(a_{k-1} - a_k)\\
\frac{\mathrm{d}T_k}{\mathrm{d}t} &= q\,(T_{k-1} - T_k) + \chi_k\,\frac{UA_{ch}\,(T_g - T_k)}{12\,M_c\,c_p(w_k)}\\
\frac{\mathrm{d}c_k}{\mathrm{d}t} &= q\,(c_{k-1} - c_k) + r_k\\
\frac{\mathrm{d}g_k}{\mathrm{d}t} &= q\,(g_{k-1} - g_k) - r_k + \delta_{k,18}\,\frac{\dot{n}}{M_c}
\end{aligned} \tag{7.3}$$

where $q = \dot{m}/M_c$, χ_k is one in the twelve chiller cells and zero elsewhere, $\delta_{k,18}$ places the dose in the first cell of the dissolution section (cells are numbered from 0), and the inlet values at $k = -1$ are the blend's composition, temperature and air with no CO_2. Transfer between free and dissolved CO_2 follows Equation (7.4).

$$r_k = k_L a\,\frac{g_k}{g_k + g_s}\,(c_k^* - c_k) - k_b\,\max(0,\; c_k - c_k^*) \tag{7.4}$$

Here c_k^* is Equation (7.2) at the cell's temperature and air and at P_c inside the dissolution section or P_t beyond it. The first term dissolves gas at $k_L a = 0.35\,\mathrm{s^{-1}}$ while it is plentiful, and the factor with $g_s = 0.015\,\mathrm{mol/kg}$ makes the interfacial area vanish with the last bubble. The second term releases gas from a liquid holding more than c_k^* at $k_b = 0.02\,\mathrm{s^{-1}}$, with or without bubbles present. Dissolved and free CO_2 are separate states everywhere, so a bubble exists in the model only where CO_2 was injected and has not dissolved, or where the liquid is supersaturated. The application's own check integrates a still cell with plenty of gas and finds the dissolved CO_2 after $1/k_L a$ within 0.054 % of the first-order 63.2 % of its approach.

7.3.3 Deaerator, Blend and Chiller

The deaerator is two well-mixed stages that strip air-saturated water toward the equilibrium of the vacuum space, as shown in Equation (7.5).

$$\begin{aligned} \frac{\mathrm{d}A_1}{\mathrm{d}t} &= \frac{q_w}{V_d}(A_{in} - A_1) + k_L a_d\,(A^* - A_1) \\ \frac{\mathrm{d}A_2}{\mathrm{d}t} &= \frac{q_w}{V_d}(A_1 - A_2) + k_L a_d\,(A^* - A_2) \\ A_{in} &= H_A(T_w)\,P_{atm} \\ A^* &= H_A(T_w)\,\max\left(0,\; P_v - p_w(T_w)\right) \end{aligned} \tag{7.5}$$

where A_1 and A_2 are the stage air concentrations in mol/m^3, $q_w = \dot{m}_w/\rho_w$ is the water volume flow with $\rho_w = 999.7\,\mathrm{kg/m^3}$, $V_d = 0.5\,\mathrm{m^3}$ per stage, $k_L a_d = 0.15\,\mathrm{s^{-1}}$, and the vacuum follows its disturbance through $\tau_{vac}\,\mathrm{d}P_v/\mathrm{d}t = P_{vac} - P_v$ with $\tau_{vac} = 20\,\mathrm{s}$. At 40 mbar the water leaves with 0.31 mg/L of oxygen, taken as 37 % of the dissolved air by mass. The valves and the blend follow Equation (7.6).

$$\begin{aligned} \dot{m}_w &= x_w\,\dot{m}_{w,max}\sqrt{P_w/100}, \qquad \dot{m}_s = x_s\,\dot{m}_{s,max} \\ w_{-1} &= \frac{\dot{m}_s\,w_s}{\dot{m}}, \qquad \tau_{syr}\frac{\mathrm{d}B_{syr}}{\mathrm{d}t} = B_{batch} - B_{syr} \\ T_{-1} &= \frac{\dot{m}_w c_p(0)\,T_w + \dot{m}_s c_p(w_s)\,T_{syr}}{\dot{m}_w c_p(0) + \dot{m}_s c_p(w_s)} \end{aligned} \tag{7.6}$$

where $\dot{m} = \dot{m}_w + \dot{m}_s$, $w_s = B_{syr}/100$, P_w is the water header in percent of design, and the blend's air is $a_{-1} = (\dot{m}_w A_2/\rho_w + \dot{m}_s a_{syr})/\dot{m}$, with a_{syr} the 4 mg/L of air the syrup carries. Each valve responds to its command as a first-order lag, 1 s for water and syrup, 5 s for glycol and 0.5 s for make-up and vent. The glycol side of the chiller is one well-mixed temperature fed by a refrigeration plant of limited duty, as shown in Equation (7.7).

$$\begin{aligned} M_g c_g \frac{\mathrm{d}T_g}{\mathrm{d}t} &= F_g c_g\,(T_{gs}^e - T_g) + \sum_k \chi_k \frac{UA_{ch}}{12}\,(T_k - T_g) \\ T_{gs}^e &= \max\left(T_{gs},\; T_g - \frac{Q_{ref}}{F_g c_g}\right), \qquad F_g = x_g F_{g,max} \end{aligned} \tag{7.7}$$

The plant cools the returning glycol by at most $Q_{ref} = 320\,\mathrm{kW}$, so beyond that load the glycol supply itself warms and more glycol flow no longer helps. At the design flow the chiller takes the 13.02 °C blend to 3.00 °C with 228 kW and the glycol valve at 56.3 %. A 26 °C water feed raises the blend to 25.24 °C, which needs 506 kW at 20 m³/h, and 320 kW would leave it at 11.19 °C at that rate or hold 3.0 °C up to 12.6 m³/h.

7.3.4 Carbonator Pressure and the Dose

The dissolution section is held above the tank by a back-pressure that falls with the square of the flow, and the CO_2 dosing valve passes gas in proportion to the root of the difference of the squared pressures on its two sides, as shown in Equation (7.8).

$$\begin{aligned} P_c &= P_t + \Delta P_c\left(\frac{\dot{m}}{\dot{m}_0}\right)^2, \qquad \dot{n}_{max} = K_c\sqrt{P_{sup}^2 - P_c^2} \\ \tau_n \frac{\mathrm{d}\dot{n}}{\mathrm{d}t} &= \min(\dot{n}_{cmd}, \dot{n}_{max}) - \dot{n} \end{aligned} \tag{7.8}$$

where $\dot{m}_0 = \rho_B(B_{rec})\,Q_0$ is the design mass flow, $\Delta P_c = 1.2\,\mathrm{bar}$, $K_c = 0.178\,\mathrm{mol/(s\,bar)}$ with the pressures in bar under the root, $\tau_n = 1\,\mathrm{s}$ and the command $\dot{n}_{cmd}$ is clamped to 0 to 300 kg/h. At the design flow the section runs at 4.20 bar(g), where pure CO_2 could hold 7.3 vol at 3 °C, and the valve can pass 241 kg/h against the 147 kg/h dosed. At a 5.0 bar(g) supply it can pass 84.5 kg/h, 57 % of the 149 kg/h that the recipe needs at the full rate.

7.3.5 Buffer Tank and Headspace

The tank is one well-mixed liquid, with its own holdup of rising bubbles, under a headspace of CO_2 and air, as shown in Equation (7.9).

$$\begin{aligned}
\frac{dM_L}{dt} &= \dot{m} - \dot{m}_f, \qquad \frac{dS_L}{dt} = \dot{m}\,w_{29} - \dot{m}_f\,w_L \\
\frac{dN_c}{dt} &= \dot{m}\,c_{29} - \dot{m}_f\,c_L + M_L\,(r_s + r_b) \\
\frac{dN_a}{dt} &= \dot{m}\,a_{29} - \dot{m}_f\,a_L + M_L\,r_{sa} \\
\frac{dN_b}{dt} &= \dot{m}\,g_{29} - M_L\,r_b - N_b/\tau_{rise} \\
\frac{dH_L}{dt} &= \dot{m}\,T_{29} - \dot{m}_f\,T_L + UA_t\,(T_{amb} - T_L)/c_p(w_L)
\end{aligned} \tag{7.9}$$

where M_L is the liquid mass, S_L its sugar, N_c and N_a its dissolved CO_2 and air, N_b the CO_2 in its bubbles and $H_L = M_L T_L$ its heat content, with $c_L = N_c/M_L$, $a_L = N_a/M_L$, $b_L = N_b/M_L$ and $w_L = S_L/M_L$. The filler draws $\dot{m}_f = \rho_L F$. The liquid exchanges CO_2 with the headspace across its surface at $r_s = k_L a_s (c_s - c_L)$, with $c_s = H_C(T_L)\,y_c\,(P_t - p_w)/\rho_L$, and air likewise at r_{sa}, and it exchanges CO_2 with its own bubbles at $r_b = k_L a_t\,b_L/(b_L + g_s)\,(c_t^* - c_L) - k_b \max(0, c_L - c_t^*)$, where c_t^* is Equation (7.2) at the tank. Bubbles rise into the headspace with $\tau_{rise} = 15\,\text{s}$. The headspace is an ideal gas at the liquid's temperature, as shown in Equation (7.10).

$$\begin{aligned}
\frac{dn_c}{dt} &= \dot{n}_{mu} - y_c(\dot{n}_v + \dot{n}_p) + N_b/\tau_{rise} - M_L\,r_s \\
\frac{dn_a}{dt} &= -(1 - y_c)(\dot{n}_v + \dot{n}_p) - M_L\,r_{sa} \\
P_t &= \frac{(n_c + n_a)\,R\,T_L}{V_T - V_L} + p_w(T_L)
\end{aligned} \tag{7.10}$$

Here $y_c = n_c/(n_c + n_a)$, $V_L = M_L/\rho_B(B_L)$, the headspace is kept to at least 2 % of V_T, and the make-up, vent and purge flows are $\dot{n}_{mu} = x_{mu} K_{mu} \sqrt{P_{sup}^2 - P_t^2}$, $\dot{n}_v = x_v K_v \sqrt{P_t^2 - P_{atm}^2}$ and $\dot{n}_p = K_p \sqrt{P_t^2 - P_{atm}^2}$, with bar under the roots. A rising level compresses the headspace. At the 50 % design level the headspace is 2.5 m^3, so 0.32 m^3 of product pushed in with no gas removed would raise it from 4.01 to 4.60 bar absolute. The make-up and the vent have their own setpoints, 3.0 bar(g) and 3.3 bar(g), so the pressure floats between them without spending CO_2, and the purge bleeds 0.61 kg/h of CO_2 at 3.0 bar(g) so that the air the product gives up does not collect. Liquid above 95 % of the tank leaves through the vent line and is booked as overflow, the filler stops itself below 8 % and restarts above 15 %, and standing product in the chiller colder than its freezing point, 0.061 °C lower per °Bx, is flagged as a freeze risk.

7.3.6 What the Filler Receives and What AT-105 Reports

The filler's product is judged on the true tank state. Its breakout margin and foam index follow Equation (7.11).

$$\begin{aligned}
m &= P_t - \left[p_C(c_L, T_L) + k_N\,p_A(a_L, T_L) + p_w(T_L)\right] \\
FI &= 100\,\text{clamp}\big(1 - m/M_{ok},\ 0,\ 1\big)
\end{aligned} \tag{7.11}$$

where $k_N = 5$ counts dissolved air five times over, on the model's reasoning that air bubbles do not redissolve and each one seeds CO_2 breakout, and $M_{ok} = 1.6\,\text{bar}$. The index is 0 with a margin of 1.6 bar or more, 50 at the limit of 0.8 bar and 100 at breakout. The foam index and its limit are an illustrative indicator of fill stability built from that margin, and they have not been validated against a filler. A liter the filler draws is in specification while $|B_f - 10.80| \le 0.10\,°\text{Bx}$, $|V_f - 3.80| \le 0.15\,\text{vol}$

Table 7.2 Parameters of the Line. Visible Entries Are Sliders in the Show Model Panel, and Hidden Entries Never Reach It

Symbol	Meaning	Value	Class
B_{rec}, V_{rec}	recipe	10.80 °Bx, 3.80 vol	visible
—	tolerances at the filler	±0.10 °Bx, ±0.15 vol	visible
—	foam index limit	50	visible
M_{ok}, k_N	foam index scale, air weight	1.6 bar, 5	hidden
V_d	deaerator stage volume	0.5 m^3	hidden
$k_L a_d$	deaerator stripping	0.15 s^{-1}	visible
τ_{vac}	vacuum settling	20 s	hidden
T_{syr}, a_{syr}	syrup temperature, air	20 °C, 4 mg/L	hidden
τ_{syr}	syrup tank mixing	150 s	visible
$\dot{m}_{w,max}$, $\dot{m}_{s,max}$	valve capacities	8.0, 2.0 kg/s	hidden
—	water, syrup valve lags	1, 1 s	visible
UA_{ch}, $F_{g,max}$	chiller, glycol valve	55 kW/K, 30 kg/s	visible
T_{gs}, Q_{ref}	glycol supply, refrigeration	−2 °C, 320 kW	visible
M_g, c_g	glycol holdup	150 kg, 3600 J/(kg K)	hidden
—	glycol valve lag	5 s	hidden
$k_L a$, k_b	dissolution, breakout	0.35, 0.02 s^{-1}	visible
g_s	free gas halving the area	0.015 mol/kg	hidden
ΔP_c	carbonator back-pressure	1.2 bar	visible
K_c	dosing valve	0.178 mol/(s bar)	visible
—	dose range, lag	300 kg/h, 1 s	hidden
V_T, $k_L a_s$	tank, surface transfer	5.0 m^3, 5×10^{-5} s^{-1}	visible
$k_L a_t$, τ_{rise}	bubbles in the tank	0.02 s^{-1}, 15 s	hidden
UA_t, T_{amb}	tank heat gain	60 W/K, 22 °C	hidden
K_{mu}, K_v, K_p	make-up, vent, purge	0.25, 0.60, 0.001	hidden
—	overflow, low-level stop	95, 8 %	hidden
Q_0	filler design demand	20 m^3/h	visible
—	filler ramp lag	3 s	hidden
—	AT-103, AT-105 lags	4, 10 s	visible
—	flow, TT-104, PT, LT, AT-001	0.5, 3, 0.5, 2, 30 s	hidden
M_c, v_1	cell mass, one volume	5.2 kg, 44.615 mol/m^3	fixed

K_{mu}, K_v and K_p in mol/(s bar).

and $FI \leq 50$, with $B_f = 100\, w_L$ and $V_f = \rho_L c_L / v_1$, and the scorecard books each liter as in or out of specification, with the reason.

The analyzer AT-105 at the last cell of the dissolution section measures its sample's equilibrium pressure and temperature and converts them with Henry's law at the temperature its own thermometer reports, as shown in Equation (7.12).

$$V_{AT} = \frac{H_C(T+\beta)}{v_1}\Big[p_C + p_A + p_w(T) - p_w(T+\beta) - \mu\, p_A\Big] \tag{7.12}$$

where β is the bias of that thermometer, and $\mu = 1$ for a multiple-expansion method that separates the air and 0 for the single-expansion default, which reads the dissolved air's pressure as CO_2. At the 1.3 mg/L of air in case 1 that error is 0.054 vol, and a 2 °C bias lowers H_C by 6.1 %, about 0.23 vol at the recipe.

🕮 **Pause.** Try Exercise 7.2 before reading on.

7.3.7 Numbers, Initialization and Numerics

The parameters are listed in Table 7.2. Each scenario opens at the steady state of its own configuration, solved at reset, and a setting changed before the run starts solves it again, as Reset does, so a tank pressure setpoint chosen then starts the tank at that pressure. The flows are those the filler demand asks for at the recipe ratio, the glycol temperature puts TT-104 on its setpoint through the chiller's cell recursion, the dose is found by bisection so that the analyzer reads its setpoint, the dissolved and free CO_2 of every cell are found by bisection, and the tank's dissolved CO_2, bubbles and headspace

composition by nested bisection and a fixed-point iteration. At reset no state derivative outside the ledgers exceeds 2.7×10^{-15} in any of the nine scenarios. The engine of Section 1.6.1 advances the plant by a fixed 0.1 s step of the fourth-order Runge-Kutta method, the controllers act every 0.2 s, the historian keeps one row a second, and the default seed is 20261021. The speed selector spans 1 to 60 times real time, 10 by default. Halving the step, with the controllers still acting every 0.2 s, moves the CO_2 vented in 600 s of case 5 by 3×10^{-6} % and the worst foam index in 900 s of case 2 by 1.3×10^{-7}. Doubling the cells to 60 of 2.5 L moves the filler's CO_2 in case 1 by 0.0004 vol and the lowest filler Brix of case 3 by 0.000 03 °Bx. Over the 2400 s challenge the water, sugar, CO_2 and air ledgers close to 8.2×10^{-15} relative, and closure below 1×10^{-9} is one of the 39 checks the application passes.

7.3.8 Assumptions

The line is plug flow through 5 L cells with the numerical mixing of upwind advection, and bubbles move with the liquid without displacing it, although at the injector the gas volume is comparable to the liquid's. The last traces of free gas depend on the grid, 0.012 % of the dose at AT-105 with 30 cells and 0.0008 % with 60, which matters to no reported result.

Henry's law uses the pure-water constants with a constant van 't Hoff slope. Against the correlation of about 100 low-pressure studies [6] the model's solubility is 3.5 % low at 25 °C, 6.7 % low at 3 °C and 7.8 % low at 0 °C, so it overstates the CO_2 pressure of a product at a given content. The salting-out by sugar, a few percent at 11 °Bx, would act the other way and is not modeled, and neither is the non-ideality of CO_2 at a few bar.

Air is one pseudo-species with a fixed oxygen share, and the deaerator is two well-mixed stages with one stripping coefficient. The dissolution coefficient, the bubble-area function, the breakout rate and the tank's surface coefficient are illustrative constants rather than correlations for a particular mixer or tank, and the dissolution coefficient does not rise with the turbulence of the flow.

The filler is a volumetric demand that can be slowed, held or starved, and its own mechanics, counter-pressure, snift and fill height, are outside the model. The tank is well mixed, its headspace an ideal gas at the liquid's temperature. The glycol side is one temperature, the refrigeration plant a duty limit, and freezing is flagged rather than modeled. Densities are those at 20 °C, valves are linear with first-order actuators and no stiction, and the pumps are represented only by the valve capacities. All sizes, capacities and limits are teaching values for one recipe, not a design basis.

7.4 The Margin Against Foaming

Three measurements fix the design point of case 1. The dissolution section's $1/k_La$ is 2.9 s against a residence of 7.2 s, so 59 % of the dose is still gas in the first cell, 11 % in the third, 0.84 % in the fifth and 0.012 % at AT-105, and the product reaches the tank 27.0 s after the blend point. In the tank the product is undersaturated by more than a bar, in equilibrium with 2.70 bar under a 4.01 bar headspace, so it gains 0.037 vol through its surface in its 7.5 min residence and reaches the filler at 3.79 vol. Counting its air five times over and its water vapor, the product's gas pressure is 2.90 bar, a margin of 1.12 bar and a foam index of 30.

The margin is set by the temperature, the dissolved air and the top pressure, and by the CO_2 only through the recipe, as shown in Figure 7.2. With 1.28 mg/L of air the product reaches the foam limit at 6.3 °C under 3.0 bar(g) and at 11.2 °C under 3.5 bar(g), and it breaks out at 13.9 °C and 18.1 °C. Half a bar of top pressure therefore moves the foam limit by 4.9 °C, and near the design point the margin falls by 0.09 bar for every degree. That shift comes mostly from the slope of the product's CO_2 pressure, about 0.1 bar per degree, so it moves little with the illustrative constants, from 4.5 to 5.4 °C for a limit drawn anywhere between 0.4 and 1.2 bar of margin and 4.85 °C with the air counted once. Raising the temperature setpoint of case 1 in steps on the running line finds the same limits within a few tenths of a degree, near 6.5 °C under 3.0 bar(g) and near 11.4 °C under 3.5 bar(g). They lie slightly above the curves because the tank's product holds 3.78 vol rather than 3.80.

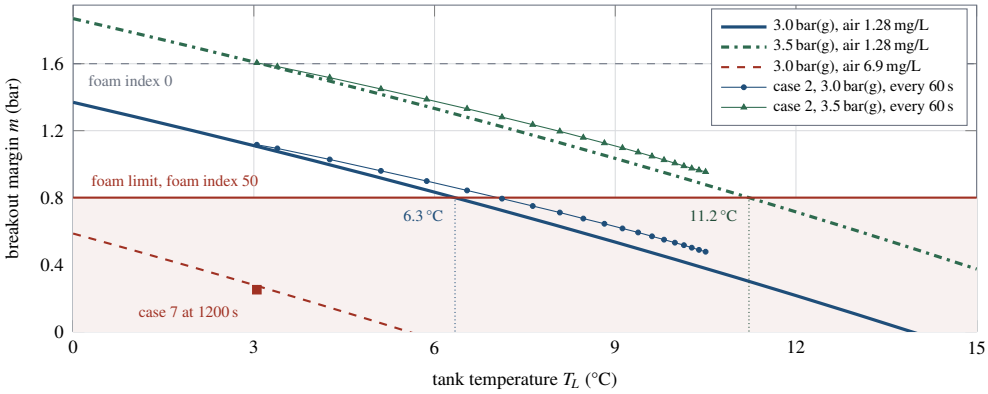


Figure 7.2 Breakout margin of the recipe product against the tank temperature. The three curves are Equation (7.11) for 3.80 vol at 10.80 °Bx with the tank at its setpoint and a fixed dissolved air: solid at 3.0 bar(g), dash-dotted at 3.5 bar(g), and dashed at 3.0 bar(g) with the higher air of case 7. The shaded band lies below the foam limit. The thin paths of markers are the engine's own tank states in case 2 every 60 s to 1200 s, circles as shipped and triangles with the tank at 3.5 bar(g) from the start, and as the tank warms they run slightly above the curves because water at 26 °C brings less air and the tank pressure floats inside the make-up and vent gap.

Air enters the margin five times over. At the 6.9 mg/L that the weak deaeration of case 7 reaches, the air's term is 1.02 bar against 0.19 bar at 1.28 mg/L, and the product passes the limit at any temperature on the chart. The analyzer sees none of this, because the CO_2 content that it measures does not change.

Three levers act on the margin. Cooling lowers the product's CO_2 pressure but is bounded by the 320 kW refrigeration plant. The top pressure raises the tank's side of the margin and costs a headspace of CO_2 and a higher make-up and vent. A throughput cut lowers the cooling load until 320 kW suffices and costs volume. A controller that holds the CO_2 content perfectly uses none of them, which is why the shipped case 2 foams while the CO_2 and pressure loops hold their setpoints.

7.5 Implemented Controller

Eight loops and a line coordinator, FY-100, run every $\Delta t_c = 0.2$ s, two physics steps, with the integrals advancing by that period. They follow the order of work of Figure 1.10, with the overflow, freeze and filler logic after the integration, so a setting changed between two executions reaches the plant at the next one.

7.5.1 Instruments

No loop reads a cell or the tank contents. Each transmitter takes the true value at its own location, lags it exactly to first order and adds first-order colored noise. The lags are 0.5 s for FT-101 and FT-102, 4 s for the refractometer AT-103, 3 s for TT-104, 10 s for AT-105, 0.5 s for PT-201 and FT-105, 2 s for LT-201 and 30 s for the dissolved oxygen meter AT-001. Noise standard deviations are 0.1 % of reading on FT-101 and FT-102, 0.2 % on FT-105, 0.005 °Bx on AT-103, 0.02 °C on TT-104, 0.008 vol on AT-105, 0.003 bar on PT-201 and 0.05 % on LT-201. FT-101, FT-102 and FT-105 read zero below 0.1 m^3/h, 0.02 t/h and 0.5 kg/h. The product flow the loops use is inferred, $Q_m = (\dot{m}_{w,m} + \dot{m}_{s,m})/\rho_B(B_{103,m})$, from the two mass meters and the refractometer. The dosing controller reports when its valve is at capacity, the one plant status any loop reads.

Table 7.3 Loop Pairings, Directions and Shipped Tunings

Tag	CV	MV	γ	K_c	τ_I, output
FIC-101	Q_w	x_w	$+1$	2.5	1 s, 0–100
FFC-102	$\dot{m}_s$	x_s	$+1$	14	1.5 s, 0–100
AIC-103	B_{103}	ΔR	$+1$	0.003	6 s, ±0.02
TIC-104	T_{104}	x_g	-1	8	40 s, 0–100
AIC-105	V_{AT}	ΔV	$+1$	0.6	15 s, ±0.6
PIC-201	P_t	x_{mu}	$+1$	150	120 s, 0–100
PIC-202	P_t	x_v	-1	150	120 s, 0–100
LIC-201	L	Q_{SP}	$+1$	0.5	300 s, 0–25

Gains in % per m³/h, % per t/h, kg/kg per °Bx, % per °C, vol per vol, % per bar, % per bar and m³/h per %, in the order of the rows.

7.5.2 The PI Algorithm

All eight loops use the shared PI law in its velocity form, as shown in Equation (7.13).

$$\begin{aligned} e_k &= \gamma(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\, e_k \\ S_k &= S_{k-1} + (P_k - P_{k-1}), \qquad v_k = b_k + S_k + I_{k-1} \\ u_k &= \mathrm{slew}\big(\mathrm{clamp}(v_k, u_{min}, u_{max}),\, u_{k-1},\, \rho_u\big) \\ I_k &= I_{k-1} + \Big[\frac{K_c}{\tau_I}\, e_k + \frac{u_k - v_k}{\tau_I}\Big]\Delta t_c \end{aligned} \tag{7.13}$$

where $\gamma = -1$ for the two loops whose valve must open as the measurement rises and $+1$ for the other six, b_k is an output bias, and ρ_u is a rate limit that only the level loop uses. This is the framework's law of Section 1.6.2, Equations (1.9) to (1.12), with $\gamma = a$, $b_k = u_b$ and the slew inside the saturation. No loop here uses derivative action, so S_k equals P_k, and the loops that integrate use back-calculation with a tracking time of τ_I. In MANUAL each loop tracks by setting $I_k = u_{op} - b_k - S_k$, so both transfers are bumpless. The pairings and shipped tunings are listed in Table 7.3.

7.5.3 Ratio and Brix Trim

LIC-201 sets the throughput Q_{SP}, and the coordinator splits it into the two flow setpoints at the ratio in force, as shown in Equation (7.14).

$$\begin{aligned} \dot{m}_t &= Q_{SP}\, \rho_B(B_{SP}), \qquad R = R_{SP} + \Delta R \\ Q_{w,SP} &= \frac{\dot{m}_t}{(1+R)\,\rho_w}, \qquad \dot{m}_{s,SP} = R\,\rho_w\,\hat{Q}_w \end{aligned} \tag{7.14}$$

where $R_{SP} = 10.8/49.2 = 0.2195\,\mathrm{kg/kg}$ makes 10.80 °Bx from 60.0 °Bx syrup, and $\hat{Q}_w$ is the measured water flow $Q_{w,m}$ on the default basis or the water setpoint $Q_{w,SP}$ on the setpoint basis. On the first the syrup follows the water it sees, and on the second both flows are asked for at once, the two bases of Section 1.4.4. AIC-103 trims ΔR on the refractometer, and ΔR adds to the ratio that multiplies the water flow, as in Equation (1.6). Both trims integrate only while the product flow exceeds 20 % of design, while the throughput setpoint, lagged by 15 s and its rate by 10 s, moves slower than 1 m³/h per minute, and once 0.25 m³ has passed since the flow last fell below 5 %. Outside those conditions they track, because a Brix or CO_2 error during a ramp or a flush says nothing about the syrup or the dose. A flow setpoint below its meter's cutoff shuts the valve.

7.5.4 The CO_2 Dose and the Dosing Balance

AIC-105 does not set the dose. The dose is the measured product flow times the setpoint volumes plus a trim, as shown in Equation (7.15).

$$\dot{n}_{cmd} = Q_m \rho_1 (V_{SP} + \Delta V), \qquad V_{bal} = \frac{\dot{n}_m}{\rho_1 Q_m} \tag{7.15}$$

where $\rho_1 = 1.9635\,\mathrm{kg/m^3}$ per volume, ΔV is the PI output on AT-105, and $\dot{n}_m$ is FT-105, the dosing controller's own mass-flow reading. The ratio part moves the dose with the flow at once, and the trim corrects the volumes. While the dosing valve is at capacity and the error is positive, AIC-105 runs with $\tau_I = \infty$, conditional integration against a limit the trim cannot move [11]. The dosing balance V_{bal} is an independent measurement of the same content, the CO_2 actually dosed per liter, in the manner of redundant flow and inventory data reconciled against a balance [16]. With the check on, a disagreement with AT-105 of more than 0.15 vol that lasts 60 s raises an alarm and returns the dose to the ratio alone, ΔV decaying with a 30 s time constant until the two agree within 0.075 vol.

7.5.5 Tank Pressure and Level

PIC-201 holds 3.0 bar(g) with the make-up valve and PIC-202 opens the vent only above 3.3 bar(g), each on its own setpoint, so the pressure floats in the gap between them. At reset the vent, resting shut below its setpoint, starts with no integral, so that measurement noise cannot crack it open. LIC-201 sets the throughput with the filler's speed as its output bias, so the throughput follows the filler and the level loop only trims. A change of the feedforward setting is offset in the integral, so the throughput does not jump, and the throughput setpoint is rate limited to 10 $\mathrm{m^3/h}$ per minute, with its upper limit lowered to the coordinator's rate cap. One $\mathrm{m^3/h}$ of imbalance moves the level 0.33 % per minute, an integrating plant.

7.5.6 The Coordinator

FY-100 stops, cuts and holds. A filler demand below 5 % of design is a stop, and with stop handling on the coordinator closes both flow valves, the dose and the glycol valve at the same instant, holding TIC-104's output for the restart. The feasibility test uses the dosing valve's own capacity curve with the measured supply and tank pressures, as shown in Equation (7.16).

$$\begin{aligned} Q_{feas} &= \max\left\{ Q \le 25 : \dot{n}_{max}(Q) \ge 1.05\, Q V_{SP}\, \nu_1 \right\} \\ \dot{n}_{max}(Q) &= K_c \sqrt{P_{sup,m}^2 - P_c(Q)^2} \\ P_c(Q) &= P_{t,m} + \Delta P_c\, (Q/Q_0)^2 \end{aligned} \tag{7.16}$$

where Q is in $\mathrm{m^3/h}$, converted to $\mathrm{m^3/s}$ in the inequality. With the cut on, the rate cap is Q_{feas}, no lower than 6 $\mathrm{m^3/h}$, while the chiller cap falls at 2 $\mathrm{m^3/h}$ per minute whenever the glycol valve is saturated and TT-104 is more than 0.2 °C above its setpoint, and rises at 1 $\mathrm{m^3/h}$ per minute while TT-104 is less than 0.1 °C above its setpoint and the glycol valve is below 95 %. The filler draws the smaller of its demand and the cap. When the supply is less than 0.3 bar above the tank setpoint, or Q_{feas} is below 6 $\mathrm{m^3/h}$, no rate can make the recipe, and after 10 s of that the hold stops the filler and records why, releasing it after 30 s of feasibility. In MANUAL the operator sets the five valves and the dose, the filler draws its demand, and the coordinator neither cuts nor holds.

7.5.7 Simulation, Control and Internal Models

The model of Section 7.3 is the simulation model. The application identifies no control model of its own, and open-loop steps in MANUAL from the case 1 line-up measure the plants that the loops see, each step made at 60 s and timed from that instant. A 10 % step of the glycol valve moves the true product temperature at TT-104 by −0.039 °C per percent, 63 % of the way in 16.2 s. A 10 kg/h step of the dose adds 0.254 vol at the analyzer's location, 63 % in 8.9 s, and 18.7 s after the analyzer's

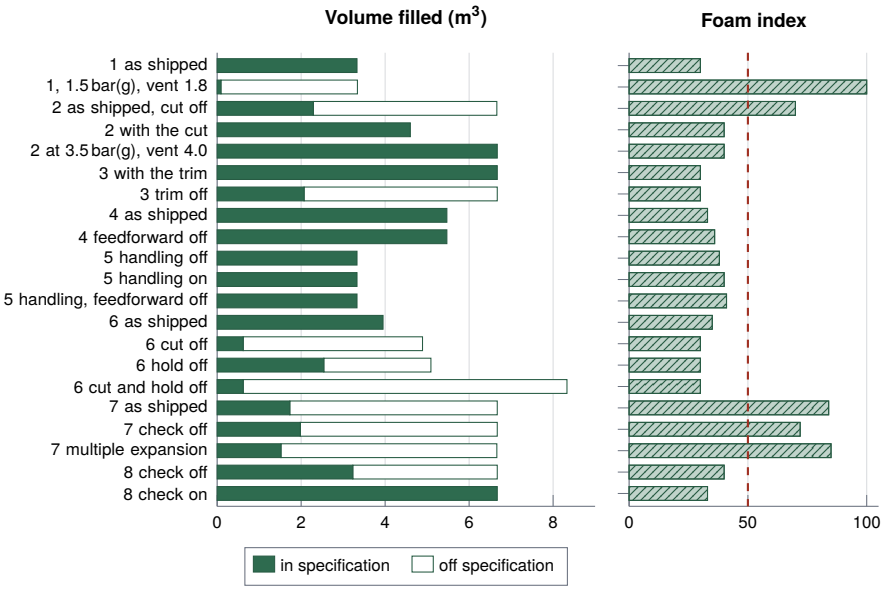


Figure 7.3 Table 7.5 redrawn: for every case and variant, the volume the filler took in and off specification, and the worst foam index, whose limit of 50 is the dashed line. Runs last 1200 s except where the table's caption says otherwise. Notice that the off-specification bars of case 1 at low top pressure and of cases 2 and 7 come with a foam index past its limit, while those of cases 3, 6 and 8 come with a foam index inside it: there the product failed on its Brix or its CO_2 content.

lag. A 2 % step of the syrup valve adds 0.338 °Bx at the refractometer, 63 % in 18.2 s. The filler sees the CO_2 through the tank, 63 % of the change it shows an hour after the step in 454 s, close to the tank's 7.5 min residence. The controllers' internal models are the recipe split of Equation (7.14), the ideal-gas volume ρ_1 in the dose, the valve curve of Equation (7.16), the density fit in Q_m and Henry's law inside the analyzer, and none of them contains the margin.

7.6 Scenarios and Measured Results

Table 7.4 lists the eight cases and the shift. Each case opens on the reference configuration, which has the filler feedforward, stop handling, the cut, the hold, both trims and the dosing-balance check on, the syrup on the measured-water basis and the single-expansion analyzer, and it changes only what its row names. The numbers below are the engine's own, read at the default seed from runs without the interface, and every time is in simulated seconds. A configuration that differs from the case's is set at $t = 0$ before the run starts, as a learner does after choosing the case, and a change meant as a step is made after the run has started. The scorecard's counts for each case are collected in Table 7.5, and Figure 7.3 draws them as bars. The exported file holds the filler's true Brix, CO_2, margin and foam index beside the refractometer and analyzer readings, with the running totals of product, CO_2 vented and cooling (Section 1.6.4).

7.6.1 Nominal Production and the Top Pressure

Case 1 blends 17.08 m^3/h of water deaerated to 0.31 mg/L of oxygen with 3.75 t/h of syrup into 20 m^3/h of product, doses 147 kg/h of CO_2 and uses 7.47 kg of CO_2 per m^3 filled. AT-105 reads 3.80 vol while the product carries 3.75 vol, because the single-expansion analyzer counts the 1.3 mg/L of dissolved air as CO_2 and the trim, at –0.054 vol after reset, holds that reading. Lowering the top pressure to 1.5 bar(g), with the vent at 1.8, a minute into the run shows what the margin is for. The vent releases 4.86 kg of CO_2 in the next minute, the margin turns negative 71 s after the change, the

Table 7.4 The Eight Cases and the Shift

#	Scenario and event	Teaching point
1	Nominal production	where the gas dissolves
2	Water 12 to 26 °C at 60 s, cut off	cooling is solubility
3	Syrup batch 58.0 °Bx at 60 s	the trim, the tank filters
4	Filler to 60, then 100 %	which flow is the master
5	Filler stop, 60 s to 360 s, handling off	a stop breathes CO_2
6	Supply to 5.0, then 2.6 bar(g)	cut, then hold
7	Vacuum 40 to 300 mbar	air foams, fools AT-105
8	AT-105 thermometer +2 °C, check off	a biased analyzer
9	2400 s shift	graded challenge

Table 7.5 Truth-Log Scorecard for Each Case and Variant. Runs Last 1200 s, Except 600 s for Case 1 (From the Change at 60 s in Its Second Row), 900 s for Case 5 and 1500 s for Case 6

Case and configuration	In spec (m^3)	Off spec (m^3)	Vented (kg)	Foam index
1 as shipped	3.33	0	0.10	30
1, 1.5 bar(g), vent 1.8 at 60 s	0.09	3.25	6.81	100
2 as shipped, cut off	2.29	4.37	0.21	70
2 with the cut	4.60	0	0.20	40
2 at 3.5 bar(g), vent 4.0	6.67	0	0.23	40
3 with the trim	6.67	0	0.20	30
3 trim off	2.07	4.60	0.20	30
4 as shipped	5.47	0	0.21	33
4 feedforward off	5.47	0	2.00	36
5 handling off	3.33	0	1.36	38
5 handling on	3.33	0	0.16	40
5 handling and feedforward off	3.33	0	6.78	41
6 as shipped	3.95	0	0.26	35
6 cut off	0.62	4.27	0.26	30
6 hold off	2.54	2.55	0.26	30
6 cut and hold off	0.62	7.71	0.26	30
7 as shipped	1.73	4.94	0.20	84
7 check off	1.98	4.69	0.20	72
7 multiple expansion	1.52	5.14	0.20	85
8 check off	3.23	3.44	0.20	40
8 check on	6.67	0	0.20	33

foam index reaches 100, and 3.25 m^3 of the 3.33 m^3 filled in the ten minutes after the change would foam while 6.81 kg of CO_2 leaves through the vent.

7.6.2 A Cooling Shortage Is a Solubility Shortage

In case 2 the water warms at 60 s, the glycol valve is wide open from 90 s, and the refrigeration plant delivers its full 320 kW, so its glycol supply warms from −2.0 °C to 6.5 °C by 600 s, when TT-104 reads 11.2 °C and the tank has reached 8.5 °C. The CO_2 loop holds the content at 3.80 vol, but the margin falls from 1.12 to 0.68 bar by 600 s, the foam index passes 50 at 413 s and reaches 70 by 1200 s, when the tank is at 10.5 °C and 3.8 vol are in equilibrium with 3.40 bar. The plant view at 900 s, with the foam index at 66, is shown in Figure 7.4, while the filler's CO_2 is still on its band. By 1200 s, 4.37 m^3 of the 6.67 m^3 filled would foam.

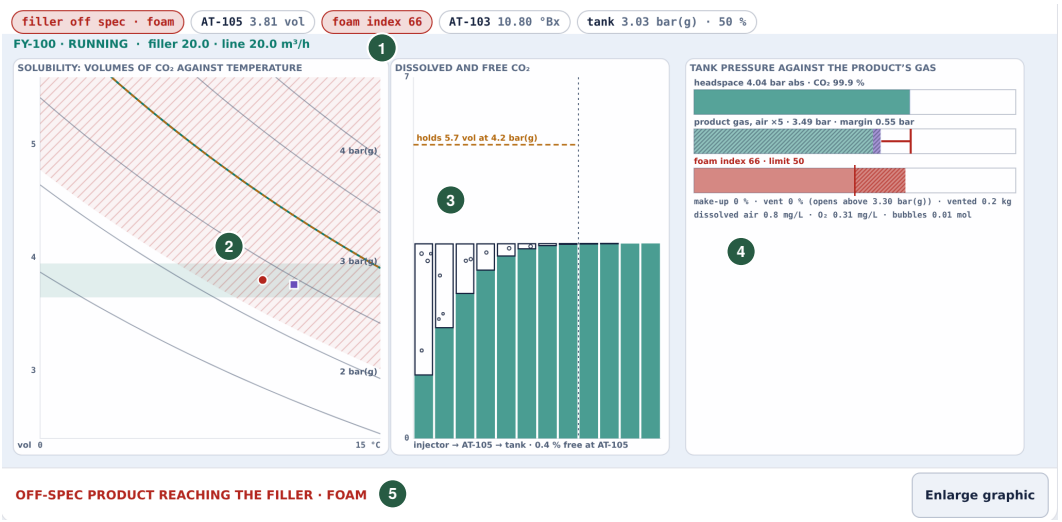


Figure 7.4 The carbonation line in case 2, warmer feed, at $t = 15.00$ min (900 s, a clock outside the crop), 840 s after the water feed warmed from 12 to 26 °C, with the throughput cut off as shipped, cropped to the Carbonation view. (1) The status chips and the FY-100 line, with the foam index at 66 and the line running at the filler's 20 m³/h. (2) The solubility chart: the tank's point, marked off specification, sits inside the hatched foam region beside the carbonator outlet's point (the run has them at 9.8 °C and 3.80 vol, and at 11.2 °C and 3.76 vol), and the highlighted 3 bar(g) curve carries the tank's own dashed saturation line. (3) The dissolution profile: free gas shrinks cell by cell under a section that holds 5.7 vol at 4.2 bar(g), with 0.4 % of the CO_2 still free at AT-105. (4) The tank panel: a headspace of 4.04 bar absolute at 99.9 % CO_2 against a product gas pressure of 3.49 bar, a margin of 0.55 bar, the foam gauge at 66 past its limit of 50, both valves shut and 0.2 kg vented. (5) The status line: off-spec product reaching the filler, foam. Outside the crop, over a 20 min window, the trends show the filler's true CO_2 at 3.804 vol and AT-105 at 3.806 vol on the recipe band, the true CO_2 at the analyzer at 3.763 vol, the dose at 148.5 kg/h against its 149.4 kg/h feedforward, the foam index climbing through its limit at 413 s to 66, the tank pressure floating at 3.03 bar(g) inside its gap and the level at 49.9 %, and the loop scorecard gives the glycol valve saturated for 83.7 % of the run.

TRY IT IN THE APP **2 · Warmer Feed**

Load the scenario, press Run and run to 1200 s. Watch the analyzer AT-105: it stays on 3.80 vol throughout. Then switch to the Carbonation view and watch the foam index. It passes its limit of 50 at 413 s and reaches 70, and 4.37 m³ of the 6.67 m³ filled would foam.

The three levers of Section 7.4 give three answers. With the cut on, the coordinator starts slowing the line and the filler at 90 s, as soon as the glycol valve saturates. Every liter stays in specification and the tank peaks at 4.9 °C, but only 4.60 m³ are filled, and the rate does not settle. Between 1200 and 2400 s of a longer run the filler's draw cycles between 11.69 and 13.41 m³/h with a period of about 227 s, around a mean of 12.72 m³/h against the 12.6 m³/h that 320 kW can hold. The cap falls while the glycol valve is saturated and TT-104 is hot, rises once TT-104 has recovered, and the glycol holdup delays both, so the cut overshoots in each direction. Volume 4, Chapter 4 meets the same difficulty in a feed rate held against constraints that are measured late. With the cut off and the tank at 3.5 bar(g), vent 4.0, from the start, the tank still warms to 10.5 °C, but the foam index peaks at 40 and all 6.67 m³ are in specification. The pressure's CO_2 lies in the headspace, which holds 2.4 kg more at 3.5 bar(g) than at 3.0 bar(g) and starts the case full, so the run itself uses 7.43 kg per m³ against 7.44,

while the cut spends 22.7 kWh of cooling per m^3 filled against 15.7.

Pause. Try Exercise 7.1 before reading on.

7.6.3 The Brix Trim and the Tank as a Filter

In case 3 a 58.0 °Bx batch enters the syrup tank at 60 s and mixes in with its 150 s time constant. At the recipe ratio the blend would settle at 10.44 °Bx. The true Brix at the refractometer, which sees each liter 16.2 s after the blend point, reaches its lowest value, 10.715 °Bx, at 161 s, and AIC-103 raises the ratio by 0.0094 kg/kg, 4.3 % more syrup, by 1200 s. The tank, holding 7.5 min of production, averages the dip, and the filler's worst is 10.777 °Bx at 412 s. With the trim off the filler leaves the band at 372 s and reaches 10.48 °Bx by 1200 s, and 4.60 m^3 of the 6.67 m^3 filled are off specification.

7.6.4 Following the Filler

In case 4 the filler slows to 60 % at 60 s and returns at 600 s. The feedforward brings the throughput from 20 to 12 m^3/h in 49 s at the ramp limit, the level stays between 49.2 and 51.0 % and the tank between 2.96 and 3.07 bar(g), so nothing is vented but the purge. On the measured-water basis the syrup lags the water through the ramps, and the true Brix at the refractometer peaks at 10.943 °Bx at 125 s and dips to 10.648 °Bx at 634 s, which the tank averages to within 0.010 °Bx at the filler. On the setpoint basis it stays between 10.772 and 10.823 °Bx. The measured-water basis earns its place when the water valve cannot follow its setpoint. With the advanced water header dip to 30 % fired at 700 s, the measured-water basis keeps the filler at 10.80 °Bx, while the setpoint basis drives it to 11.33 °Bx and fills 1.69 m^3 off specification by 1200 s. Without the feedforward the level loop alone lets the level swing from 42.9 to 58.4 %, the pressure reaches 3.35 bar(g), and 1.8 kg of CO_2 is vented beyond the purge.

> **TRY IT IN THE APP** **4 · Filler Slowdown**
>
> Run the scenario and watch the tank level while the filler slows to 60 % and returns: it stays between 49.2 and 51.0 %. Reload the scenario with its reload button, set Filler-speed feedforward Off on the FY-100 card before 60 s and run again. The level loop alone lets the level swing from 42.9 to 58.4 %, and 1.8 kg of CO_2 is vented beyond the purge.

7.6.5 A Stop That Breathes

In case 5 the filler stops at 60 s for 300 s, with stop handling off, and the feedforward ramps the blend down at 10 m^3/h per minute. The 0.32 m^3 made meanwhile lifts the level to 56.4 % at 204 s, squeezes the headspace by an eighth and raises the pressure to 3.39 bar(g) at 134 s, over the vent setpoint, so the vent and the purge release 1.36 kg of CO_2 in 900 s, against 0.16 kg from the purge alone when the stop is handled. The restart sends a standing slug past the refractometer, whose true Brix swings from 11.15 to 10.48 °Bx. With stop handling the feeds, the dose and the glycol close together at 60 s, the level does not rise, only the purge is vented, and the product standing in the chiller cools no lower than –0.16 °C, above its freezing point of –0.66 °C at 10.8 °Bx. The restart has its own cost, since the filler draws 20 m^3/h at once while the blend ramps back, and the level dips to 43.5 % at 481 s. Without the feedforward and without stop handling the level climbs to 69.7 %, the pressure to 3.47 bar(g), and 6.8 kg of CO_2 is vented. With stop handling but no feedforward the restart is left to the level loop, which brings back only 1.6 m^3/h by 395 s, so the product trickling through the chiller cools to –0.68 °C and is flagged as a freeze risk, and the level sinks to 29.0 % at 766 s.

Pause. Try Exercise 7.3 before reading on.

7.6.6 An Infeasible Recipe

In case 6 the CO_2 supply sags from 9.0 to 5.0 bar(g) at 60 s. The coordinator reads the valve curve with PT-501 and PT-201, finds that about 13.95 m^3/h can be carbonated (13.87 to 14.01 m^3/h with the noise on PT-201), and slows the line and the filler to that rate from 61 s, so the product stays in

specification. At 600 s the supply drops to 2.6 bar(g), below what the tank needs, no rate is feasible, and the filler is held at 611 s. The tank keeps its product while it waits, its pressure sagging to 2.94 bar(g) by 1200 s with a 1.04 bar margin left, and the hold is released at 1230 s, 30 s after the supply returns. The line fills 3.95 m^3 of the 8.33 m^3 the filler asked for, none of it off specification. Without the cut the dose saturates and off-specification product is filled from 112 s, 4.27 m^3 of it. Without the hold the uncarbonated blend reaches the filler from 660 s, 2.55 m^3 falling to 2.61 vol by 1229 s, and with neither 7.71 m^3 are off specification, some of it at 0.78 vol. The hold loses volume, and the off-specification product it prevents could not have been recovered. The steady balance decides which recipes are reachable at all, as it decides which compositions are reachable in Volume 2, Chapter 3.

7.6.7 Instruments That Can Be Fooled

In case 7 the deaerator's vacuum rises to 300 mbar at 60 s, the oxygen after the deaerator rises from 0.31 to 3.04 mg/L, and the dissolved air at the filler climbs from 1.3 to 6.9 mg/L by 1200 s. Its saturation pressure, 0.04 bar at the start, 0.10 bar at 300 s and 0.20 bar at 1200 s, counts five times in the foam index, which passes its limit at 311 s, and 4.94 m^3 of the 6.67 m^3 filled would foam. The single-expansion analyzer counts the same air as CO_2 and reads 0.32 vol high. The dosing balance sees it disagree with the CO_2 dosed per liter, alarms at 165 s and returns the dose to the ratio, so the CO_2 at the filler stays in specification. With the check off the trim cuts the dose until the filler's CO_2 has fallen to 3.55 vol by 1200 s, and 4.01 m^3 are off for CO_2 as well as foam. The multiple-expansion analyzer never makes the error, but 5.14 m^3 still foam, because only the vacuum can remove the air.

In case 8 AT-105's own thermometer reads 2 °C high from 60 s, with the check off. Believing its sample warmer, the analyzer credits its pressure with 6.1 % less CO_2 and reads low, and the trim raises the dose until the reading is back at 3.80 vol. The true CO_2 reaches 3.99 vol at the analyzer and 4.01 vol at the filler by 1200 s, over the 3.95 vol limit from 581 s, and 3.44 m^3 of the 6.67 m^3 filled are off specification while every instrument reads normally. With the check on, the analyzer and the dose per liter differ by more than 0.15 vol for a minute, the alarm comes at 127 s, and the filler stays in specification, at worst 3.84 vol. An independent balance that the plant already has catches both faults, as redundant flow and inventory data can [16]. In Volume 11, Chapter 13 a drifting analyzer likewise takes a plant outside its permit while the loop it feeds reads on target.

> **DID YOU KNOW? Every Instrument Reads Normally**
>
> From the simulator: with its thermometer 2 °C high and the dosing balance off, the analyzer of case 8 fills 3.44 m^3 off specification and 3.23 m^3 in it. With the balance on, all 6.67 m^3 are in specification (Table 7.5).

7.6.8 The Shift Challenge

The challenge runs 2400 s. The filler slows to 70 % at 300 s and returns at 600 s, the water warms to 26 °C from 700 to 1800 s, the filler jams from 1200 to 1440 s and speeds up to 110 % at 1860 s, and a 58.8 °Bx batch arrives at 1600 s. It starts without the filler feedforward, stop handling or the cut, with a loose level loop ($K_c = 0.3$, $\tau_I = 900$ s) and the vent at 3.05 bar(g). The target asks for at least 11.5 m^3 in specification, at most 0.05 m^3 off, at most 3 kg of CO_2 vented, a foam index never above 50 while filling, a level never above 80 % with no overflow, and at most 16 kWh of cooling per m^3 filled, with the model parameters and disturbances as given. A changed model parameter or disturbance voids the attempt. Table 7.6 compares the measured configurations.

As shipped the line fills 10.98 m^3 in specification and 0.82 m^3 that would foam, vents 10.6 kg of CO_2, lets the level reach 70.7 % and the foam index 58, and uses 15.2 kWh of cooling per m^3. The suggested configuration fills all 11.80 m^3 in specification, with the foam index peaking at 39 while the tank warms to 9.6 °C, the level between 43.3 and 50.6 %, 0.77 kg vented and 14.1 kWh per m^3. Each element pays for itself. With the feedforward and stop handling but the tank left at 3.0 bar(g), 3.16 m^3 still foam, and with the cut on at 3.0 bar(g) every liter is in specification but only 9.52 m^3 are

Table 7.6 Challenge Configurations Over 2400 s of a Filler Demand of 11.80 m^3

Configuration	In spec (m^3)	Off (m^3)	Vented (kg)	Foam index	Level (%)
as shipped	10.98	0.82	10.62	58	40.7–70.7
suggested[a]	11.80	0	0.77	39	43.3–50.6
suggested, 3.0 bar(g), vent 3.3	8.64	3.16	1.47	68	43.3–50.6
the same, cut on	9.52	0	0.41	45	44.3–50.5

[a] feedforward and stop handling on, cut off, 3.5 bar(g), vent 4.0, level loop $K_c = 0.5$ and $\tau_I = 300$ s

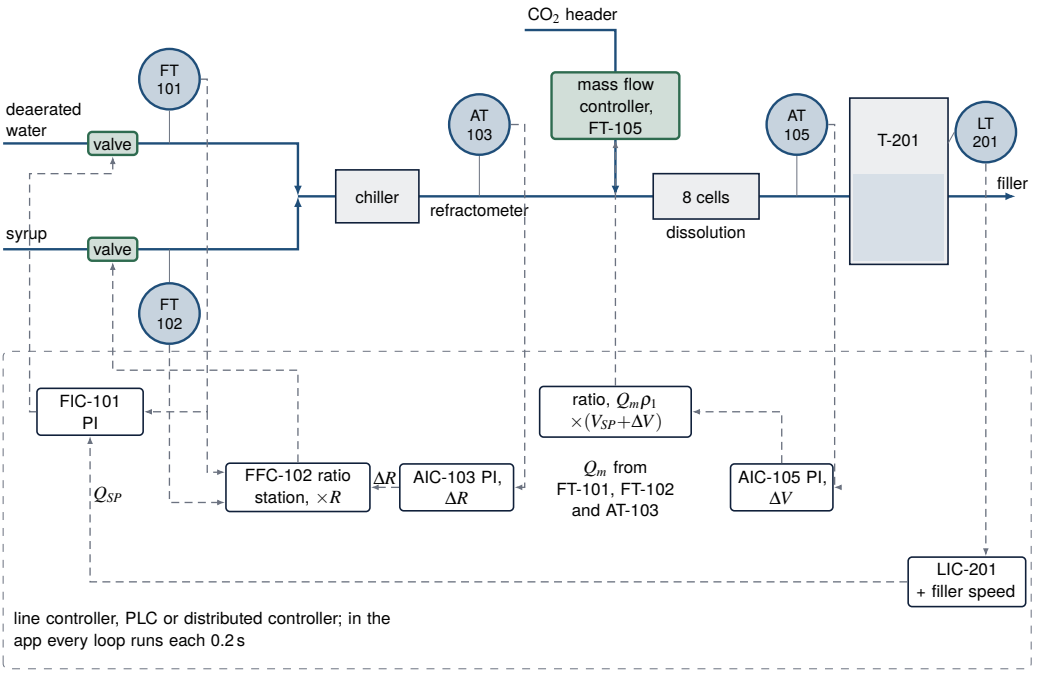


Figure 7.5 The two ratio structures of the line on real hardware. The water mass meter FT-101 is the wild flow: FIC-101 holds it at its share of the throughput Q_{SP} that LIC-201 sets with the filler speed as its feedforward, and the ratio station FFC-102 multiplies it by the ratio, trimmed by AIC-103 on the refractometer, to give the syrup setpoint. The dose block multiplies the measured product flow by the setpoint volumes plus the trim of AIC-105 on the pressure-and-temperature analyzer, and sends the result to the gas mass flow controller at the injector. Solid arrows are process streams and dashed gray arrows are signals. Notice that the analyzer only trims a dose that already follows the flow, so a wrong analyzer moves the product off specification without upsetting any loop.

filled. The coordination keeps the level and the CO_2, and the top pressure keeps the rate. Against the target, the shipped configuration fails four of the bounds, the in-specification volume by 0.52 m^3 and the off-specification volume more than 16 times over. The suggested one passes with 0.30 m^3, 2.2 kg, 11 points of foam index and 1.9 kWh per m^3 to spare.

7.7 Technology in Practice

Figure 7.5 draws the two ratio structures of Section 7.5, the trimmed syrup ratio and the CO_2 dose in ratio to the measured flow, with their hardware.

What Measures It

Four measurements carry the ratios. The water and syrup mass meters give the wild and the captive flow, the refractometer AT-103 trims the syrup ratio, and the CO_2 analyzer AT-105 trims the dose. The analyzer is the one that can be fooled. It infers a content from a pressure and a temperature, and the international method for sparkling products shows how strongly that conversion leans on temperature: a pressure read at 3 °C is multiplied by 1.68 to give its value at 20 °C [17]. In the model AT-105 is Equation (7.12) behind a 10 s lag, with its thermometer bias β and the single-expansion reading of air as its two ways to be wrong.

> **TECHNOLOGY Sensor: Equilibrium Pressure and Temperature**
>
> A CO_2 meter of this kind does not see the gas directly. It brings its sample to equilibrium, reads that pressure and the sample temperature, and converts them with Henry's law, whose constant depends on temperature [13]. The international wine method reads the same equilibrium pressure in a sealed bottle with a needle gauge, either a Bourdon tube or a piezoelectric cell. A class 1 gauge of 1000 kPa full scale reads within 10 kPa. A reading taken away from 20 °C is multiplied by a correction factor, 1.68 at 3 °C, which shows how much the temperature weighs [17]. An error in that temperature, or air counted as CO_2, shifts the reading with nothing on its trend to show it, as cases 7 and 8 of this chapter show.

What Moves It

The dose leaves the CO_2 header through a gas mass flow controller, whose own sensor is FT-105 and whose valve curve, Equation (7.8), sets the most it can pass against the carbonator pressure. Its calibration assumes one gas, which suits a header of pure CO_2 [18]. The water, syrup and glycol valves are modeled as linear with first-order lags of 1 s, 1 s and 5 s, and the make-up and vent valves at 0.5 s.

> **TECHNOLOGY Final Element: A Gas Mass Flow Controller**
>
> A thermal mass flow controller combines a flow sensor, a throttling valve and its own control electronics in one small body. The sensor holds a heated element in the gas beside an unheated one, and the heat the flow carries away measures the mass flow. The calibration holds for one gas only, because it rests on that gas's specific heat, and a change from turbulent to laminar flow shifts it. Such meters are less accurate than Coriolis meters but far cheaper, and suit purified gases of stable composition [18]. The chapter's dose controller is of this kind: FT-105 is its own reading, and its valve reports when it is at capacity.

Where the Controller Runs

Programmable logic controllers act as local controllers under a supervisory control scheme, or as the primary controller of a smaller system, and their programs hold logic, timing, counting, PID control and arithmetic [19]. Control cycles range from milliseconds to minutes [19]. In the app every loop and the coordinator run each 0.2 s, against a dose that reaches 63 % of its effect at the analyzer's location in 8.9 s. The structures map onto standard blocks. A signal multiplying relay, or ratio station, multiplies the wild flow by a factor and sends the product to the captive flow controller as its setpoint [18]. FFC-102 is that block with the factor $R_{SP} + \Delta R$ from AIC-103, and the dose command of Equation (7.15) is a second one whose factor is $V_{SP} + \Delta V$.

> **IN PRACTICE Check the Gauge Before the Product**
>
> The international method for sparkling wine, resolution OENO 21/2003, reads a bottle only after its temperature has been stable for at least 24 h, and only once shaking has brought the pressure to a constant value. The gauges are verified at least once a year on a test bed against a reference gauge of a higher class traced to national standards, comparing the two at rising and at falling pressures and adjusting the tested gauge when they differ [17]. Here the dosing balance is the analyzer's second opinion.

What Is Modernizing It

Three developments bear on this line. The first is a direct optical reading of dissolved CO_2. Infrared sensors usually need a gas sample or an equilibrated headspace, while attenuated total reflectance reads the liquid itself through an evanescent wave at a crystal face. For dissolved CO_2 it has so far been used mainly with laboratory instruments, and compact devices are now being shown [20]. The second is rapid, standardized assessment of bubbles and foam by near-infrared spectroscopy and computer vision in place of slow manual tests [21]. The third is the industrial internet of things, with networked instruments and computing at the edge that acts on data where it is collected [19]. The same connectivity raises the need for resilience, safety and security [19].

7.8 Limitations

The foam index is an illustrative indicator built from the breakout margin with dissolved air weighted five times, and it has not been validated against a filler. The filler itself is a demand, without counter-pressure, snift or fill-height mechanics.

The solubility is the pure-water compilation value with a constant van 't Hoff slope, 3.5 to 7.8 % below a critically evaluated correlation between 25 and 0 °C [6], without the salting-out by sugar or the non-ideality of CO_2. The dissolution, breakout and surface rates are constants chosen for the teaching point rather than correlations for a mixer or tank.

> ⚠ **OPEN CHALLENGE** **How Much Gas a Sweet Drink Holds**
>
> Sugars lower the solubility of CO_2 in water, and the more sugar, the lower it is. Models of the gas in sugar, water and alcohol are hard to calibrate because measured data are limited, and for the mixture with alcohol one main gap is the lack of data at low pressures [22]. An analyzer that converts pressure to content with a pure-water law inherits that gap. Progress would be measured solubility for real recipes at filling pressures, built into the analyzer's conversion and into the margin a coordinator protects.

The instruments fail only through the thermometer bias and the air the analyzer misreads, with no laboratory sampling schedule, valves are linear and free of stiction, and every size and limit is a teaching value for one recipe. The warm spell of case 2 and the thermometer error of case 8 arrive as steps, which a plant would not deliver, but steps separate one cause from the others.

> ⚠ **OPEN CHALLENGE** **Judging Foam Without Trial and Error**
>
> The traditional ways of judging the bubbles and foam of carbonated drinks are slow and costly for small and medium producers, and partly subjective. A review calls for rapid, standardized methods such as near-infrared spectroscopy and computer vision, and notes that trial and error still dominates the industry [21]. The foam index of this chapter has not been validated against a filler. Progress would be an online measure of fill stability that a controller can close a loop on, in place of a margin computed from a model.

7.9 Conclusion

A soft drink carbonation line is a plant in which the analyzer measures a content and the filler needs a margin, and the simulator of this chapter logs the true tank state beside every instrument so that the two can be compared. At the design point the product in the tank at 3.05 °C is 1.12 bar from breakout, and the foam limit lies at 6.3 °C under 3.0 bar(g) and at 11.2 °C under 3.5 bar(g). A 26 °C water feed that outruns the 320 kW chiller takes the tank to 10.5 °C while the analyzer stays on 3.80 vol, and 4.37 m^3 of 6.67 m^3 would foam. A throughput cut saves the product by filling 4.60 m^3 at a rate that cycles every 227 s, while half a bar of top pressure fills all 6.67 m^3. A supply below the tank's own pressure is held rather than filled, avoiding 2.55 m^3 of off-specification product, and a dosing balance catches both an analyzer that reads dissolved air as CO_2 and one whose thermometer, 2 °C high, would otherwise fill 3.44 m^3 over the CO_2 limit. The shift improves from 0.82 m^3 off specification to

all 11.80 m^3 in specification. The model is plug flow with an illustrative foam index and constant transfer coefficients. A fuller model would add a filling model with counter-pressure, the salting-out by sugar, a dissolution coefficient that follows the turbulence, and a laboratory sampling schedule. Within those limits the line shows that a gas held in solution by pressure defines a margin, not a setpoint.

KEY POINTS

- The analyzer measures how much gas is in the drink, but the filler needs a margin of pressure, and a warmer drink loses that margin while the analyzer reads its setpoint.
- A trimmed ratio blend and a dose in ratio to the measured flow hold the recipe, while the filler-speed feedforward keeps the level and the vent quiet.
- Cooling, top pressure and throughput are the three levers on the margin: half a bar of top pressure fills all 6.67 m^3 where the throughput cut fills 4.60 m^3.
- A dosing balance catches both an analyzer that counts air as CO_2 and one whose thermometer reads high.

Where else this applies. Chapter 6 also computes a dose from measured flows and trims it on a slow analyzer, and Chapter 4 trims a ratio on a measured quality.

References

[1] M. Shachman, "The syrup proportioning system," in *The Soft Drinks Companion*. CRC Press, 2004, pp. 133–156.

[2] D. P. Steen, "Carbon dioxide, carbonation and the principles of filling technology," in *Carbonated Soft Drinks: Formulation and Manufacture*, D. P. Steen and P. R. Ashurst, Eds. Wiley, 2006, pp. 112–143.

[3] T. A. Haley and S. J. Mulvaney, "Advanced process control techniques for the food industry," *Trends in Food Science & Technology*, vol. 6, no. 4, pp. 103–110, 1995.

[4] J. Priestley, "Observations on different kinds of air," *Philosophical Transactions of the Royal Society of London*, vol. 62, pp. 147–264, 1772.

[5] W. Henry, "Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures," *Philosophical Transactions of the Royal Society of London*, vol. 93, pp. 29–274, 1803.

[6] J. J. Carroll, J. D. Slupsky, and A. E. Mather, "The solubility of carbon dioxide in water at low pressure," *Journal of Physical and Chemical Reference Data*, vol. 20, no. 6, pp. 1201–1209, 1991.

[7] G. Liger-Belair, "Carbon dioxide in bottled carbonated waters and subsequent bubble nucleation under standard tasting condition," *Journal of Agricultural and Food Chemistry*, vol. 67, no. 16, pp. 4560–4567, 2019.

[8] G. Liger-Belair, M. Parmentier, and P. Jeandet, "Modeling the kinetics of bubble nucleation in champagne and carbonated beverages," *The Journal of Physical Chemistry B*, vol. 110, no. 42, pp. 21 145–21 151, 2006.

[9] S. F. Jones, G. M. Evans, and K. P. Galvin, "Bubble nucleation from gas cavities — a review," *Advances in Colloid and Interface Science*, vol. 80, no. 1, pp. 27–50, 1999.

[10] R. M. Price, P. R. Lyman, and C. Georgakis, "Throughput manipulation in plantwide control structures," *Industrial & Engineering Chemistry Research*, vol. 33, no. 5, pp. 1197–1207, 1994.

[11] 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.

[12] J. M. Nooth, "The description of an apparatus for impregnating water with fixed air; and of the manner of conducting that process," *Philosophical Transactions of the Royal Society of London*, vol. 65, pp. 59–66, 1775.

[13] R. Sander, "Compilation of Henry's law constants (version 4.0) for water as solvent," *Atmospheric Chemistry and Physics*, vol. 15, no. 8, pp. 4399–4981, 2015.

[14] W. K. Lewis and W. G. Whitman, "Principles of gas absorption," *Industrial & Engineering Chemistry*, vol. 16, no. 12, pp. 1215–1220, 1924.

[15] T. C. Rees, "An automatic method for the determination of carbon dioxide in bottled beer," *Journal of the Institute of Brewing*, vol. 75, no. 5, pp. 465–468, 1969.

[16] R. S. Mah, G. M. Stanley, and D. M. Downing, "Reconcillation and rectification of process flow and inventory data," *Industrial & Engineering Chemistry Process Design and Development*, vol. 15, no. 1, pp. 175–183, 1976.

[17] International Organisation of Vine and Wine, "Carbon dioxide method of analysis in wines: Overpressure measurement of sparkling wines," Resolution OENO 21/2003, Compendium of International Methods of Wine and Must Analysis, Paris, 2003.

[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] M. R. Hasan, A. Nikeghbal, S. Tran, F. S. Sium, S. Noh, H. Kim, and C. H. Mastrangelo, "A MEMS microbolometer-based ATR mid-infrared sensor for direct dissolved CO_2 detection and UV-induced sediment carbon assay in aquatic environments," *Sensors*, vol. 26, no. 9, p. 2689, 2026.

[21] C. Gonzalez Viejo, D. D. Torrico, F. R. Dunshea, and S. Fuentes, "Bubbles, foam formation, stability and consumer perception of carbonated drinks: A review of current, new and emerging technologies for rapid assessment and control," *Foods*, vol. 8, no. 12, p. 596, 2019.

[22] S. Al-Hasnaawei, K. M. Batoo, H. Doshi, A. Yadav, R. Roopashree, A. Kashyap, K. Jayabalan, S. Ray, and A. A. Ibrahim, "Prediction of carbon dioxide solubility in sugar-water-alcohol solutions at high pressure for application to sparkling drinks," *Scientific Reports*, vol. 16, p. 25168, 2026.

7.10 Exercises

7.1 Concept check · 2 min

In case 2 the CO_2 loop holds AT-105 on 3.80 vol while 4.37 m^3 of the 6.67 m^3 filled would foam. What does the analyzer fail to see?

(a) The free CO_2 that has not yet dissolved when the product reaches AT-105
(b) A rise of the CO_2 content that the 10 s lag of AT-105 hides
(c) The dissolved air, which the single-expansion method reads as CO_2
(d) The warmer tank, which raises the gas pressure of an unchanged content

7.2 Concept check · 1 min

At the design point of case 1, AT-105 reads 3.80 vol while the product at its location carries 3.75 vol. Why does it read high?

(a) It counts the pressure of the dissolved air as CO_2
(b) It samples free CO_2 that has not yet dissolved
(c) Its thermometer reads high, so it credits more CO_2
(d) Its 10 s lag leaves it behind a rising content

7.3 Concept check · 2 min

In case 5 the filler stops for five minutes with stop handling off, and 1.36 kg of CO_2 leaves through the vent and the purge in 900 s, against 0.16 kg with stop handling on. Where does the extra CO_2 come from?

(a) The dose carries on after the stop and overcarbonates the tank
(b) Product made during the ramp-down squeezes the headspace over the vent setpoint
(c) The idle tank warms and its product breaks out into the headspace
(d) The make-up valve opens to hold the setpoint while the level falls

7.4 Theory and calculation · 6 min

Use Equation (7.1) with $H_C^\circ = 3.3 \times 10^{-4}\,\mathrm{mol/(m^3\,Pa)}$, $T_C^\circ = 298.15\,\mathrm{K}$, $\theta_C = 2400\,\mathrm{K}$ and $\nu_1 = 44.615\,\mathrm{mol/m^3}$. Since $V = \rho c/\nu_1$, the CO_2 pressure of a product is $p_C = V\nu_1/H_C(T)$.

(a) Find H_C and the CO_2 pressure of 3.80 vol at 3.0 °C and at 10.5 °C. *Checkpoint:* $H_C = 6.27 \times 10^{-4}$ and 4.98×10^{-4} in mol/(m^3 Pa), $p_C = 2.70$ bar and 3.40 bar.

(b) The product at the filler in case 1 holds 3.79 vol at 3.05 °C, its air exerts $p_A = 0.038$ bar and water vapor $p_w = 0.0076$ bar, and the headspace is at 3.0 bar(g), or 4.013 bar absolute. Find the effective gas pressure of Equation (7.11), the margin and the foam index. *Checkpoint:* 2.90 bar, 1.11 bar and 30.

(c) At 6.3 °C the same air gives $5p_A = 0.202$ bar and $p_w = 0.0095$ bar. Find the margin and foam index of 3.80 vol under 3.0 bar(g) and under 3.5 bar(g). *Checkpoint:* 0.80 bar and 50, and 1.30 bar and 19.

7.5 Theory and calculation · 5 min

The recipe makes 10.80 °Bx from 60.0 °Bx syrup by the mass ratio R_{SP} of Equation (7.14).

(a) Show that $R_{SP} = 10.8/49.2$, and find the Brix of the blend if a 58.0 °Bx batch arrives and the ratio is not changed. *Checkpoint:* 0.2195 and 10.44 °Bx.

(b) Find the ratio that makes 10.80 °Bx from 58.0 °Bx syrup, the trim ΔR it needs and the percentage of extra syrup. *Checkpoint:* 0.2288, $\Delta R = +0.0093$ and 4.2 %.

(c) Find the dose command of Equation (7.15) at 20 m^3/h for $V_{SP} = 3.80$ vol, first with $\Delta V = 0$ and then with the trim of case 1, $\Delta V = -0.054$ vol. *Checkpoint:* 149.2 kg/h and 147.1 kg/h.

7.6 Theory and calculation · 6 min

The dosing valve of Equation (7.8) passes $\dot{n}_{max} = K_c\sqrt{P_{sup}^2 - P_c^2}$ with $K_c = 0.178$ mol/(s bar), the pressures absolute in bar, $P_{atm} = 1.013$ bar and $M_{CO_2} = 44.01$ g/mol. The tank is at 3.0 bar(g), and $P_c = P_t + 1.2\,(Q/20)^2$ in bar with Q in m^3/h.

(a) At the design flow $P_c = 4.20$ bar(g). Find the valve capacity in kg/h with the supply at 9.0 bar(g) and at 5.0 bar(g), and the second as a share of the 149 kg/h the recipe needs. *Checkpoint:* 241 kg/h, 84.5 kg/h and 57 %.

(b) With the supply at 5.0 bar(g), test the feasibility condition of Equation (7.16) at 13 and at 14 m^3/h, and bracket Q_{feas}. *Checkpoint:* Feasible at 13 (111.8 against 101.8 kg/h), not at 14 (109.2 against 109.7 kg/h), so Q_{feas} just under 14.

(c) The supply then falls to 2.6 bar(g). What rate is feasible, and what does FY-100 do? *Checkpoint:* None. The filler is held.

7.7 App challenge · 9 min

Load scenario 2, Warmer feed, and use the default seed. Change nothing, run to 1200 s and export the CSV file. Then press the reload button, set Tank pressure setpoint to 3.5 bar(g) and Vent setpoint to 4.0 bar(g) before Run, run to 1200 s and export again.

(a) From the first run report the Filled in spec, Filled off spec and Worst foam index tiles, the first row where `foam_index` exceeds 50, and `T_tank_C` and `margin_bar` at 1200 s. *Checkpoint:* 2.29 and 4.37 m^3, 70, 413 s (within 1 s), 10.5 °C and 0.48 bar.

(b) Report the same tiles, `T_tank_C` and `margin_bar` for the second run. *Checkpoint:* 6.67 and 0.00 m^3, 40, 10.5 °C and 0.95 bar.

(c) The top pressure setpoint rose by 0.5 bar, but the margins at 1200 s differ by 0.48 bar. Read `P_true_barg` at 1200 s in both files and explain. *Checkpoint:* 3.02 bar(g) and 3.50 bar(g).

7.8 App challenge · 8 min

Load scenario 8, Analyzer bias, in which the dosing-balance check ships Off, and use the default seed. Run to 1200 s and export the CSV file. Then press Reset, set Dosing-balance check On on the AIC-105 card before Run, run to 1200 s and export again.

(a) From the first file read `V_meas_vol`, `V_true_an_vol`, `V_filler_vol` and `V_balance_vol` at 1200 s, the first row where `V_filler_vol` exceeds 3.95, and the Filled off spec tile. *Checkpoint:* About 3.80 (within 0.03, it is noisy), 3.99, 4.01 and 4.00 vol, 581 s and 3.44 m^3.

(b) From the second file find the first row with `balance_alarm_flag` equal to 1, the highest `V_filler_vol` and the Filled off spec tile. *Checkpoint:* 128 s, 3.84 vol and 0.00 m^3.

Explain why the first run fills product over the limit while every instrument reads normally, and what the check compares.

7.9 App challenge · 9 min

Load the challenge, a shift of filler demand, and use the default seed.

(a) Change nothing and run the full 2400 s. Report the Filled in spec, Filled off spec, CO_2 vented, Worst foam index and Highest level tiles, and name the bounds of the target that fail. *Checkpoint:* 10.98 and 0.82 m^3, 10.6 kg, 58 and 70.7 %. Four bounds fail.

(b) Press Reset and before Run set Filler-speed feedforward On, Stop handling On, Level loop gain 0.5 m^3/h per %, Level loop integral time 300 s, Tank pressure setpoint 3.5 bar(g) and Vent setpoint 4.0 bar(g). Run 2400 s and report the same tiles, the Cooling tile and the cooling per m^3 filled. *Checkpoint:* 11.80 and 0.00 m^3, 0.8 kg, 39, 50.6 %, 166 kWh and 14.1 kWh per m^3. Target met.

(c) Press Reset, set Tank pressure setpoint to 3.0 bar(g) and Vent setpoint to 3.3 bar(g), keep the other changes, and run again. *Checkpoint:* 8.64 and 3.16 m^3, 1.5 kg and 68.

Say what each change of (b) protects.

7.10 Application area · 3 min

Section 7.3 states that the model's CO_2 solubility is 6.7 % below a correlation of about 100 studies at 3 °C, and that the salting-out by sugar is not modeled.

(a) Does the model place the foam limit at a lower or a higher temperature than a product that follows the correlation would have? Estimate the difference for 3.80 vol under 3.0 bar(g), given that the correlation's solubility moves the limit from 6.3 to about 8.5 °C. *Checkpoint:* Lower, by about 2 °C.

(b) Which way does the missing salting-out act, and what should a plant engineer do before trusting the model's limit on a real line? *Checkpoint:* It lowers the solubility and partly cancels (a). Calibrate on the plant.

7.11 Application area · 4 min

Case 2 offers two ways to keep the product in specification through a warm water feed: the throughput cut, which fills 4.60 m^3 in 1200 s, and a top pressure of 3.5 bar(g), which fills all 6.67 m^3. The higher pressure costs about 2.4 kg of CO_2 once, held in the headspace.

(a) Over the 1200 s of the case, how much product does the cut give up, and what is the CO_2 cost of the pressure per m^3 of that product? *Checkpoint:* 2.07 m^3, about 1.2 kg per m^3.

(b) Figure 7.2 gives a margin of 0.71 bar at 12 °C under 3.5 bar(g). If a hotter feed took the tank to 12 °C, would the pressure alone keep the product in specification, and what would the engineer add? *Checkpoint:* No. Foam index 55. Add the cut, or more pressure within the tank's rating.

7.12 Review the agent · 5 min

An engineer asks an AI assistant how to stop the foaming of case 2. It replies: “The foam index passes 50 at 413 s because the margin is too small. Raise the PIC-201 setpoint from 3.0 to 3.5 bar(g). The margin rises by 0.5 bar, so the foam index falls by about 31 points and the whole run stays in specification. No other change is needed.” Find the flaw, and say how you would check the claim in the app.

CHAPTER 8

Timing the Quench to Hold mRNA Lipid Nanoparticle Size at Higher Throughput



The controller moves 1 the feed line into the closed mixing skid

The controller reads 1 the bench analyzer that measures particle size

What disturbs it 1 samples waiting their turn, the delay before a result

The mixing, the aging loop and the quench run out of sight inside the sealed cabinet, while the size that decides release is measured on a drawn sample at the bench minutes later. Illustration, not to scale.

IN PLAIN TERMS **Stopping Small Particles at the Right Size**

The main idea. Two liquids meet in a mixer, and tiny carrier particles form in the stream at once. Those particles keep growing until a flood of fresh buffer stops them, so the time between the mixer and the flood is what sets their size.

The challenge. The only instrument that reports the size draws a sample and answers minutes later. By then the liquid it sampled has left the plant, and anything made since was made blind.

Why it matters. Size is one of the qualities the product is judged on, so the plant has to know it while the liquid is flowing rather than after the tank is full.

THREE WAYS TO READ THIS CHAPTER

Explore	13 min	the first page, the history of the mixer as a flow device and the fact on parallel channels (Section 8.1), the chapter's own account of ethanol, jets and the arrest of fusion (Section 8.2), the two try-it segments and the fact from the simulator (Section 8.6) and the key points
Study	60 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 8.6), Technology in Practice (Section 8.7), Limitations (Section 8.8) and the two open challenges

The control question The particles of this skid grow for as long as the mixed stream ages before a quench arrests them, so a loop volume and a total flow act on one quality attribute. What does timing the quench buy that ratio control on four feeds cannot, and what can a blend of channels behind one set of meters hide?

Learning objectives

- Solve the four-feed ratio block for its flows, and show why two feeds cannot hold the charge ratio when the flow ratio moves.
- Compute the aging time and the predicted particle size from the loop volume, the total flow and the fitted size law.
- Explain why the ethanol content and not the quench ratio arrests growth, and compute the encapsulation at a given charge ratio.
- Compare release on a delayed size result with release on a soft sensor, and say why the soft sensor's bias gain stays small.
- Judge a scale-out blend that meets every window while its channels do not.

Prerequisites Sections 1.4.4 (ratio station and trim), 1.4.2 (unmeasured disturbances), 1.6.2 (velocity form) and 1.6.4 (scenarios and the challenge). Chapter 7 for a ratio blend trimmed on an analyzer.

The app https://apmonitor.com/apps/mrna-lnp/

See also Volume 5, Chapter 7 for residence-time tracking and diversion on a continuous drug line, Volume 5, Chapter 3 for a composition loop through a slow analyzer, and Volume 5, Chapter 4 and Volume 5, Chapter 10 for soft sensors that late measurements correct.

8.1 Introduction

The size of a lipid nanoparticle can be moved without touching its formulation. In the simulated skid of this chapter, an aqueous stream of mRNA in acetate buffer meets an equal flow of lipid in ethanol in an impinging-jet mixer at 40 mL/min per channel, and the mixed stream spends 117 ms in a 78 µL loop of tubing before a second jet floods it with two volumes of buffer. The particles leave at 139.9 nm with a modeled size spread $\sigma(\ln d)$ of 0.349 and 92.7 % nominal encapsulation. A 293 µL loop gives the same material 440 ms and takes it to 215.8 nm at the same flows, the same charge ratio and the same mixing, while raising the quench from two volumes to 3.5 volumes moves the size by 0.46 nm.

Nanoparticle production of this kind is often approached as one mixing step, which obscures the relation between the process inputs and the particle properties [1]. The simulator is built against the form of that belief a control student arrives with, that the flow ratio and the mixing intensity decide the particle size while the total flow decides only the throughput. In this model the jets make small particles whose size follows the aqueous buffer strength, and those particles then fuse for as long as the ethanol content lets them. The total flow is therefore the clock as well as the throughput, because the aging time is the loop volume divided by it, and the flow ratio acts mostly through the ethanol content. At a 3:1 flow ratio, with the charge ratio held at 6.01 by the ratio block, the same skid makes 98.0 nm particles instead of 139.9 nm.

The class of problem is a quality attribute fixed by the age of the material rather than by the state of a vessel. A continuous drug process must follow each portion of material through a residence-time distribution, and its control objective is the satisfaction of every critical quality attribute rather than a set of steady values [2]. The vocabulary for that is as old as the description of residence times in a continuous flow system [3], and every tubular hold, aging coil or maturation line shares the shape. Volume 5, Chapter 7 follows material in volume through such a distribution to a divert valve on a flow synthesis line. However, demonstrations of timed arrest are laboratory runs at fixed recipes [1], and the control literature on continuous manufacturing treats traceability and quality attributes in general terms [2], so neither supplies a plant on which the age distribution, each channel's true particle size and a delayed assay can be read side by side while a controller decides.

This chapter asks whether timing the quench buys anything that ordinary ratio control does not already buy. Every policy runs on one simulated formulation skid with the same instruments, feeds and faults, and each is judged on particle-quality variability and accepted productivity across flow perturbations and channel counts. The methods are all established ones: the size law, the tanks-in-series distribution, ratio control, a bias-corrected soft sensor and a constrained grid search. The results are measurements on a synthetic skid. Over six paired seeds, coordinated ratio control on the 78 µL loop runs out of buffer strength at 441.7 mL/min of formulated bulk, while a supervisor free to change the loop as well reaches 600.0 mL/min, a margin of 158 mL/min. In the passing run 13 500 mL of the 15 409 mL accepted was made while at least one channel sat outside its own window, although the blend that reached the tank met every criterion. The model behind these numbers closes its volume, RNA and lipid ledgers to 10^{-14} of throughput, its size map is checked against withheld tables, and a measurement boundary keeps every live decision on declared instruments.

Section 8.2 follows the formulation method from ethanol injection to the timed quench. Section 8.3 sets out the skid: the ratio block on four feeds, the volume clock and its age distribution, the size law, the collection line, the parameters and the boundary between what is measured and what is not. Section 8.4 records how the size law was calibrated and where its domain ends. Section 8.5 gives the flow loops, the size trim and its soft sensor, the three policies and the release rule. Section 8.6 runs the nine cases and the challenge, Section 8.7 sets the simulated instruments beside the real ones, Section 8.8 states what the model cannot support, and Section 8.9 collects the numbers. The exercises of Section 8.10 work through the ratio block, the size law, three of the cases and the challenge.

> ⧖ HISTORY **The Mixer Becomes a Flow Device, 1999 Onward**
>
> Particles of this kind were first made by hand, by pipetting or vortexing one phase into the other. In 1999 a T-shaped junction was introduced as an alternative to that macroscopic mixing for lipid-based particles, and the preparation became a flow process in a channel [4]. Industrial scale went a different way. A confined impinging-jet mixer collides two high-velocity streams in a chamber, and it relies on rapid mixing at Reynolds numbers between 100 and 2000, where inertia dominates and the flow is turbulent [4]. Throughput is then raised by running mixers in parallel rather than by building a larger one. One practice of that industry is the subject of this chapter: a dilution placed in line just downstream of the mixing zone floods the stream with aqueous buffer, which drops the ethanol content to 20 percent by volume or less and stabilizes the particle size [4].

8.2 Ethanol, Jets and the Arrest of Fusion

Lipid vesicles entered the laboratory in 1965, when phospholipids swollen in water were shown to form lamellar structures across which univalent ions diffuse slowly [5]. Ethanol was the solvent that made them a process rather than a preparation. Injecting an ethanolic lipid solution into an aqueous phase produced single-bilayer vesicles without sonication in 1973 [6], and the step from that injection to a manufacturing route was taken when plasmid DNA was encapsulated by combining a lipid-ethanol stream with an aqueous DNA stream in a T-shaped chamber, a stepwise dilution that dropped the ethanol below the concentration at which the lipid stays soluble and gave particles under 200 nm with encapsulation above 80 % [7]. The chemistry that made the route useful arrived separately. Ionizable cationic lipids, designed against a proposed mechanism of action, carried small interfering RNA at doses as low as 0.01 mg/kg in rodents [8], and the same particle class later carried mRNA into clinical use [9, 10].

Mixing time became the design variable once the vesicles were made in flowing streams. A confined impinging-jet mixer was characterized through a competitive reaction set as a chemical ruler, over characteristic times from 320 ms down to 5 ms [11]. Microfluidic mixing in milliseconds then produced limit-size particles, the smallest aggregates compatible with the packing of their constituents, with diameters engineered from 20 to 80 nm by composition [12], and the same technique gave siRNA particles of 20 nm and larger with essentially complete encapsulation and polydispersity indexes as

low as 0.02 [13]. Cryogenic microscopy and X-ray evidence then described how the particles form, with ionizable lipid not complexed to cargo promoting small vesicular structures that coalesce into larger particles with dense cores [14]. Coalescence is the mechanism the present model exploits, because a process that can start it can also stop it.

> **+ DID YOU KNOW? A Hundred Channels at Once**
>
> Scale-out is real practice. An array of 128 mixing channels operating simultaneously gave more than 100 times the production rate of a single channel, with no loss of the particles' physical properties or of their biological potency [4].

The step this simulator is built on is the deliberate separation of the two stages. Mixing at a symmetric flow ratio, with the higher ethanol content that implies, holds the particles in a fusion-prone condition, and an aqueous buffer injected after a defined residence time arrests them at the size they have reached, with the size and the morphology controlled independently and without giving up the other quality attributes [1]. That source is the anchor of the application. Its reported mean sizes, tabulated over four residence times and five buffer strengths for several cargoes, are the data the simulator's size law was fitted to, and its qualitative statements about ethanol, about PEG-lipid saturation as what finally arrests fusion, and about a conductivity probe as a cheap handle on the buffer strength are what the application's three declared extension factors stand on. The flowsheet modeled here was assembled for teaching, out of evidence that is public. It is not a reproduction of any commercial process, no product or company is named, and the quality limits in the application are teaching values rather than a specification. Real limits for a particle product come from the product's own specification and from regulatory guidance, which for liposome drug products asks for particle size and distribution profile among the characteristics of the drug product specification [15].

8.3 The Skid as Modeled

The skid is drawn in Figure 8.1. Four pumps feed the first impinging-jet mixer IJM-1 of each channel: P-101 carries the mRNA stock at concentration c_P of counted phosphate groups in buffer S_1, P-102 a buffer concentrate S_2, P-103 water for injection, and P-104 the lipid solution in ethanol at c_N of counted ionizable nitrogen. Each has its own flow meter and its own flow loop, and a conductivity probe CT-101 reads the aqueous blend. The mixed stream crosses one of five aging loops between 3.3 µL and 667 µL, reaches IJM-2 where pump P-105 injects R_q volumes of quench buffer, and flows through a collection line of four well-mixed cells to the divert valve XV-401, the hold tank and a buffer exchange that removes the ethanol and hands over the formulated bulk. One, two or four identical channels run off the same manifold.

Table 8.1 sorts the variables by what they do. Four operator demands act through one ratio block, four flow loops and the aging loop in service, and three quality attributes and a rate are what the skid is judged on. Units are SI inside the model, which displays mL/min, µL, nanometers, mol/m^3 and bar.

8.3.1 Four Demands on Four Feeds

The operator asks for a total flow per channel, a flow ratio, a charge ratio and an aqueous buffer strength. The ratio block FY-100 solves the four feed flows that realize all four at once, as shown in Equation (8.1). It is the ratio station of Section 1.4.4 with three ratios held instead of one, and each of its outputs is the setpoint of a flow loop.

$$
\begin{aligned}
Q_o &= \frac{\mathrm{TFR}}{1+\mathrm{FRR}}, \qquad Q_a = \mathrm{TFR} - Q_o = Q_1 + Q_2 + Q_3 \\
Q_1 &= \frac{Q_o\, c_N}{(N/P)\, c_P}, \qquad Q_2 = \frac{S\,Q_a - Q_1 S_1}{S_2} \\
Q_3 &= Q_a - Q_1 - Q_2, \qquad Q_q = R_q\,\mathrm{TFR}
\end{aligned}
\tag{8.1}
$$

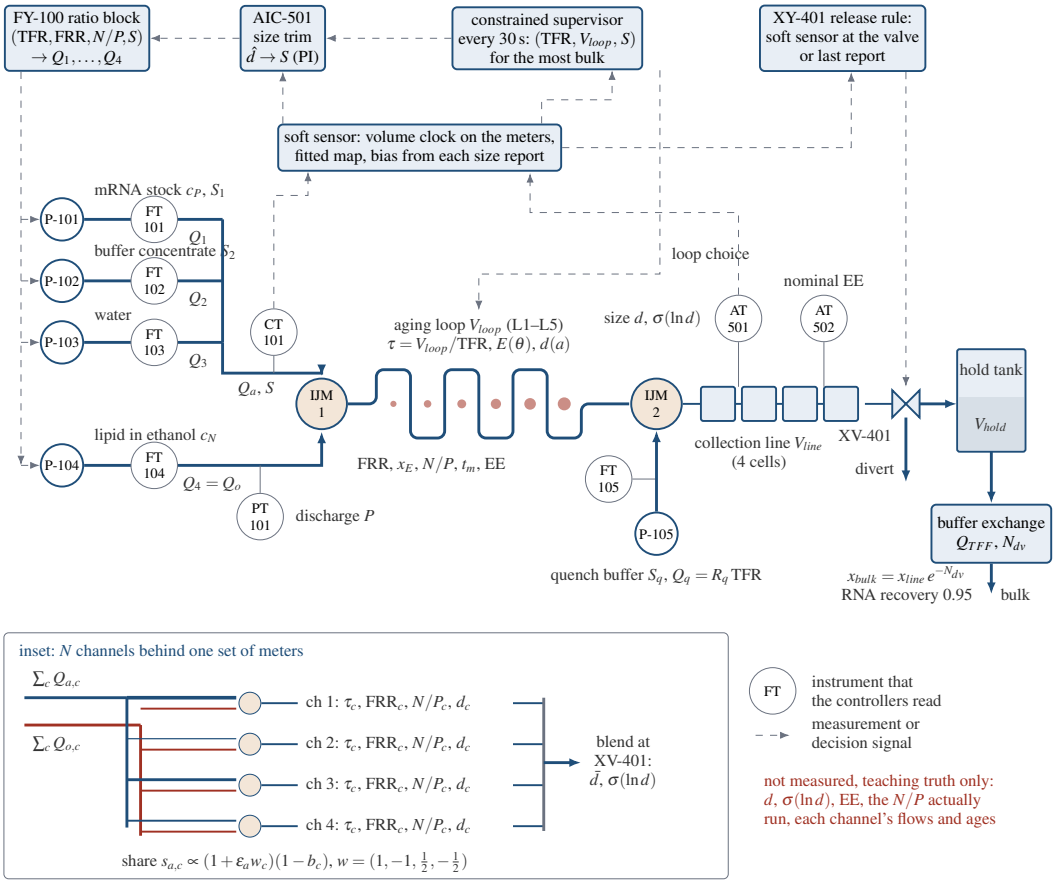


Figure 8.1 One formulation channel of the simulated skid, with the manifold of the parallel channels as an inset. Four metered feeds set the total flow TFR, the flow ratio FRR, the charge ratio N/P and the aqueous buffer strength S at IJM-1, which also fixes the ethanol fraction x_E, the mixing time t_m and the nominal encapsulation EE. The aging loop V_{loop} is the only place the particles grow, over the age distribution $E(\theta)$ of its volume clock. IJM-2 arrests them with Q_q, the collection line carries the eight transported quantities to XV-401, and the buffer exchange takes the residual ethanol out. Instruments the controllers read are circled, signals are dashed, and the legend lists the states of the model that no instrument on the skid reports. In the inset the organic header is the lower line of each pair. Process model only, not a piping diagram. Not to scale.

where c_N and c_P are molar concentrations of counted ionizable nitrogen and counted RNA phosphate rather than of lipid and RNA molecules. With an mRNA stock and a lipid solution alone the first and third relations share Q_1, so tripling the flow ratio at fixed feed compositions divides the charge ratio by three. Measured on the application with two feeds, a 1:1 to 3:1 step takes N/P from 3.00 to 1.00, while the four-feed block holds 6.000 to 6.000. The buffer concentrate and the water are therefore the two extra degrees of freedom. The block keeps Q_2 between zero and $Q_a - Q_1$ and Q_3 at or above zero, and it reports when either limit binds: at FRR $= 0.50$ the mRNA stock alone fills the aqueous stream and the realized strength is 26 mol/m^3 of the 100 mol/m^3 asked. The quench dilutes both components in the same ratio, so it cannot change N/P.

At the design point the declared feed concentrations give $c_N = 5\,\mathrm{g/L} \times 0.50/0.648\,\mathrm{kg/mol} = 3.858\,\mathrm{mol/m^3}$ and $c_P = 0.425\,\mathrm{g/L}/0.330\,\mathrm{kg/mol} = 1.288\,\mathrm{mol/m^3}$, and Equation (8.1) returns 9.986, 3.501, 6.514 and 20.000 mL/min with 80 mL/min of quench.

Table 8.1 Variable Classification and the Design-Point Values

Variable	Role	Design	Range
d	CV, particle size (true)	139.9 nm	target ±15 nm
$\sigma(\ln d)$	CV, modeled size spread	0.349	≤ 0.45
EE	CV, nominal encapsulation	92.7 %	≥80 %
S	MV of the size trim	100 mol/m^3	25–300
V_{loop}	MV of the supervisor	78 µL	3.3–667
TFR	MV of the supervisor	40 mL/min	10–80
$Q_1 \dots Q_4$	MV, feed flows	9.99/3.50/6.51/20.0	0–160 mL/min
Q_q	MV, quench flow	80 mL/min	0–1200 mL/min
$u_1 \dots u_5$	MV, pump speeds	25 %	0–100 %
XV-401	MV, accept or divert	accept	—
FRR	demand, flow ratio	1.0	0.5–5
N/P	demand, charge ratio	6.0	3–10
R_q	demand, quench ratio	2.0	0.5–4
ε_a, ε_o	DV, maldistribution	0	±30 %
b_c	DV, channel restriction	0	0–80 %
δ_L	DV, lipid lot and drift	0	±30 %
δ_S	DV, concentrate lot	0	±30 %
A, f	DV, pump pulsation	0	0–60 %, 0–5 Hz
ϕ	state, lipid volume fraction	2.49×10^{-3}	—
W, V_2	states, log-size moments	—	—
a	state, material age	6.1 s at the valve	—
c_R, e_R	states, total and held RNA	—	—
x_E	state, ethanol fraction	16.6 % in line	—
salt	state, buffer in the stream	—	—
Q_i	states, pump deliveries	as above	—
P	output, discharge pressure	2.51 bar	alarm 6 bar

🕮 **Pause.** Try Exercise 8.2 before reading on.

8.3.2 The Volume Clock and Its Age Distribution

The aging loop is between 5 ms and 1000 ms long while the physics step is 50 ms, so the loop is not integrated. It is a transport clock on cumulative volume with a normalized residence-time distribution, as shown in Equation (8.2).

$$\begin{aligned} \Theta(t) &= \int_0^t \mathrm{TFR}(t')\,\mathrm{d}t', \qquad \theta = \frac{\Theta(t) - \Theta(t-a)}{V_{loop}} \\ E(\theta) &= \frac{n^n \theta^{n-1} \mathrm{e}^{-n\theta}}{(n-1)!}, \qquad \int_0^\infty E(\theta)\,\mathrm{d}\theta = 1 \end{aligned} \tag{8.2}$$

where n is the number of equal stirred volumes the loop is modeled as, a teaching value of 24 whose distribution has $\sigma/\tau = 0.204$. A packet's age a is found by inverting the cumulative volume, never by taking the instantaneous V_{loop}/TFR. The application evaluates E at nine equal-probability nodes, whose dimensionless throughputs run from 0.699 to 1.347 and whose mean is renormalized to exactly one, so the material reaching IJM-2 is a volume average over a distribution of ages. At a constant flow the clock is exact and $\tau = V_{loop}/\mathrm{TFR}$, which is 117.0 ms for the 78 µL loop at 40 mL/min.

🕮 **Pause.** Try Exercise 8.1 before reading on.

8.3.3 The Size Law

Each node carries the fitted map of Equation (8.3), which is the model form of the anchor: an initial self-assembly size that follows a modified Gompertz curve in the aqueous buffer strength, plus a fusion-driven growth term that is a power law in the age.

$$\begin{aligned} d(a) &= d_0(S) + R(S)\,a^p, \qquad u(S) = \mathrm{e}^{-\mathrm{e}^{-D(S-C)}} \\ d_0(S) &= a_0 + a_1 u(S), \qquad R(S) = r_0 + r_1 u(S) \end{aligned} \tag{8.3}$$

where a is in milliseconds and d in nanometers. The seven coefficients were fitted inside the application, as Section 8.4 records, because the anchor's own coefficients are in supplementary material that could not be read.

Growth is fusion, so it conserves particle volume. The model transports the lipid volume fraction ϕ with no source term anywhere and derives the particle number from it and the diameter, as shown in Equation (8.4).

$$\begin{aligned} \frac{\mathrm{d}d}{\mathrm{d}a} &= p\,R(S)\,a^{p-1} f_E(x_E)\, f_d(\phi)\, f_P(d) \\ f_E &= \left(\frac{x_E}{0.5}\right)^{6}, \quad f_d = \frac{\phi}{\phi_{ref}}, \quad f_P = \min\!\left(1, \frac{900-d}{500}\right)^{+} \\ N &= \frac{6\phi}{\pi d^3}, \qquad \left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{\text{growth}} = 0 \end{aligned} \tag{8.4}$$

where x_E is the ethanol volume fraction and ϕ_{ref} the lipid volume fraction of the design mixed stream. The three factors are declared teaching closures, each resting on one qualitative statement of the anchor rather than on data: ethanol raises membrane fluidity and lowers the barrier to fusion, fusion is a collision process whose rate falls with dilution, and PEG-lipid accumulating on the shrinking surface eventually arrests it. Their measured weights differ by three orders of magnitude. Two volumes of quench multiply f_E by 0.00137 and f_d by 0.333, so the arrest is the ethanol's, and f_P is exactly one below 400 nm and is never engaged by any shipped case. The quadrature of the growth integral runs in $u = a^p$ rather than in a, because the integrand carries an a^{p-1} singularity at zero. In that variable the integrand is bounded and the plant reproduces Equation (8.3) to round-off whenever the ethanol content is constant along a packet's path.

8.3.4 What the Collection Line Carries

Everything downstream of the quench junction is a state. Each channel's collection line is four well-mixed cells carrying eight quantities that mix linearly, as shown in Equation (8.5).

$$\begin{aligned} \frac{\mathrm{d}c_k}{\mathrm{d}t} &= \frac{Q_{line}}{V_{cell}}\,(c_{k-1} - c_k) + s_k, \qquad V_{cell} = \frac{V_{line}}{4} \\ d &= d_{ref}\,\mathrm{e}^{W/\phi}, \qquad \sigma(\ln d)^2 = \frac{V_2}{\phi} - \left(\frac{W}{\phi}\right)^2 \end{aligned} \tag{8.5}$$

where the eight are ϕ, $W = \phi \ln(d/d_{ref})$, $V_2 = \phi[(\ln d/d_{ref})^2 + v_0]$, the age a with source term one, total and encapsulated RNA c_R and e_R, the ethanol fraction x_E and the buffer in the whole stream. The size distribution is modeled as lognormal in d, so two streams mix as two lognormals do and the reported size is the geometric mean diameter. The growth source terms are $s_W = \phi g/d$ and $s_{V_2} = 2\phi \ln(d/d_{ref})\, g/d$ with g from Equation (8.4), which is why fusion moves W and V_2 and never ϕ. The quantity $\sigma(\ln d)$ is the modeled width of that lognormal. It is not the cumulants polydispersity index a light-scattering instrument reports [16], the transform between the two is not modeled, and the application never prints a polydispersity index for its own output. The distinction is the one a measurement study of dynamic light scattering makes when it warns that central values taken incautiously from a scattering distribution lead to interpretation errors [17].

8.3.5 Encapsulation, Pressure and the Exchange

Nominal encapsulation is set at the mixer and then carried along, as shown in Equation (8.6).

$$\begin{aligned} \mathrm{EE} &= \mathrm{EE}_{max}\,\frac{1}{1+(t_m/t_{m,1/2})^2}\cdot\frac{1}{1+(2.6\,P/N)^4} \\ t_m &= t_{m,0}\left(\frac{Q_0}{\mathrm{TFR}}\right)^{3/2}, \quad P = 10^5 + k_{res}(1+3b_c)Q_{ch}^{1.8} \\ \frac{x_{bulk}}{x_{line}} &= \mathrm{e}^{-N_{dv}}, \qquad N_{dv} = \ln\!\left(\frac{x_{line}}{x_{max}}\right) + m_{dv} \end{aligned} \tag{8.6}$$

where t_m is the jet mixing time, scaled from its design value by engulfment so that it falls as $\mathrm{TFR}^{-3/2}$, Q_{ch} is the channel's whole flow including the quench, b_c a channel restriction and x_{max} the teaching limit on residual ethanol. The mixing term is fitted to the single degraded operating point the anchor quotes, 60 % encapsulation at 0.4 mL/min against 82 % at 8 mL/min, with the exponent declared rather than fitted because one end point cannot identify both. The charge term collapses below about four nitrogens per phosphate. This is nominal encapsulation, the fraction an exclusion assay cannot reach, not intact functional RNA, and no statement about potency or transfection is made or implied anywhere in the application. The diafiltration relation is exact for a constant-volume exchange, and at the design point 16.62 % ethanol in the line needs 3.50 diavolumes to reach the 0.5 % teaching limit, against the 3.81 the controller runs with its margin.

The manifold splits each feed among the channels in shares that always add up to one, so it can neither create nor lose flow. The aqueous share of channel c is proportional to $(1+\varepsilon_a w_c)(1-b_c)$ with the pattern $w=(1,-1,\frac{1}{2},-\frac{1}{2})$, and the organic share follows ε_o in the same way. The total flow meters see nothing of either.

8.3.6 Parameters, Evidence, Initialization and Numerics

Table 8.2 gives every parameter with its value and the label its evidence earns. Sizes, delays and limits that no source fixes are teaching values chosen for the lesson, and they are labeled as such in the application's Model dialog and on its Instructions page as well.

Every case opens lined out on its own settings. Equation (8.1) is solved open loop, the loop and the line are filled with material of the age that flow would give it, the meters and the analyzers open on that state, and the soft sensor's history is primed back over twice the longest loop. Only a start-up case opens with its pumps stopped. States advance by fourth-order Runge-Kutta at a fixed 0.05 s, a step set by the pump drives ($\tau = 0.8$ s) and one collection-line cell (3 mL at 120 mL/min, 1.5 s), and the controllers execute every 0.2 s. Halving the step moves the particle size of the quench case by 0.00 nm, of the pulsation case by 0.05 nm and of the challenge by 0.18 nm, and the accepted bulk by less than 10^{-4} relative. The challenge's 1500 s needs 1.8 s of computing headless and 37.5 s of wall time in a browser at the top of the speed ladder of Section 1.6.1, where the full 40-fold rate is realized.

8.3.7 Measurements and the Truth Boundary

The controllers, the soft sensor, the supervisor and the release rule read one packet and nothing else: five flow meters with a 0.5 s lag and 0.4 % noise, their manifold total, the conductivity probe with 1.5 mol/m^3 of noise and whatever drift a case applies, the discharge pressure transmitter, and two sampled analyzers. The size analyzer draws a sample every 90 s and reports 60 s later, so a result is between 60 and 150 s old, and the encapsulation analyzer reports 420 s after its own sample. Noise is drawn once per scheduled measurement, never inside the derivative function and never while drawing. The true particle size, the true spread, the true encapsulation, the age distribution, the charge ratio actually run and every per-channel flow are states of the model with no instrument behind them. They appear in the scene as dashed teaching truth, in the scorecard, and in the exported file under names prefixed 'truth_', and one of the application's own checks fails when a controller touches any of them.

8.3.8 Conservation and Numerical Checks

The application's own test suite prints the residual of every balance rather than clipping it. Over the disturbed four-channel case the volume, RNA and lipid ledgers close to 3.9×10^{-15}, 7.4×10^{-15} and 2.0×10^{-15} of what was fed. Because the aging loop is evaluated rather than integrated, its own hold-up is carried as two explicit inventory states per channel. Without them the RNA and lipid balances closed only to 2.4×10^{-4} during a transient, which was the loop's hold-up and not a solver defect. The residence-time distribution integrates to one at every node count tested, growth has no source term in ϕ, and the priming execution feeds nothing, advances no clock, draws no sample and updates no bias.

Table 8.2 Parameters and the Evidence Each One Carries

Symbol	Meaning	Value	Source
a_0, a_1	assembly size, offset and span, nm	61.84, 65.98	fitted
r_0, r_1	growth, the same, nm/ms^{p}	0.154, 1.653	fitted
C, D	Gompertz center mol/m^3, slope	81.03, 0.0256	fitted
p	growth exponent in the age	0.7785	fitted
x_{fit}	ethanol fraction the map was fitted at	0.50	source
n_E	ethanol exponent in f_E	6	teaching
ϕ_{ref}	design lipid volume fraction	2.50×10^{-3}	derived
d_{knee}, d_{stop}	PEG guard, nm	400, 900	source, teaching
n, M	tanks in series, quadrature nodes	24, 9	teaching
v_0	spread one aging leaves, $(\ln d)^2$	0.12	teaching
EE_{max}	encapsulation plateau	0.96	teaching
$t_{m,0}$	jet mixing time at 40 mL/min, ms	0.30	teaching
$t_{m,1/2}$, n_{mix}	mixing half point ms, exponent	406, 2	fitted, declared
c_{RNA}	mRNA stock, g/L	0.425	derived
c_L	lipid solution, g/L	5.0	source
S_1, S_2, S_q	stock, concentrate, quench mol/m^3	25, 500, 100	source, teaching
V_{loop}	five loops, µL	3.3–667	source times
V_{line}	collection line per channel, mL	12	teaching
Q_{max}, $Q_{max,q}$	pump capacities, mL/min	160, 1200	teaching
τ_{pump}, τ_{FT}	drive lag, meter lag, s	0.8, 0.5	teaching
k_{res}	channel resistance (SI)	2.7×10^{15}	teaching
P_{alarm}, P_{trip}	pressure alarm and trip, bar	6, 8	teaching
V_{hold}	hold tank, L	2.0	teaching
Q_{TFF}	exchange throughput, mL/min	700	teaching
rec_{TFF}	exchange RNA recovery	0.95	teaching
t_{SR}, t_{cyc}	size result, sampling, s	60, 90	source, teaching
t_{EE}	encapsulation result, s	420	teaching
σ_{FT}, σ_{cond}	meter, probe noise	0.4 %, 1.5 mol/m^3	teaching
σ_{size}	size assay noise	2.5 %	teaching
Δd	size window, nm	15	teaching
σ_{max}	modeled spread limit	0.45	teaching
EE_{min}	encapsulation minimum	80 %	source standard
x_{max}	residual ethanol, % v/v	0.5	teaching
N/P window	charge ratio, low and high	5.0, 7.0	teaching

Table 8.3 Verification of the Implementation

Check	Result
Volume ledger, fed minus line change minus out	3.9×10^{-15}
RNA ledger, the same	7.4×10^{-15}
Lipid ledger, the same	2.0×10^{-15}
Lipid volume fraction during growth, $\lvert d\phi/dt \rvert$	0 exactly
Residence-time distribution, mean throughput	1.000000000000
Charge ratio across a 1:1 to 3:1 step, four feeds	$< 10^{-9}$
Nine nodes against 2001, mean size at 440 ms	0.75 per mille
Step halved, quench case size	0.00 nm
Step halved, pulsation case size	0.05 nm
Step halved, challenge size | bulk	0.18 nm | $< 10^{-4}$
Playback speed, one run against thirty chunks	bit-identical
Each case loaded three ways	identical
Priming step: volume fed, samples, bias updates	0, 0, 0
Wall time, 1500 s challenge, headless	1.8 s
Wall time, the same in a browser at 40×	37.5 s

8.3.9 Assumptions

The size map is fitted to one published table: one formulation family of four lipids, one cargo type, pH 5.5, one mixer geometry, buffer strength 25 to 300 mol/m^3 and aging 5 to 440 ms. The application draws that box and says when the operating point has left it. Fixed pH is a real restriction, because the anchor shows pH to be an independent handle on size and morphology, and its own pH 4.5 table is the

one this map does not reproduce. Morphology is not modeled at all, and no bleb fraction, transfection or potency is predicted. Encapsulation is nominal and is set at the mixer rather than evolving. The loop's age distribution is that of 24 stirred volumes in series, whereas a real coil carries a parabolic velocity profile and secondary flow, whose dispersion is a different and usually wider shape around the same mean [18]. The buffer exchange is a declared recovery map with an exact diafiltration law, not a membrane model, and the retention of a 100 nm particle by a real device is outside it. Every parallel channel uses the same per-channel model with its own actual flows, so more throughput across more channels is not evidence that a larger channel would behave the same way. All quality limits are teaching values for this simulation, and a simulated pass is not product release.

8.4 Calibration and Its Domain

The anchor publishes the model form of Equation (8.3) and its own five coefficients in supplementary material whose endpoint returned HTTP 403, so the coefficients could not be read. The seven coefficients of Table 8.2 were therefore fitted inside the application, by least squares on $\ln d$, to the twenty mean particle sizes the anchor's Figure 2b5 tabulates for one cargo at pH 5.5 over four residence times and five buffer strengths [1]. The fit reproduces that table to 5.11 % rms with a worst cell of 10.7 %. A second derivation by damped Gauss-Newton from 200 random starts, written independently of the first, lands on the same coefficients to six significant figures and the same sum of squares, 5.2764×10^{-2}.

Agreement with the source is a fit inside its domain plus two withheld checks, not a validation. The two cargo tables that were not fitted, a second mRNA and a yeast RNA at the same pH, are reproduced at 8.76 % and 7.96 % rms with no refitting, as shown in Figure 8.2. The fourth table, taken without cargo at pH 4.5, is not reproduced at all (28.15 % rms, worst cell 79.7 %), which is why pH 5.5 is a declared boundary of the domain rather than an untested extrapolation.

Two further checks were run for this chapter and are not in the application. The first withholds one buffer row at a time and refits: the interior rows are predicted at 4.8 to 6.4 % rms, while the 25 mol/m^3 and 300 mol/m^3 edges are predicted at 20.5 and 16.5 %, because each edge is held by its own row and is not interpolated. The second withholds one residence-time column at a time. The 50, 117 and 440 ms columns are predicted at 9.6, 10.9 and 90.6 % rms, and the 5 ms column at 34.8 %. The power law cannot be extrapolated to the longest time from the shorter ones, and 90.6 % is the quantitative case for the window guard that stops the supervisor from leaving the box. One independent check of the reading itself: the anchor's Figure 4 quotes five case studies at pH 5.5 by their conditions and sizes as 70.7, 67.6, 113, 100 and 78 nm, and the corresponding table cells read 70.7, 67.7, 112.8, 100.7 and 78.6 nm.

The three extension factors of Equation (8.4) carry the map to conditions the table does not cover, and their weights were measured by changing each one for the plant, the soft sensor and the supervisor together and rerunning every case. The dilution factor carries almost nothing. Removing it entirely moves the quench case's 215.8 nm by 0.4 nm and tripling it by 0.7 nm, and no verdict changes. The PEG guard is inert in every shipped case, since none reaches 400 nm even with the window guard off. The ethanol exponent carries the rest. At $n_E = 3$ and 12 the direction of every lesson survives and the numbers do not: the 3:1 size moves from 98.0 to 103.3 and 97.4 nm, the separation between the richest and leanest channel of the four-channel case from 36.5 to 29.6 and 72.8 nm, and at $n_E = 12$ the challenge's suggested settings no longer pass, making 358.6 mL/min with 7.2 L of off-specification material accepted. Every quantitative statement this chapter makes about a flow-ratio effect rests on that one teaching exponent.

8.5 Implemented Control System

Every loop executes at $\Delta t_c = 0.2$ s, which is also the instrument scan. Within one physics step the order is the controllers, then the integrator, then the plant's own bookkeeping, and last the outputs and the historian.

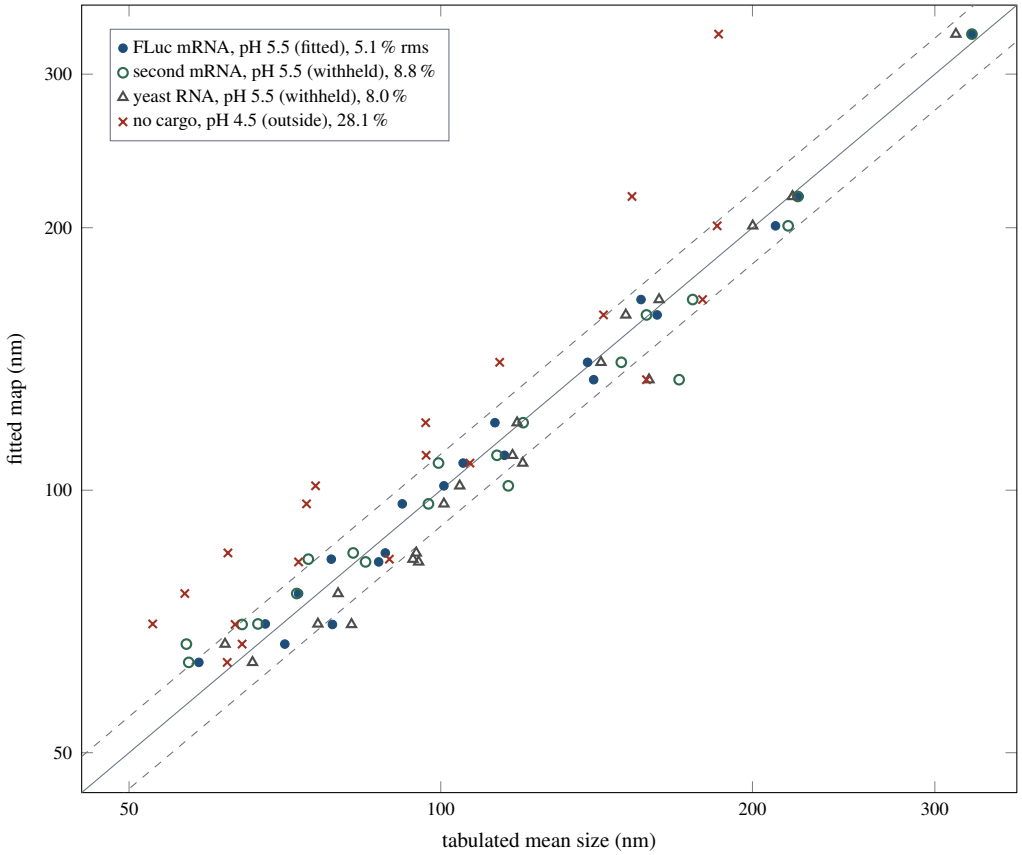


Figure 8.2 The fitted map against the tabulated mean sizes it was fitted to and the three tables withheld from the fit: filled circles for the fitted table, open circles and triangles for the two withheld cargo tables, and crosses for the table without cargo. Dashed lines are $\pm 10\,\%$. The two withheld cargo tables at pH 5.5 are reproduced at 8.8 and 8.0 % rms with no refitting, while the pH 4.5 table lies off the diagonal at 28.1 %, which fixes pH 5.5 as a boundary of the declared domain.

8.5.1 The Flow Loops

Each of the five pumps runs on the velocity-form PI algorithm of Section 1.6.2, with the error taken on the metered flow in mL/min, the output in percent of pump speed clamped to 0 and 100, back-calculation into the integral term over a tracking time of τ_I, and bumpless transfer from MANUAL. The four feed loops share $K_c = 1.2\,\%$ per mL/min and $\tau_I = 3.0\,\mathrm{s}$, and the quench loop runs at 0.25 and 4.0 s, with no derivative action anywhere. The ratio block's demand passes through a rate limit on the total flow, a declared 120 mL/min per minute, which is what makes the start-up case a lesson rather than a five-second event. A pump asked for more than its capacity reports the request beside what was applied.

8.5.2 The Size Trim and the Soft Sensor

AIC-501 is a slow PI on the same algorithm that moves the aqueous buffer strength to hold the size, with $K_c = 1.2\,\mathrm{mol/m^3}$ per nanometer, $\tau_I = 90\,\mathrm{s}$ and output limits of 25 and 300 $\mathrm{mol/m^3}$. That knob was chosen because the ratio block can move it without touching the flow ratio or the charge ratio, which is the whole reason for the two extra feed pumps. It reads either the last reported size or the

soft sensor. The same shape, a ratio blend trimmed on a composition analyzer, holds the sweetness of a soft drink in Chapter 7.

The soft sensor rebuilds, from the meters and the probe alone, the ratios on the label concentrations, its own volume clock with the same normalized distribution the plant uses, the fitted map at each of the nine ages, and the encapsulation closure. Each size result then updates a bias on the logarithm of the diameter, as shown in Equation (8.7).

$$b \leftarrow b + \lambda \left[\ln d_{lab} - (\ln \hat{d}_s + b) \right], \qquad \hat{d} = \hat{d}_{raw}\, \mathrm{e}^{b} \tag{8.7}$$

where $\hat{d}_s$ is the sensor's own prediction for the sample that was drawn, recovered from a ring of past predictions, rather than its prediction now. The same rule corrects the spread and the encapsulation on their own channels, bounded at 0.5, 0.4 and 0.15. The gain λ is 0.15 per result and was chosen by measurement, not by data: the result is 60 s old and arrives every 90 s while the trim acts in one scan, so the loop is delay-dominated, and at $\lambda = 0.4$ the challenge's accepted off-specification volume rises from a mean of 43.1 mL to 197.4 mL over the same six paired seeds at the same bulk rate. A soft sensor whose bias a late laboratory corrects also carries the moisture loop of Volume 5, Chapter 10, and Volume 5, Chapter 4 reconciles a residence-time predictor with an analyzer that reports hours after the action.

🕮 **Pause.** Try Exercise 8.3 before reading on.

8.5.3 Three Policies and a Release Rule

The policies read the same packet and the same hardware. FIXED RATIOS sends the operator's numbers to FY-100 and corrects nothing. COORDINATED FLOW AND QUENCH adds AIC-501 and automatic diavolumes. CONSTRAINED SUPERVISOR adds the choice of operating point: every 30 s it evaluates nine total flows from 20 to 80 mL/min crossed with the five aging loops, bisects the buffer strength that puts the predicted size on target, and takes the point of greatest bulk rate subject to the predicted size, the modeled spread, the encapsulation, the pressure alarm, the exchange throughput, and, with the window guard on, the box the map was fitted in, as shown in Equation (8.8).

$$\begin{aligned} \max_{\mathrm{TFR},\, V_{loop},\, S} \quad & \min\left(Q_{line}, Q_{TFF}\right) rec_{TFF} \\ \text{s.t.} \quad & |\hat{d} - d_{SP}| \le m_s \Delta d \\ & \hat{\sigma} \le \sigma_{max}, \quad \hat{\mathrm{EE}} \ge \mathrm{EE}_{min} \\ & P(\mathrm{TFR}) \le P_{alarm}, \quad Q_{line} \le Q_{TFF} \\ & 5\,\mathrm{ms} \le \tau \le 440\,\mathrm{ms} \\ & 25 \le S \le 300\,\mathrm{mol/m^3} \end{aligned} \tag{8.8}$$

where $m_s = 0.6$ is a margin on the size window and $\hat{\sigma}$ adds the variance of $\ln d$ over the age nodes to the within-node spread. Keeping the present loop is worth 8 % more bulk and no loop change may follow another within 150 s, because moving the loop puts a transient through the line. When nothing is feasible the search takes the point that misses the size by least from inside the box and names *size* as the binding constraint. It is a constrained steady-state search on a cadence, not a receding-horizon optimization, and nothing in the application is called model predictive control.

The release rule at XV-401 is a separate decision with three settings: accept everything, accept while the last reported size is inside the window, or accept while the soft sensor says the material now at the valve is inside it. The third asks the sensor what it predicted one line volume ago, computed from the meters, because the sensor predicts the quench junction while the valve is 12 mL per channel downstream. Each rule keeps 20 % of every window in hand by default, diverts while the pressure trip is latched, and holds the valve shut until the line has turned over twice since the feeds started.

Table 8.4 The Nine Teaching Cases and the Challenge

#	Case and events	Teaching point
1	Flow ratio 1:1 to 3:1 at 240 s, charge ratio held	ethanol, not mixing, sets the window
2	Total flow to 20 at 200 and 72 mL/min at 500 s	the total flow is the clock
3	Lipid lot 22 % under its label at 150 s	a feed nothing meters
4	Aging loop 78 to 293 µL at 240 s, then quench 3.5 volumes	the opening experiment
5	P-104 head pulsates ±40 % at 180 s	a spread the meter cannot see
6	Release on the reported size, flow to 21 mL/min at 200 s	a result a minute old
7	Four channels, manifold 12 % out at 180, channel 2 restricted at 480 s	one meter, four channels
8	Start-up from stopped pumps	material with another history
9	Supervisor asked for 380 nm	a target outside the window
10	Challenge: four channels, four faults, 1500 s	the whole problem

8.5.4 Plant, Controller and Internal Models

The application keeps three models apart. The plant is everything in Section 8.3. The controller's model is the fitted map on label concentrations with its learned bias, which does not know a lipid lot that is not its label, a concentrate that is weak, a maldistribution or a restriction. The identified model is the relation a learner fits by hand between the aging time and the size, which is what cases 2 and 4 are for. The supervisor evaluates the controller's model only, including for the discharge pressure, so a hidden restriction never reaches it.

8.6 Nine Cases and a Challenge, Measured

The cases appear in Table 8.4. One reference configuration underlies them, the design point of Table 8.1 run on two channels under fixed ratios with the soft-sensor release rule, and each row names only what it changes. Measurement was done on the engine alone, at its default seed, on the step the browser uses, after the last behavior change. Loading a case first, picking it from the dropdown after another one and restarting it give the same trajectory to the last bit. Each case credits one decision, an action taken and a state the skid reached because of it, as Section 1.6.4 describes.

8.6.1 What Sets the Size

Case 4 is the opening experiment. At 200 s on the 78 µL loop the skid makes 139.9 nm with $\sigma(\ln d) = 0.349$ and 92.7 % encapsulation, and the loop change at 240 s takes it to 215.8 nm with $\sigma(\ln d) = 0.356$ and the same encapsulation by 420 s, as shown in Figure 8.3. That is 3.76 times the aging time for 1.54 times the size, because only the growth term grows: at 100 mol/m^3 of buffer the map assembles particles at 97.5 nm and adds 42.7 and 119.7 nm over the two ages, a factor of 2.81 that is $3.76^{0.779}$ to three figures.

> **TRY IT IN THE APP** **4 · Quench Timing · the Opening Experiment**
>
> Load the scenario and press Run. Watch the true particle size alone. It sits near 139.9 nm while the skid runs on the 78 µL loop, and when the loop goes to 293 µL at 240 s it climbs to 215.8 nm by 420 s. No flow, no ratio and no charge ratio moved.

Raising the quench ratio from 2.0 to 3.5 volumes on the same loop moves the size by 0.46 nm, because the ethanol is already below 17 % and the growth it allows is spent. Case 2 reaches the same conclusion along the other path: the 78 µL loop gives 234 ms at 20 mL/min and 65 ms at 72 mL/min, and the size follows at 171.1 nm and 124.5 nm rather than in proportion. Case 1 moves the flow ratio instead. At 3:1 the ethanol content falls from 50 to 25 %, f_E falls to 0.0156, and the size falls to 98.0 nm while the ratio block holds the charge ratio at 6.01 and every one of the four feed pumps moves.

8.6.2 Faults That No Flow Meter Sees

Three cases turn on measurements that do not exist. In case 3 a lipid lot 22 % under its label takes the true charge ratio from 6.00 to 4.62 and the true encapsulation to 87.6 % with every flow meter exactly on setpoint. The first encapsulation result that shows it arrives 11.5 min after the lot went on, and the size itself moves by less than 0.1 nm. In case 5 a pump head pulsating ±40 % at 1.2 Hz leaves its average delivery unchanged, so the meter sees almost nothing through its own lag, while the modeled spread rises from 0.349 to 0.400 and the mean from 140.5 to 150.4 nm, because the jets get a different flow ratio every half second and the loop carries every one of them forward. In case 6 the release rule reads the last reported size and the flow drops to 21 mL/min, which takes the aging to 223 ms and the size to 168 nm: 239.5 mL of off-specification material reaches the hold tank by 420 s, against 1.2 mL when the same rule reads the soft sensor.

Case 8 shows what a start-up makes. With the pumps ramping together the first material reaches 191.8 nm at 10 s, because the loop holds a fixed volume while the flow is still a fraction of its design value, and the valve first accepts at 21.9 s and keeps all but 0.8 mL of that material out of the tank. Case 9 asks for a size the fitted map cannot deliver. The largest in-window size is 333.4 nm, at the corner of the box, and a logged sweep of 5992 points over flows, loops and buffer strengths reaches 380 nm at none of them, missing by 46.6 nm at best. The supervisor saturates at 300 mol/m^3 on the 293 µL loop, reports *size* as the binding constraint and stays inside the box. With the guard off it picks the 1000 ms loop, predicts 371.6 nm and leaves 89 % of its age distribution outside the fitted window.

8.6.3 Scale-Out and What the Blend Hides

Four channels behind one set of meters is the application's second axis. In case 7 a 12 % aqueous maldistribution leaves the manifold total at 159.6 mL/min, inside the noise of its own meters, while the four channels run at flow ratios of 1.13, 0.89, 1.07 and 0.95 and make 126.6, 163.6, 132.9 and 150.8 nm. The blend is 142.8 nm with $\sigma(\ln d) = 0.364$ and meets every window. The only online sign of a restriction is pressure: channel 2 closing to 40 % of its share takes PT-101 from 2.49 to 4.29 bar against its 6 bar alarm, and the channels separate further, to 125.3 and 190.1 nm.

> **TRY IT IN THE APP** **7 · Four Channels, One Flow Meter**
>
> Run the scenario to 480 s and compare the blend with the channels. The manifold total stays at 159.6 mL/min while the four channels make 126.6, 163.6, 132.9 and 150.8 nm, and the blend is 142.8 nm inside its window. Then run past 480 s and watch PT-101 rise from 2.49 to 4.29 bar.

Figure 8.4 sweeps that fault from an even split to 30 %. The worst channel leaves the 15 nm window near 8 % and is 89.7 nm off target at 30 %, while the blend of four channels is 14.6 nm off and the blend of two is 22.9 nm off at the same fault. The blend's spread crosses its 0.45 limit between 24 and 27 % with two channels and reaches only 0.446 at 30 % with four. More channels therefore make the blend look better at the same fault: at 24 % the four-channel blend is 9.8 nm off target with three of its four channels outside their windows, against 14.8 nm off with both channels outside when there are two, and the two-channel blend is the only one that ever leaves the size window itself, which it does at 27 %. Averaging more streams is not the same as making them alike, which is the reason the application scores the blend and reports, separately, how much of the accepted volume was made while a channel was outside its window.

8.6.4 The Challenge and the Policy Comparison

The challenge runs four channels at a 200 nm target through 1500 s and four faults in sequence: the aqueous manifold drifts 12 % out at 180 s, a lipid lot 8 % under its label goes on at 420 s, a buffer concentrate 10 % under its label follows at 620 s, the lipid strength then drifts at −12 % per hour, and the lipid pump starts pulsating at 1080 s. The target is 560 mL/min or more of formulated bulk, at most 150 mL of off-specification material accepted, residual ethanol under 0.5 % and no overflow.

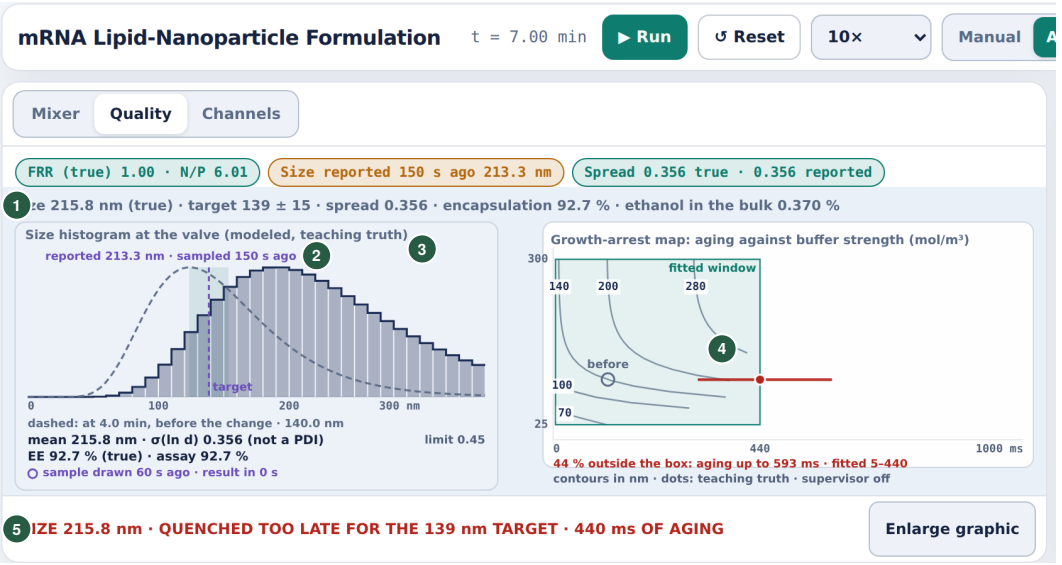


Figure 8.3 Case 4 at $t = 420$ s, three minutes after the aging loop went from 78 to 293 µL, on the Quality view in AUTO with fixed ratios and the soft-sensor release rule. The histogram at the valve is centered on 215.8 nm with $\sigma(\ln d) = 0.356$, and the dashed curve beside it is the distribution the same skid made at 240 s on the shorter loop, 140.0 nm. The at-line result, the filled circle, reads 213.3 nm from a sample drawn 150 s earlier, and its next sample is already 60 s old. The growth-arrest map puts the operating point on the 440 ms edge of the fitted box with 44 % of the age distribution beyond it. Encapsulation is 92.7 % true against an assay reporting the same value seven minutes after its own sample, a delay Table 8.2 prints, and the status line names the cause of the miss against the 139 nm target the case ships with. The picture is the application's window cropped to the header and the Quality view, with the Control panel, the trends and the scorecard cut. Callouts: (1) the summary line with the true size against the target and its window, (2) the at-line result and the age of its sample, (3) the size histogram at the valve with the dashed distribution from before the loop change, (4) the operating point on the growth-arrest map at the edge of the fitted box, (5) the status line.

The allowance is three line volumes, because when an unmeasured feed steps the material already between the mixer and the valve was made at the old condition and no release rule can reach it.

Table 8.5 compares every policy on the same plant, the same faults, the same instruments and six paired seeds. As shipped, fixed ratios at 32 mL/min per channel with the valve accepting everything make 372.5 mL/min and fail on rate at every seed, for the physical reason the case teaches: the quench triples the volume, so four channels at 32 mL/min present 384 mL/min to an exchange that passes 700, and the rest of the ceiling is unused. Raising the flow is what costs the size. Coordinated ratio control buys it back with the buffer strength and reaches 441.7 mL/min at 44 mL/min per channel, with the trim already at 273 mol/m^3 of the 300 mol/m^3 it has. At 48 mL/min it falls back to 415.7 mL/min, because the trim saturates and the product leaves the window. The supervisor, free to lengthen the loop as well, moves to the 293 µL loop at 56 mL/min per channel and reaches 600.0 mL/min, where the exchange and not the chemistry is the limit. **The timed quench is therefore worth 158 mL/min of bulk that ordinary ratio control cannot buy**, with the same hardware, feeds, faults and information. That margin is the answer to the question the application was built to ask, and it is a measurement on this plant and this formulation family, not a general claim.

The ablations in Table 8.5 say which layer does what. The size trim is what makes the rate: without it the supervisor's own point gives 368.8 mL/min and leaks 522.4 mL, because nothing corrects the

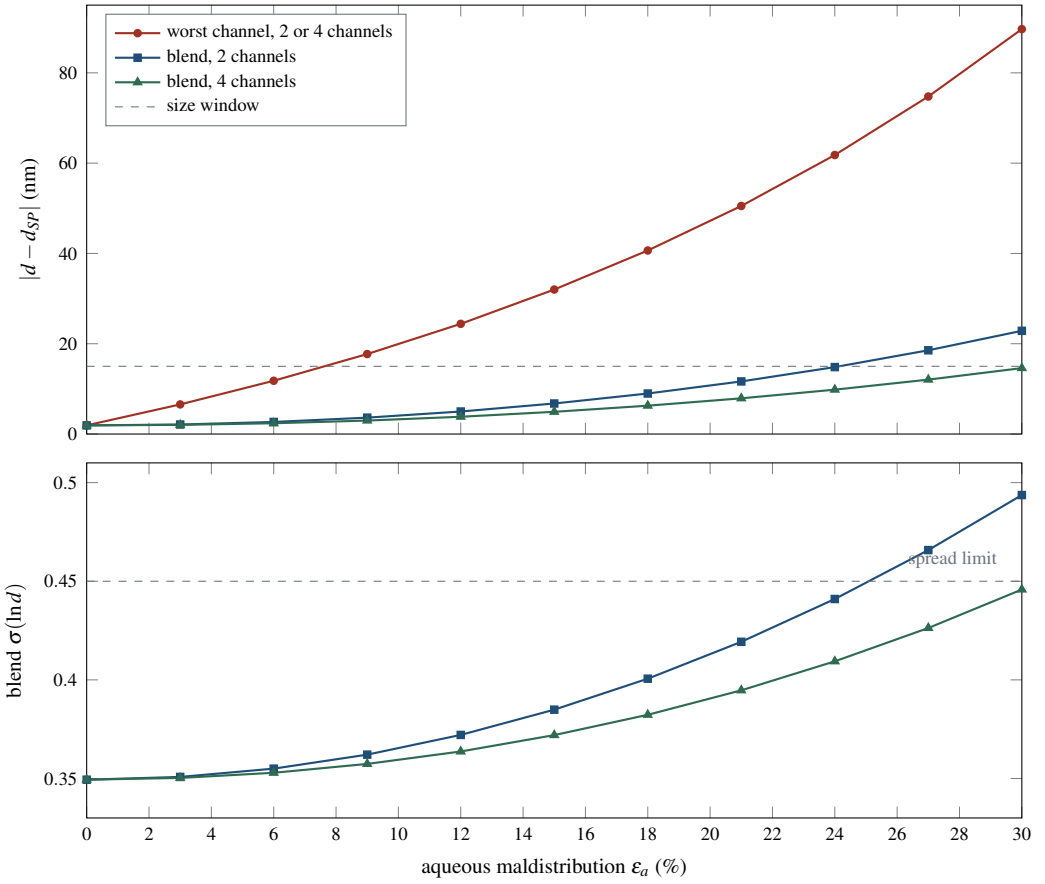


Figure 8.4 A steady aqueous maldistribution at the design recipe, read at 170 s. Above, the largest departure of any single channel from the target and the departure of the blend that leaves the valve, for two and for four channels. Below, the blend's modeled spread. Circles mark the worst channel, squares the blend of two channels and triangles the blend of four, and the dashed lines are the limits. The worst channel leaves the 15 nm window near 8 % while the blend stays inside it, and the blend's spread catches the fault only when it is severe.

concentrate lot. The conductivity probe is what makes the quality: the same supervisor on the metered blend runs 35.7 mL/min faster and leaks 776.7 mL, passing none of the six seeds, because no flow meter sees a drum that is 10 % weak while the probe on the blend line does. The window guard is worth more than either: without it the supervisor extrapolates, the trim chases a map that is wrong there, and the rate collapses to 200.4 mL/min with 1928.5 mL leaked. Figure 8.5 puts five of those rows against the challenge's two targets.

Table 8.5 The Challenge Over 1500 s, Six Paired Seeds. Rate Is Accepted Bulk per Minute and Off-Spec the Accepted Volume Outside a Window

Policy	Pass	Rate (mL/min)	Off-spec (mL)	Spread worst
fixed ratios, 32 mL/min, no gate (shipped)	0/6	372.5	0.0	0.441
with the soft-sensor gate	0/6	314.9	0.0	0.441
with the gate on the reported size	0/6	288.6	0.0	0.441
coordinated, 32 mL/min	0/6	314.4	526.8	0.446
coordinated, 40 mL/min	0/6	377.0	65.7	0.439
coordinated, 44 mL/min	0/6	441.7	0.0	0.434
coordinated, 48 mL/min	0/6	415.7	0.0	0.430
the supervisor's point, held by hand	0/6	619.2	1297.2	0.423
supervisor on the probe (suggested)	**6/6**	**600.0**	**43.1**	0.423
on the metered blend instead	0/6	635.7	776.7	0.491
without the size trim	0/6	368.8	522.4	0.411
without the bias update	0/6	600.3	349.7	0.422
with the bias update at 0.4	4/6	598.5	197.4	0.425
without the window guard	0/6	200.4	1928.5	0.647
with the gate on the reported size	5/6	581.0	51.5	0.423
with no release margin	6/6	628.1	47.9	0.423
with the valve accepting everything	6/6	644.8	63.9	0.423

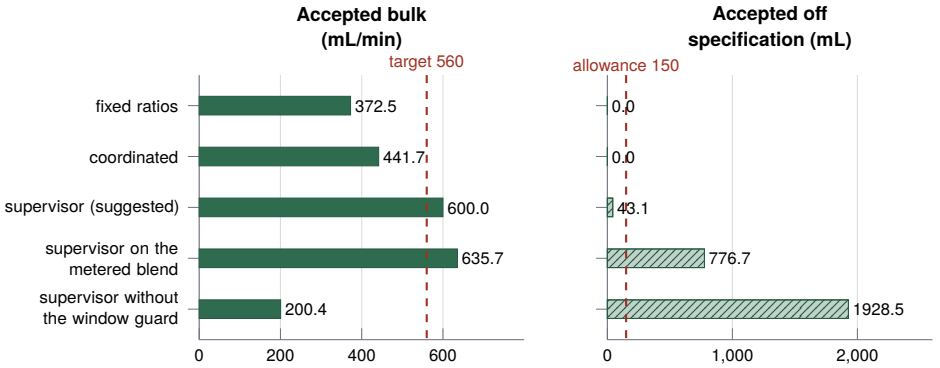


Figure 8.5 Table 8.5 redrawn: five of its policies on the same plant, the same four faults and six paired seeds. Solid bars are the accepted bulk rate, hatched bars the accepted volume made outside a window, and the dashed lines are the challenge's rate target and its off-specification allowance. Only the suggested supervisor meets both, and it passes six of six seeds. The two rows below it run faster or further and fail on quality, which is why the conductivity probe and the window guard are each necessary.

> ✦ **DID YOU KNOW?** **The Gate That Buys Nothing**
>
> From the simulator: with the supervisor in charge, a valve that accepts everything makes 644.8 mL/min and still passes six of six seeds, against 600.0 mL/min when the gate is in service (Table 8.5). Here the diversion rule costs rate and buys no quality.

Two results in that table are negatives and are reported as such. The release gate adds nothing here: with the same supervisor and the valve accepting everything the challenge makes 644.8 mL/min and still passes six of six, so the gate costs about 45 mL/min of rate and buys no quality. Its value appears where the control layer is blind, in cases 6 and 8 and in the metered-blend row, not where the control layer already holds the product inside its window. And the release margin, which the suggested settings keep at 20 % of every window, is worth 4.8 mL of off-specification material for 28 mL/min of rate, which is close to a wash.

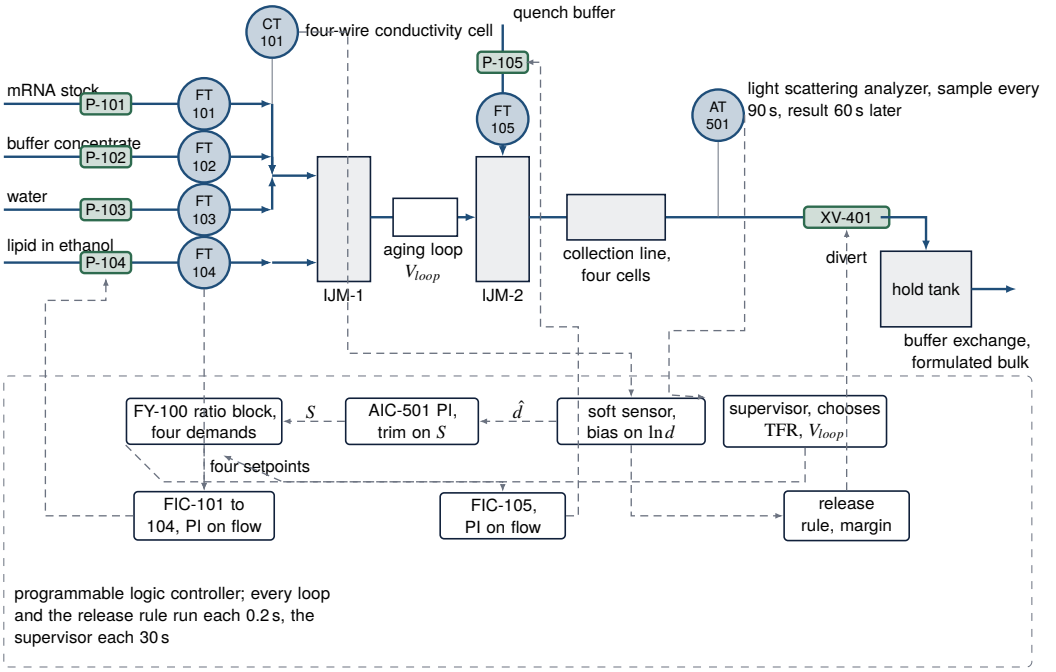


Figure 8.6 One channel of the skid on real hardware. Four metered feeds set the four demands at once through the ratio block FY-100, whose outputs are the setpoints of the four flow loops, and the quench pump follows the total flow in ratio. Only the buffer strength S is trimmed, by AIC-501 on the soft sensor, which reads the conductivity cell and the flow meters and is corrected each time the sampled analyzer answers. The supervisor sets the total flow and the aging loop, and the release rule drives the divert valve. Solid arrows are process streams and dashed gray arrows are signals. Notice that the only instrument that sees the product itself sits downstream of everything the controller moves.

Robustness was checked on twelve seeds rather than six. The shipped settings fail at all twelve. The suggested settings meet the target at eleven, with a mean rate of 598.1 mL/min and a mean of 65.4 mL accepted off-specification. The twelfth is a documented miss rather than noise: after the pulsation at 1080 s the true mean creeps to 215 to 219 nm against a 185 to 215 nm window while the soft sensor reads 193 to 202 nm and the last reported size 193 nm, so neither instrument asks the valve to divert and 295.2 mL is accepted against the 150 mL allowance. It is the blindness of case 5 arriving at the end of a long run, and every retuning that removed it cost rate at other seeds. Over the 31 seeds of the final validation the suggested settings pass 27 at both steps, and the four misses accept 166 to 466 mL.

The scoring deserves its own sentence, because the passing run is not as clean as its scorecard. At 1500 s the suggested run's four channels are at 171.0, 253.5, 185.3 and 225.4 nm against the 200 nm target, the blend is 206.3 nm with $\sigma(\ln d) = 0.412$, and 13 500 mL of the 15 409 mL accepted was made while at least one channel was outside its window. The criterion stays on the blend, because the blend is what leaves the valve and reaches the tank, and a light-scattering measurement of that tank would report the blend's mean and width. The application therefore scores the blend and reports the other figure beside it as teaching truth, in a scorecard tile and in the exported file.

8.7 Technology in Practice

Figure 8.6 puts the structure of Section 8.5 on real hardware: four metered feeds under one ratio block, the quench in ratio to the total flow, the size trim and the release decision at the valve.

What Measures It

Three kinds of instrument carry the loops. The flow meters make the ratios possible, because a ratio block can only hold what something measures. The conductivity probe CT-101 stands in for the aqueous buffer strength, and a probe of that kind serves concentration analysis over limited ranges only [19]. Its service problem is fouling: ions bonding to the electrodes of a two-electrode cell read as a liquid less conductive than it is, while a four-wire cell measures the voltage across the liquid alone and ignores any coating on its inner electrodes [19]. The drift a case applies to CT-101 is that error. The third instrument is the analyzer that decides the product.

> **TECHNOLOGY** **Sensor: Light Scattering on a Drawn Sample**
>
> The common industrial size measurement for these particles is dynamic light scattering. A laser lights a dilute sample, and the speed at which the scattered intensity fluctuates gives a diffusion coefficient and from it the hydrodynamic diameter, over a range of 0.5 to 1000 nm [4]. The number that reaches the controller depends on how the measurement is analyzed. The default output of a common inversion method departs from the cumulant z-average typically by more than 15 percent, the width descriptors are unreliable, and the governing standard ISO 22412:2017 defines the size only as the central value of the distribution [17]. The method is also biased toward larger particles and sensitive to aggregates and contaminants [4], so one stray result sits high on a trend and the next sample does not repeat it.

What Moves It

Five pumps are the final elements, and nothing else on the skid moves except the divert valve. Each pump is a metered feed whose delivery follows a command signal over a range set at the pump [19], modeled here as a first-order drive with $\tau = 0.8\,\mathrm{s}$ that reports when a request passes its capacity. The aging loop is the supervisor's second actuator, and the valves that select it are not modeled, which is why a loop change is given a 150 s lockout rather than a stroke time. XV-401 is the last element in the chain and the only one whose action is a decision rather than a flow.

> **TECHNOLOGY** **Final Element: Pumps That Meter Their Own Flow**
>
> A metering pump carries its pump mechanism, its motor and its drive electronics in one package, and a 4 to 20 mA signal sets its delivery from zero to the full flow of a range that knobs on the pump establish [19]. That is what makes a feed a controlled variable instead of a hand setting. The cheaper kind drives a plunger with a solenoid at variable intervals, so its delivery is a train of discrete pulses rather than a continuous flow, which is simple and reliable wherever non-continuous flow is permissible [19]. A formulation skid is not such a place. A head that pulses at constant average delivery leaves the meter reading its setpoint while the jets see a different ratio every half second, which is the fault of case 5.

Where the Controller Runs

A skid like this one runs on a programmable logic controller, which serves either as a local controller under a supervisory scheme or as the primary controller of a smaller system, with control cycles in the range of milliseconds to minutes [20]. In the application every flow loop, the size trim and the release rule execute each 0.2 s, while the supervisor evaluates its grid every 30 s. The structure maps onto standard blocks. A signal multiplying relay, or ratio station, multiplies a wild flow signal by a constant and sends the product to a captive flow controller as its setpoint [19]. FY-100 is that block solved for four demands at once, and AIC-501 is an ordinary slow loop whose output is one of those demands rather than a valve position.

IN PRACTICE **What a Divert Valve Has to Be Able to Say**

Guidance for continuous manufacturing asks that the residence-time distribution be characterized, because it fixes the time available for transport and transformation and so lets each portion of material be tracked [21]. Diversion follows from that tracking. Criteria should be established that trigger the start and the end of a diversion period and the restart of collection, and how much material is diverted depends on the process dynamics, that distribution, the control strategy, the severity of the disturbance and where the sampling and diversion points sit [21]. The release rule at XV-401 is exactly such a criterion, and its margin and its two line volumes of hold-off are the designer's answers to the last two factors.

What Is Modernizing It

Three developments bear on this skid. The first is better size metrology. Multi-angle dynamic light scattering represents the size distribution more accurately and is routinely used in formulation development and quality control [4]. The second is prediction between samples. Modeling, digital twins, automation and real-time monitoring are the opportunities named for this industry [4], and the soft sensor of Section 8.5 is the plain version of one: a fitted map on the meters, corrected by a bias each time the laboratory answers. The third is the network under all of it. Computing at the edge acts on data where it is collected rather than shipping it to a data center [20], which is what would let a per-channel measurement act inside one line volume.

8.8 Limitations

The size law is a published model form fitted here to one published table, so every size the application reports is an interpolation inside a box of two axes at one pH, one cargo type, one mixer geometry and one formulation family, and the agreement reported in Section 8.4 is a fit plus two withheld cargo tables rather than a validation. Outside that box the three extension factors carry the prediction, and the ethanol exponent that carries the most is a teaching value fitted to no data: at an exponent of 12 the suggested challenge settings no longer pass. The quantitative size of every flow-ratio result in this chapter therefore depends on one number that a measurement could change, while the direction of each result does not.

The plant is lumped where the lesson does not need detail. The aging loop is evaluated on a volume clock rather than integrated, and that clock is first-order accurate in the step under a fast-swinging flow: against an exact integration the age of a packet under the 1.2 Hz pulsation is 8.6 % low at 0.05 s and 4.5 % low at half the step, where the instantaneous V_{loop}/TFR the model avoids would be 19.9 % low and would not converge at all. The error does not reach a lesson, because the size is a power law of the age averaged over nine nodes and then over a collection line, and halving the step moves the pulsation case by 0.05 nm. The nine-node quadrature is likewise 0.75 per mille above a 2001-node evaluation in the mean size at 440 ms, but it resolves the share of the distribution inside the box only in ninths, which overstates that share's resolution by about eleven points.

OPEN CHALLENGE **Channels That Drift Apart**

Parallel mixing lines are how throughput is won, and they are also where quality is lost. Scaling out risks altered mixing regimes and mixer-to-mixer variability, with inconsistent flow rates, pressure imbalance between the channels or a channel clogging, and the result is non-uniform product [4]. The hard part is that the blend hides it, as Figure 8.4 shows. Progress would be a measurement per channel, or a flow loop per channel rather than one manifold, so that a drifting line can be found before its material is averaged into the tank.

Encapsulation is nominal and set at the mixer, morphology is absent, and nothing in the application speaks to intact RNA, transfection or potency. The buffer exchange is a declared recovery map. Faults enter on a schedule as steps and ramps, not with the texture a plant's own upsets have. The pressure trip is an educational guard rather than certified protection, and every window, limit and delay is

a teaching value. A simulated pass is not batch release, and real limits come from the product's own specification, a pharmacopeial chapter or regulatory guidance [15]. No learning study has been run, so no claim is made that the application improves learning. What the cases establish is which experiments a learner can carry out and which misconception each one addresses.

OPEN CHALLENGE **Running for Weeks Instead of Hours**

Deposits on channel walls change the mixing and the residence time, and a surface treatment that resists them has held particle size and an encapsulation above 85 percent throughout a run [22]. The demonstration was one chip over a few hours. Production will need either much longer continuous runs or cleaning and reuse that are reliable and reproducible across hundreds of parallel units [22]. Progress would be a documented campaign length at which the size law and the age distribution are still the ones the controller assumes.

8.9 Conclusion

A formulation skid whose particle size is set by the age of the material has two actuators for one quality attribute, a loop volume and a flow, and they are not equivalent. The simulator of this chapter carries a published size law along a normalized age distribution on a volume clock, with the true size, the age distribution and each channel's own product logged beside the instruments that the controllers read. Moving the aging loop from 78 to 293 µL takes the product from 139.9 nm to 215.8 nm at the same flows and the same charge ratio, while 3.5 volumes of quench instead of 2.0 move it by 0.46 nm, so the arrest is the ethanol's and the quench ratio is a cost rather than a quality knob. On the challenge, coordinated ratio control runs out of buffer strength at 441.7 mL/min while a supervisor that may also choose the loop reaches 600.0 mL/min, a margin of 158 mL/min with the same instruments and faults. The release gate adds nothing once that control layer is in place, and the conductivity probe, the size trim and the window guard are each necessary. A 12 % maldistribution separates four channels by 37 nm while their one flow meter reads its setpoint, and 13 500 mL of the 15 409 mL the passing run accepted was made with a channel outside its own window. The map is fitted inside one box and the factors that carry it outside are teaching values. A fuller model would fit the ethanol dependence to measured data, add pH as an axis, model the morphology the anchor controls alongside the size, and replace the exchange map with a membrane model. The simulator shows what a quality attribute looks like when its controlled variable is a time.

KEY POINTS

- The size of these particles is set by how long they travel before the quench floods them, so the clock and not the mixing is the thing to control.
- The ratio block holds the flow ratio, the charge ratio and the buffer strength at once, which leaves the buffer strength free as the trim knob.
- A supervisor free to choose the aging loop as well reaches 600.0 mL/min against 441.7 mL/min for coordinated ratio control.
- One flow meter in front of four channels reads its setpoint while the channels separate by 37 nm, and the blend still meets every window.

Where else this applies. Volume 5, Chapter 7 tracks material through a residence-time distribution to a divert valve, Chapter 7 trims a ratio blend on a slow analyzer, and Volume 5, Chapter 10 runs a soft sensor that a late measurement corrects.

References

[1] C. Devos, A. Udepurkar, P. Sagmeister, A. S. Hodlewsky, J. Chen, A. Hatas, N. Ostrovsky, M. Al-Jazrawe, J. I. Ren, A. Y. Liu, R. D. Braatz, and A. S. Myerson, "Manufacturing mRNA-loaded lipid nanoparticles with precise size and morphology control," *ACS Nano*, vol. 19, no. 38, pp. 33 991–34 002, 2025.

[2] A. S. Myerson, M. Krumme, M. Nasr, H. Thomas, and R. D. Braatz, "Control systems engineering in continuous pharmaceutical manufacturing May 20–21, 2014 continuous manufacturing symposium," *Journal of Pharmaceutical Sciences*, vol. 104, no. 3, pp. 832–839, 2015.

[3] P. V. Danckwerts, "Continuous flow systems," *Chemical Engineering Science*, vol. 2, no. 1, pp. 1–13, 1953.

[4] J. Hong Hoang, M. Ott, E. Samaridou, M. Beck-Broichsitter, and J. Simon, "Industrial perspective on the manufacturing of lipid nanoparticles for nucleic acid delivery," *Pharmaceutics*, vol. 18, no. 4, p. 489, 2026.

[5] A. D. Bangham, M. M. Standish, and J. C. Watkins, "Diffusion of univalent ions across the lamellae of swollen phospholipids," *Journal of Molecular Biology*, vol. 13, no. 1, p. 238, 1965.

[6] S. Batzri and E. D. Korn, "Single bilayer liposomes prepared without sonication," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 298, no. 4, pp. 1015–1019, 1973.

[7] L. B. Jeffs, L. R. Palmer, E. G. Ambegia, C. Giesbrecht, S. Ewanick, and I. MacLachlan, "A scalable, extrusion-free method for efficient liposomal encapsulation of plasmid DNA," *Pharmaceutical Research*, vol. 22, no. 3, pp. 362–372, 2005.

[8] S. C. Semple, A. Akinc, J. Chen, A. P. Sandhu, B. L. Mui, C. K. Cho, D. W. Y. Sah, D. Stebbing, E. J. Crosley, E. Yaworski, I. M. Hafez, J. R. Dorkin, J. Qin, K. Lam, K. G. Rajeev, K. F. Wong, L. B. Jeffs, L. Nechev, M. L. Eisenhardt, M. Jayaraman, M. Kazem, M. A. Maier, M. Srinivasulu, M. J. Weinstein, Q. Chen, R. Alvarez, S. A. Barros, S. De, S. K. Klimuk, T. Borland, V. Kosovrasti, W. L. Cantley, Y. K. Tam, M. Manoharan, M. A. Ciufolini, M. A. Tracy, A. de Fougerolles, I. MacLachlan, P. R. Cullis, T. D. Madden, and M. J. Hope, "Rational design of cationic lipids for siRNA delivery," *Nature Biotechnology*, vol. 28, no. 2, pp. 172–176, 2010.

[9] P. R. Cullis and M. J. Hope, "Lipid nanoparticle systems for enabling gene therapies," *Molecular Therapy*, vol. 25, no. 7, pp. 1467–1475, 2017.

[10] X. Hou, T. Zaks, R. Langer, and Y. Dong, "Lipid nanoparticles for mRNA delivery," *Nature Reviews Materials*, vol. 6, no. 12, pp. 1078–1094, 2021.

[11] B. K. Johnson and R. K. Prud'homme, "Chemical processing and micromixing in confined impinging jets," *AIChE Journal*, vol. 49, no. 9, pp. 2264–2282, 2003.

[12] I. V. Zhigaltsev, N. Belliveau, I. Hafez, A. K. K. Leung, J. Huft, C. Hansen, and P. R. Cullis, "Bottom-up design and synthesis of limit size lipid nanoparticle systems with aqueous and triglyceride cores using millisecond microfluidic mixing," *Langmuir*, vol. 28, no. 7, pp. 3633–3640, 2012.

[13] N. M. Belliveau, J. Huft, P. J. C. Lin, S. Chen, A. K. K. Leung, T. J. Leaver, A. W. Wild, J. B. Lee, R. J. Taylor, Y. K. Tam, C. L. Hansen, and P. R. Cullis, "Microfluidic synthesis of highly potent limit-size lipid nanoparticles for in vivo delivery of siRNA," *Molecular Therapy - Nucleic Acids*, vol. 1, p. e37, 2012.

[14] J. A. Kulkarni, M. M. Darjuan, J. E. Mercer, S. Chen, R. van der Meel, J. L. Thewalt, Y. Y. C. Tam, and P. R. Cullis, "On the formation and morphology of lipid nanoparticles containing ionizable cationic lipids and siRNA," *ACS Nano*, vol. 12, no. 5, pp. 4787–4795, 2018.

[15] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, "Liposome drug products: Chemistry, manufacturing, and controls; human pharmacokinetics and bioavailability; and labeling documentation. Guidance for industry," Guidance document, https://www.fda.gov/media/70837/download, Apr. 2018.

[16] D. E. Koppel, "Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants," *The Journal of Chemical Physics*, vol. 57, no. 11, pp. 4814–4820, 1972.

[17] N. Farkas and J. A. Kramar, "Dynamic light scattering distributions by any means," *Journal of Nanoparticle Research*, vol. 23, no. 5, p. 120, 2021.

[18] G. I. Taylor, "Dispersion of soluble matter in solvent flowing slowly through a tube," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 219, no. 1137, pp. 186–203, 1953.

[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] U.S. Food and Drug Administration, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, "Q13 continuous manufacturing of drug substances and drug products. Guidance for industry," ICH harmonised guideline, https://www.fda.gov/media/165775/download, Mar. 2023.

[22] O. Milić, S. M. Savić, M. Zurković, B. Stanojević, and S. Savić, "Lipid nanoparticles for gene therapy: Unresolved challenges in manufacturing, transdermal delivery, machine learning, endosomal escape, and the protein corona," *Pharmaceutics*, vol. 18, no. 8, p. 991, 2026.

8.10 Exercises

8.1 Concept check · 2 min

Which operator demand of this skid sets how long the material ages before the second junction arrests it?

(a) The flow ratio, because it fixes the mixing time in the jets

(b) The quench ratio, because it fixes how far the stream is diluted

(c) The total flow, because the age is the loop volume divided by it

(d) The charge ratio, because it fixes how much lipid is uncomplexed

8.2 Concept check · 2 min

A two-feed version of this skid carries only the mRNA stock and the lipid solution, at fixed feed concentrations. The operator steps the flow ratio from 1:1 to 3:1. What happens to the charge ratio N/P?

(a) It holds, because the quench dilutes both components equally

(b) It holds, because the ratio block solves for it

(c) It rises by a factor of three, since three volumes of aqueous arrive

(d) It falls by a factor of three, from 3.00 to 1.00

8.3 Concept check · 1 min

AIC-501 holds the particle size by moving the aqueous buffer strength rather than the flow ratio or the total flow. Why?

(a) The ratio block can move it while the flow ratio and the charge ratio stay on their demands

(b) It is the only input the conductivity probe can measure

(c) Moving the total flow would change the throughput, and the buffer strength does not

(d) The size law is linear in the buffer strength and nonlinear in the others

8.4 Theory and calculation · 7 min

The ratio block of Equation (8.1) is asked for TFR $= 40\,\mathrm{mL/min}$, FRR $= 1.0$, $N/P = 6.0$ and $S = 100\,\mathrm{mol/m^3}$, with $c_N = 3.858\,\mathrm{mol/m^3}$, $c_P = 1.288\,\mathrm{mol/m^3}$, $S_1 = 25\,\mathrm{mol/m^3}$, $S_2 = 500\,\mathrm{mol/m^3}$ and $R_q = 2.0$.

(a) Find the four feed flows and the quench flow. *Checkpoint:* 9.99, 3.50, 6.51 and 20.0 mL/min, with 80 mL/min of quench.

(b) The same demands are asked at FRR $= 3.0$. Find the new feed flows, and say which feed takes up the change. *Checkpoint:* 4.99, 5.75, 19.26 and 10.0 mL/min.

(c) At FRR $= 0.50$ the block cannot deliver $S = 100\,\mathrm{mol/m^3}$. It keeps Q_1 and the aqueous total Q_a, gives the concentrate only what the stock leaves, $Q_2 = Q_a - Q_1$, and sets Q_3 to zero. Find the strength it delivers. *Checkpoint:* About $26\,\mathrm{mol/m^3}$.

8.5 Theory and calculation · 8 min

Use the fitted map of Equation (8.3) with $a_0 = 61.84\,\mathrm{nm}$, $a_1 = 65.98\,\mathrm{nm}$, $r_0 = 0.154$ and $r_1 = 1.653\,\mathrm{nm/ms^p}$, $C = 81.03\,\mathrm{mol/m^3}$, $D = 0.0256\,\mathrm{m^3/mol}$ and $p = 0.7785$, at $S = 100\,\mathrm{mol/m^3}$.

(a) Find the assembly size $d_0(S)$ and the growth coefficient $R(S)$. *Checkpoint:* 97.5 nm and $1.047\,\mathrm{nm/ms^p}$.

(b) Find the aging time of the 78 µL and the 293 µL loop at 40 mL/min, and the size the map gives at each. *Checkpoint:* 117 and 440 ms, giving 140.2 and 217.2 nm.

(c) The age rises by a factor of 3.76 and the size by a factor of 1.54 on the application. Reconcile the two. *Checkpoint:* $3.76^{0.779} = 2.80$ on the growth term alone.

8.6 Theory and calculation · 5 min

At the design point the mixed stream is 50 % ethanol by volume before the quench. The quench adds R_q volumes of aqueous buffer per volume of mixed stream, and the exchange must bring the line's ethanol to the 0.5 % teaching limit.

(a) Find the ethanol fraction after 2.0 volumes of quench, and the factor f_E of Equation (8.4) before and after it. *Checkpoint:* 16.7 %, with 1.000 and 0.00137.

(b) Find the same at 3.5 volumes, and say what the difference is worth. *Checkpoint:* 11.1 % and 1.2×10^{-4}.

(c) Find the diavolumes needed from 16.62 % to the limit, and the residual the controller's extra 0.30 leaves. *Checkpoint:* 3.50 diavolumes, leaving 0.37 %.

8.7 App challenge · 8 min

Load scenario 4, Quench timing, the opening experiment. The aging loop goes from 78 to 293 µL at 240 s. Use the default seed and read the exported columns `truth_d_nm` and `tau_ms`.

(a) Run to 200 s, then on to 420 s, and read the true size and the aging time at each. *Checkpoint:* 139.9 nm at 117 ms, then 215.8 nm at 440 ms.

(b) Without resetting, set the quench ratio to 3.5 volumes and run to 700 s. Compare with the same run left at 2.0 volumes. *Checkpoint:* 215.5 nm against 216.0 nm.

(c) Say what part (b) costs, reading the discharge pressure in each run. *Checkpoint:* 2.49 against 4.10 bar.

Accept a size within 0.5 nm and a pressure within 0.05 bar.

8.8 App challenge · 8 min

Load scenario 6, The size result is a minute old. The release rule starts on the last reported size, and the total flow steps to 21 mL/min at 200 s. Use the default seed and read the Off-specification accepted tile.

(a) Run to 420 s and read the tile, the true size and the aging time. *Checkpoint:* 239.5 mL, with 168 nm at 223 ms.

(b) Reload the scenario, set the release rule to Soft sensor, and run the same step to 420 s. *Checkpoint:* 1.2 mL.

(c) Explain the factor of 200, using the age of the reported size.

Accept a volume within 2 mL.

8.9 App challenge · 8 min

Load scenario 7, Four channels, one flow meter. A 12 % aqueous maldistribution arrives at 180 s and channel 2 restricts at 480 s. Use the default seed and read the exported columns `truth_ch1_d_nm` to `truth_ch4_d_nm`, `truth_d_nm`, `truth_spread` and `P_bar`.

(a) Run to 470 s and read the four channel sizes, the blend and its spread. *Checkpoint:* 126.6, 163.6, 132.9 and 150.8 nm, blending to 142.8 nm at 0.364.

(b) Does the blend meet the 15 nm size window and the 0.45 spread limit? Say what an instrument at the valve would report. *Checkpoint:* Yes to both.

(c) Run on to 600 s and read the pressure and the channel sizes again. *Checkpoint:* 4.29 bar, with 125.3, 190.1, 131.2 and 148.1 nm.

Accept a size within 0.5 nm and a pressure within 0.05 bar.

8.10 Application area · 4 min

Section 8.4 reports the fitted map at 5.11 % rms on the table it was fitted to, 8.76 and 7.96 % on two withheld cargo tables at the same pH, and 28.15 % on a table taken at pH 4.5.

(a) Say what those four figures together establish, and what they do not.

(b) A leave-one-column-out refit predicts the 440 ms column at 90.6 % rms. Name the feature of the application that this number justifies, and say what it protects against.

8.11 Application area · 3 min

The challenge target asks for 560 mL/min of bulk with at most 150 mL accepted outside a window, and the suggested settings pass it at six of six paired seeds.

(a) State two reasons from the chapter why that pass is not a statement about a real product.

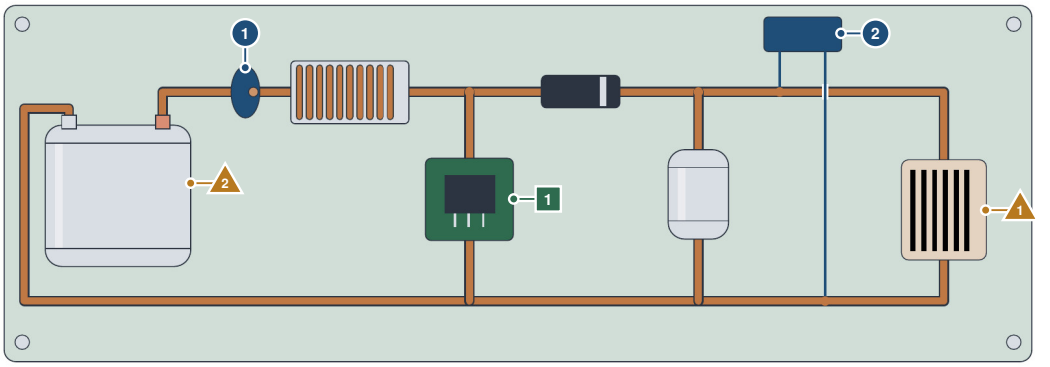
(b) The passing run made 13 500 mL of its 15 409 mL while at least one channel sat outside its own window. Say why the application still scores the blend.

8.12 Review the agent · 5 min

An AI assistant is asked how to raise throughput on this skid and replies as follows. “Hold the 78 µL loop and raise the total flow to 56 mL/min per channel. That shortens the age, so recover the size by taking the quench ratio from 2.0 up to 3.5 volumes: a longer quench leaves the particles less diluted at the junction and holds them larger. The flow loops and the exchange have the capacity.” Find the flaw and say how to check it.

CHAPTER 9

An Inner Current Loop Against the Right-Half-Plane Zero of a Boost Converter



The controller moves 1 the transistor switch, through the share of each cycle it stays on

The controller reads 1 current in the coil (fast inner loop) 2 output voltage (slow outer loop)

What disturbs it 1 load changes 2 supply voltage sags

Notice that the current is read on the coil ahead of the switch and the voltage across the output, so the inner loop watches the fast store of energy and the outer loop the slow one. Illustration, not to scale.

IN PLAIN TERMS A Power Box That Dips Before It Climbs

The main idea. A boost converter makes a higher voltage from a lower one by opening and closing a switch very fast. The longer the switch stays closed each time, the higher the output ends up.

The challenge. A longer closed time first cuts a coil of wire off from the output, so the output dips before it rises. A second, quicker loop that watches the current in the coil cannot remove the dip, but it keeps that current safe.

Why it matters. These converters sit between solar panels, batteries and the circuits they feed, and those circuits feel every dip.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	12 min	the first page, the fact on switching speed (Section 9.1), the history of averaged models (Section 9.2), the history of predictive control in converters (Section 9.5.1), the two try-it segments and the fact from the simulator (Section 9.6) and the key points
🕮	**Study**	60 min	the whole chapter with its three pause points and the twelve exercises at its end
⛰	**Practice**	25 min	the first page, the results (Section 9.6), Technology in Practice (Section 9.7), Limitations (Section 9.8) and the two open challenges

The control question Raising the duty of a boost converter starves the output before it feeds it, and that right-half-plane zero belongs to the power stage. What does it cost a voltage loop, and what does an inner current loop buy that more gain cannot?

Learning objectives

- Derive the right-half-plane zero from the averaged equations and compute it at an operating point.
- Explain why duty to inductor current is minimum phase and what that gives the outer loop.
- Measure the trade between dip and recovery time under gain, feedforward and component changes.
- Compute both feedforward terms and the output a current limit holds.
- Judge what an explicit MPC buys over the cascade and what the zero still charges it.

Prerequisites Sections 1.4.5 (inner-loop speed rule), 1.4.3 (static feedforward), 1.6.2 (velocity form and back-calculation) and 1.6.4 (scenarios and exported files). Section 1.6.3 for the MPC only.

The app https://apmonitor.com/apps/boost-converter/

See also Volume 7, Chapter 6 for an explicit MPC whose gain over a PID is state feedback, Volume 7, Chapter 7 for the terminal cost of a short horizon. Volume 9, Chapter 4 for offset-free MPC on a drum unit whose level swells.

9.1 Introduction

Almost every electrical load in a modern vehicle, laptop, photovoltaic string or battery pack is fed through a switching converter rather than a linear regulator, because chopping a voltage and filtering the result wastes a fraction of a percent where dissipating the difference would waste tens of percent [1, 2]. The boost converter is the member of that family that produces an output above its input, and it is also the member whose dynamics place the strongest limits on feedback design. Raising the duty cycle keeps the transistor closed for a longer share of each switching period, which is exactly the interval during which the inductor is disconnected from the output and the capacitor supplies the load alone. The immediate consequence of a duty increase is therefore a reduction in delivered charge, and the output falls. Only after the inductor current has grown does the larger current through a shorter delivery window overtake the loss, and the output climbs to a value above where it started.

That behavior is inverse response, and in transfer function terms it is a zero in the right half plane. A right-half-plane zero cannot be canceled by a controller, because canceling it would require an unstable pole in the controller. It cannot be moved by tuning, because it belongs to the power stage. The zero bounds the achievable crossover frequency and imposes a trade between the speed of a response and the size of its initial excursion in the wrong direction, and the classical sensitivity integrals make that trade quantitative rather than qualitative [3–5]. The same response shape appears in a boiler drum whose level rises when cold feedwater is added and in a distillation column whose bottom composition initially moves away from its final value after a reflux increase [6, 7]. The drum of Volume 18, Chapter 4 has a right-half-plane zero that, like the one of this chapter, moves with load.

Textbook treatments of inverse response usually begin from a second-order transfer function with a numerator zero placed by hand, followed by a step response that dips [6]. That presentation ties the phenomenon to an equation rather than to a mechanism. The physical statement is that a fixed quantity of charge per switching period is reallocated between two destinations, and the reallocation is instantaneous while its benefit is delayed. However, a simulation in which the zero is a coefficient supplies no mechanism to examine, and a hardware converter fast enough to be realistic reverses direction in tens of microseconds, well below what an oscilloscope laboratory in a first control course resolves.

This chapter works with a boost converter simulator, built on the framework of Section 1.6, whose right-half-plane zero is produced by the averaged circuit equations rather than inserted by hand. It quantifies the bandwidth cost of the zero and the effect of an inner current loop. The averaged model and current-mode control are both settled engineering. What the simulator adds is a single power

stage on which a single voltage controller and a voltage-to-current cascade share one actuator, one sample period and one set of limits, so the two architectures differ in loop structure and in nothing else. An explicit model predictive controller (MPC) on the same actuator, sample period and limits is a third structure, so the comparison can separate what a hard constraint on the inductor current and a model that contains the zero add to the cascade from what the zero still takes. The scenarios separate what the zero fixes from what the tuning chooses, including one that moves the zero by changing the inductor and one that deepens the dip by changing the capacitor while leaving the zero exactly where it was. Together they cover inverse response, the bandwidth limit a right-half-plane zero imposes, cascade structure, feedforward, anti-windup, actuator saturation and constrained predictive control, principles treated in the process control textbooks [5–7]. The model rests on two conservation statements, a voltage balance on the inductor and a charge balance on the capacitor, and the cascade rests on the time-scale separation between the two storage elements.

> ✦ DID YOU KNOW? **Ten Times the Frequency, a Tenth of the Inductor**
>
> Inductance and capacitance scale with the switching frequency. A national technology assessment of 2015 states that a tenfold increase in frequency gives a tenfold decrease in the required capacitance and inductance [8]. Section 9.4 shows that a smaller inductor also moves the right-half-plane zero farther from the origin.

Section 9.2 traces averaged models and current-mode control. Section 9.3 states the equations, the parameters and the numerics closely enough that the power stage can be reconstructed from the text. Section 9.4 derives the right-half-plane zero and its dependence on load, and Section 9.5 describes the cascade, the single loop, the feedforward and the explicit MPC. Section 9.6 measures the fourteen scenarios, Section 9.7 describes the hardware that carries the cascade, Section 9.8 states the limitations, Section 9.9 discusses the design trade the zero sets, Section 9.10 concludes, and the exercises of Section 9.11 close the chapter.

9.2 Background

A switching converter has no single set of differential equations. During each period the topology changes at least once, so the circuit is described by one linear system while the transistor conducts and by a different one while the diode conducts. Averaging over one switching period answered that difficulty, as the history segment below tells. The result is a continuous model that is accurate well below the switching frequency and that carries the operating point inside its coefficients. Later work extended averaging to quantities that are not slowly varying, by carrying complex Fourier coefficients of the state rather than its period average, which is what resonant converters require [9].

> ⧖ HISTORY **Averaging the Switch Away, 1973 to 1976**
>
> A converter's circuit changes every time its transistor opens or closes, and by 1976 two ways of modeling it had grown up side by side. One wrote the exact state equations of each switched circuit and relied on numerical computation. The other averaged by manipulating equivalent circuits, which gave the designer physical insight [10]. A 1973 paper had characterized the low-frequency behavior of switched dc-dc converters [11]. A 1976 conference paper joined the two traditions: average the two exact state-space descriptions over one switching cycle, then perturb and linearize. Its canonical circuit model predicted that switching puts zeros as well as poles into the response from duty to output, and it pointed to a real zero in the right half-plane of the buck-boost converter that would need proper compensation. Averaged models of this kind agreed with measured responses up to close to half the switching frequency [10].

The small-signal consequences of averaging a boost converter were catalogued soon afterwards. A canonical circuit model collects the duty-cycle dependence into controlled sources and shows that the boost and buck-boost topologies both acquire a numerator zero in the right half plane while the buck

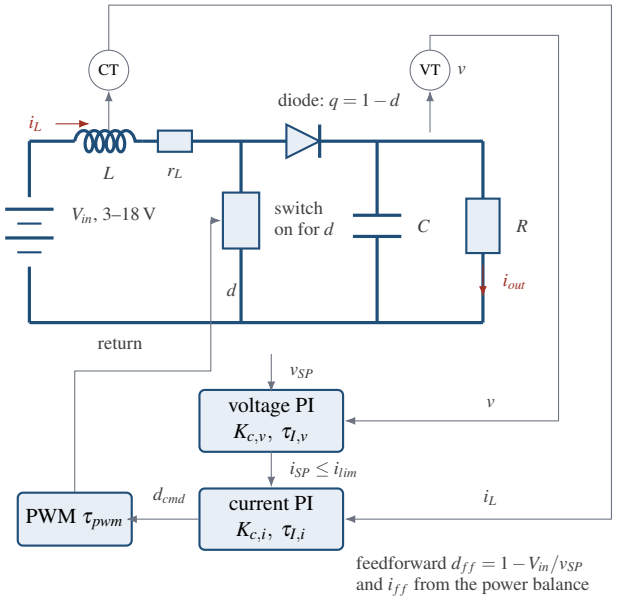


Figure 9.1 Boost power stage and the cascade that drives it. The transistor shorts the switching node for the fraction d of each switching period and the diode delivers inductor charge to C and R for the complement $q = 1 - d$. A current transformer reports i_L to the inner loop and a voltage transmitter reports v to the outer loop, whose output is the inductor-current setpoint i_{SP}. The commanded duty is clamped to $[0, d_{max}]$ and the current setpoint to $[0, i_{lim}]$, and the load current is $i_{out} = v/R$.

does not [12]. An alternative construction replaces the transistor and diode pair with a three-terminal averaged element, which allows the same zero to be read from a small-signal circuit without repeating the matrix algebra for every topology [13]. Modern treatments derive the control-to-output transfer function of the boost stage directly and give the zero as a function of load resistance, duty cycle and inductance [14].

Current-mode control developed as the practical remedy. Regulating the inductor current with a fast inner loop and letting a slower voltage loop set the current reference removes the right-half-plane zero from the loop that has to be tuned, because duty cycle to inductor current is minimum phase. Small-signal models of the arrangement followed its adoption in practice. A first small-signal treatment of current-programmed buck and boost regulators established the low-frequency behavior and the effect of the ramp added for stability [15], and a continuous-time model that also predicts the high-frequency sampling behavior of the current loop followed [16]. The engineering conclusion is long established, that the remedy for a right-half-plane zero is a different loop structure rather than a higher gain, and the simulator of this chapter makes that conclusion measurable. Section 1.4.5 gives the inner-loop speed rule that the same structure obeys on a process plant.

9.3 Converter Model

The simulated hardware is a single boost stage. Figure 9.1 places each symbol of the equations that follow onto the element it belongs to. A stiff source V_{in} drives an inductor L with series resistance r_L. The switching node is pulled to the negative rail by a transistor for a fraction d of each period and released for the complement $q = 1 - d$, during which the diode conducts and the inductor delivers charge to the output capacitor C and the load resistance R. Both states are instrumented, the inductor current i_L and the output voltage v, and both the source voltage and the load current are available to the feedforward terms.

Table 9.1 Variable Classification

Symbol	Meaning	Unit	Range
Controlled variables			
v	Output voltage, outer loop CV	V	13–48
i_L	Inductor current, inner loop CV	A	0–9
Setpoints			
v_{SP}	Output voltage setpoint	V	13–48
i_{SP}	Current setpoint, outer loop MV	A	0–9
Manipulated variable			
d	Duty cycle, inner loop MV	—	0–0.85
Disturbances			
V_{in}	Source voltage	V	3–18
R	Load resistance	Ω	4–48
States			
i_L	Inductor current	A	≥ 0
v	Capacitor voltage	V	≥ 0
d	Applied duty, lagged command	—	0–0.85
Derived outputs			
q	Delivery fraction $1-d$	—	0.15–1
i_{out}	Load current v/R	A	—
d_{ff}	Duty feedforward	—	0–0.85
i_{ff}	Current feedforward	A	0–9
η	Instantaneous P_{out}/P_{in}	%	—
E_L, E_C	Stored energy in L and in C	J	—
z	Right-half-plane zero	rad/s	—
ω_n	Natural frequency of the LC pair	rad/s	—

9.3.1 Variables and Signal Classification

The model quantities are listed in Table 9.1, separated into what is controlled, what is manipulated, what is disturbed and what is integrated. The classification is stated because the cascade assigns one signal three roles. The inductor current is a state of the power stage, the controlled variable of the inner loop, and the manipulated variable of the outer loop.

9.3.2 Averaged State Equations

Nothing in the simulator resolves an individual switching cycle. Both circuit equations are averaged over one period, with the transistor conducting for the fraction d and the diode for the complement $q = 1 - d$, which is the standard continuous-conduction averaged model [10, 14]. The inductor sees the full source voltage while the switch is closed and must also overcome the output voltage while the diode conducts, so the average opposing voltage is $q\,v$, as shown in (9.1).

$$L\frac{\mathrm{d}i_L}{\mathrm{d}t} = V_{in} - r_L\,i_L - q\,v \tag{9.1}$$

The capacitor is charged only during the diode interval and is discharged by the load at all times, as shown in (9.2).

$$C\frac{\mathrm{d}v}{\mathrm{d}t} = q\,i_L - \frac{v}{R} \tag{9.2}$$

The applied duty follows its command through a single lag that stands for the digital pulse-width modulator update and the gate driver, as shown in (9.3).

$$\tau_{pwm}\frac{\mathrm{d}d}{\mathrm{d}t} = d_{cmd} - d \tag{9.3}$$

where L is the inductance in H, C the output capacitance in F, r_L the inductor series resistance in Ω, R the load resistance in Ω, V_{in} the source voltage in V, τ_{pwm} the modulator lag in seconds and d_{cmd} the duty the controller asks for. Every term in (9.1) is a voltage and every term in (9.2) is a current, so both balance dimensionally.

The inverse response originates in the product $q\,i_L$ in (9.2). A step increase in d reduces q immediately while i_L is a state and cannot move at all in the same instant, so the delivered current $q\,i_L$ drops by exactly $i_L\,\Delta d$ and the capacitor is starved. At the default operating point a duty step of $\Delta d = 0.10$ gives an initial slope of $-i_L\,\Delta d/C = -207\,\mathrm{V/s}$. The recovery has to wait for i_L to build through (9.1), which takes on the order of $L\,i_L/(V_{in} - r_L i_L) = 7.2\,\mathrm{ms}$.

Pause. Try Exercise 9.1 before reading on.

A useful accounting of the same event is the energy balance obtained by multiplying (9.1) by i_L and (9.2) by v and adding, as shown in (9.4).

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\tfrac{1}{2}L i_L^2 + \tfrac{1}{2}C v^2\right) = i_L\,(V_{in} - r_L i_L) - \frac{v^2}{R} \tag{9.4}$$

The switch appears nowhere in (9.4), because an ideal switch stores and dissipates nothing and only decides which reservoir is being filled. During the initial dip the capacitor energy falls while the inductor energy rises faster, so the total stored energy is already increasing while the output voltage is still decreasing. The charge withheld from the output is stored in the inductor as current.

9.3.3 Steady State and the Effect of Inductor Resistance

Setting both derivatives to zero in (9.1) and (9.2) gives $i_L = v/(qR)$ and the steady-state output in (9.5).

$$v = \frac{V_{in}\,q\,R}{q^2 R + r_L}, \qquad q = 1 - d \tag{9.5}$$

With $r_L = 0$ this collapses to the familiar ideal relation $v = V_{in}/(1-d)$, and the application never uses that relation as a model. At the default point the difference matters. A duty of exactly 0.5 gives 23.226 V from (9.5) rather than the ideal 24.000 V, 0.774 V short, and the duty that actually holds 24.000 V is 0.517263. The inductor drops $r_L i_L = 0.414\,\mathrm{V}$, which the duty must make up.

Inverting (9.5) for a target voltage requires the larger root of $Rq^2 - (V_{in}R/v)\,q + r_L = 0$, as shown in (9.6).

$$q = \frac{1}{2R}\left(\frac{V_{in}R}{v} + \sqrt{\left(\frac{V_{in}R}{v}\right)^2 - 4R\,r_L}\right) \tag{9.6}$$

The smaller root is the high-current branch beyond the peak of the voltage-versus-duty curve and is not the physical operating point of a regulator. A negative discriminant means the target voltage is unreachable at any duty, in which case the initialization falls back on $q = 1 - d_{max}$. Equation (9.6) places every automatic scenario at an exact equilibrium at $t = 0$, so the only transient in a run is the one the scenario introduces.

9.3.4 Parameters, Limits and Numerics

Table 9.2 collects the seven parameters, their default values and their adjustable intervals. Three of them are hardware limits rather than tuning choices and are fixed, so the saturation statistics and the displayed limits derive from one set of declared values.

Three hard events are handled once per step by a post-step routine and never by the right-hand side itself, which keeps every intermediate integration stage on a smooth vector field. The applied duty is clamped to $[0, d_{max}]$. The output voltage is clamped at zero from below. If the averaged inductor current reaches zero the converter has left continuous conduction, the state is clamped at zero and a fault is annunciated that clears only once the current exceeds 0.05 A again, because the averaged

Table 9.2 Model Parameters and Their Adjustable Ranges

Symbol	Meaning	Value	Unit	Range
L	Inductance	20	mH	2–60
C	Output capacitance	2.0	mF	0.2–6.0
r_L	Inductor resistance	0.10	Ω	0–0.60
τ_{pwm}	PWM and driver lag	0.2	ms	0–2.0
d_{max}	Maximum duty (hidden)	0.85	—	fixed
i_{lim}	Current setpoint limit (hidden)	9.0	A	fixed
i_{trip}	Overcurrent trip (hidden)	12.0	A	fixed
Default operating point, $V_{in} = 12\,\text{V}$, $R = 12\,\Omega$				
v	Output voltage	24.000	V	—
d	Duty cycle	0.517263	—	—
i_L	Inductor current	4.1430	A	—
P_{in}	Input power	49.72	W	—
P_{out}	Output power	48.00	W	—
η	Efficiency	96.55	%	—
E_L	Inductor energy	171.6	mJ	—
E_C	Capacitor energy	576.0	mJ	—
z	Right-half-plane zero	134.8	rad/s	—
ω_n	*LC* natural frequency	77.68	rad/s	—
ζ	*LC* damping ratio	0.300	—	—

model no longer describes the circuit past that boundary. A separate overcurrent flag with hysteresis annunciates when i_L exceeds $i_{trip} = 12\,\text{A}$ and clears below 10.8 A.

Three clocks run at once, as in every framework app (Section 1.6.1). A fixed-step fourth-order Runge-Kutta integrator advances the states every $\Delta t = 0.1\,\text{ms}$, the two loops execute on a $\Delta t_c = 0.2\,\text{ms}$ cycle that stands for one period of a 5 kHz digital controller, and 500 samples of every logged quantity are recorded per simulated second. The logging rate is set by the phenomenon, because the open-loop inverse response bottoms out 7.5 ms after a duty step and is over by 15.8 ms, and a 5 mH inductor compresses the closed-loop dip into 2.5 ms. Each run exports as comma-separated data, including both states, the applied and commanded duty, both feedforward terms, the operating-point zero and natural frequency, the stored energy in each element, the saturation and fault flags, all seven parameters and both disturbance channels.

9.3.5 Assumptions

The model is switching-averaged and assumes continuous conduction throughout. It says nothing about output ripple, nothing about subharmonic oscillation, and nothing about the peak-current-mode instability that appears above 50 % duty in hardware and that a compensating ramp is added to suppress [16]. Averaging is accurate for behavior well below the switching frequency and breaks down as the closed-loop bandwidth approaches it. Nothing in the equations resolves an individual switching cycle.

The switch and the diode are ideal, with no forward drop, no on-resistance, no reverse-recovery charge and no dead time. A real diode loses roughly a volt, which matters most at low output voltage and would depress the achievable boost ratio in the brownout scenario more than the model shows. The only dissipation is the inductor series resistance r_L. Capacitor equivalent series resistance, core loss and switching loss are absent, so the reported efficiency is an upper bound and the step in output voltage that equivalent series resistance produces at the instant of a load change is missing entirely.

The load is a fixed resistor and the source is ideal. A constant-power load such as a downstream regulator presents a negative incremental resistance and would move the closed-loop poles, and a source with series inductance would add a resonance this model cannot show. Sensing is perfect and instantaneous, with no current-sense filter, no divider dynamics, no quantization and no noise, so the

derivative-free tuning used here is not challenged by measurement noise the way a real current loop is. The only actuator dynamic is the first-order modulator lag of 0.2 ms.

The parameter values are exaggerated on purpose. A 5 kHz converter with a 20 mH inductor and a 2 mF capacitor is slow by the standards of commercial designs, and no product would be built with an inverse response that lasts 16 ms. The exaggeration makes the phenomenon legible on a trend display at reduced playback speed. Every qualitative statement of Section 9.4 survives the scaling, because (9.9) is homogeneous in the component values.

9.4 Origin and Load Dependence of the Right-Half-Plane Zero

Linearizing (9.1) and (9.2) about an operating point (i_0, v_0, q_0) and writing δ for a deviation gives (9.7).

$$\begin{aligned} L\,\frac{\mathrm{d}\delta i}{\mathrm{d}t} &= -r_L\,\delta i - q_0\,\delta v + v_0\,\delta d \\ C\,\frac{\mathrm{d}\delta v}{\mathrm{d}t} &= \quad q_0\,\delta i - \frac{\delta v}{R} - i_0\,\delta d \end{aligned} \tag{9.7}$$

The two duty terms carry opposite signs, and that sign difference is the origin of the inverse response. Additional duty raises the average inductor voltage, which acts through the inductor dynamics, and simultaneously removes $i_0\,\delta d$ of delivered current, which acts immediately. Eliminating δi gives the control-to-output transfer function in (9.8).

$$\frac{\delta v(s)}{\delta d(s)} = \frac{(q_0 v_0 - r_L i_0)\,(1 - s/z)}{LC\,s^2 + \left(\dfrac{L}{R} + r_L C\right) s + \left(\dfrac{r_L}{R} + q_0^2\right)} \tag{9.8}$$

where the numerator zero is given by (9.9).

$$z = \frac{q_0 v_0 - r_L i_0}{L\,i_0} \;\approx\; \frac{q_0^2 R}{L} \;>\; 0 \tag{9.9}$$

The zero is positive for every physically reachable operating point, because $q_0 v_0$ is the average voltage the inductor pushes against and $r_L i_0$ is a small fraction of it. At the default point (9.9) evaluates to $+134.8\,\mathrm{rad/s}$ against the approximation $q_0^2 R/L = 139.8\,\mathrm{rad/s}$, and the denominator of (9.8) gives $\omega_n = 77.68\,\mathrm{rad/s}$ with $\zeta = 0.300$. The steady-state gain of (9.8) is 46.3 V per unit of duty, so half a percent more duty is worth 0.23 V of output.

Two features of (9.9) are tested by separate scenarios. The capacitance does not appear in the zero at all, so a smaller capacitor deepens the initial dip through the $1/C$ in (9.2) and leaves the zero exactly where it was. The inductance appears in the denominator, so a smaller inductor moves the zero away from the origin and relaxes the bandwidth limit. The load appears through $q_0^2 R$, which means the zero moves toward the origin as the converter is loaded more heavily, and the operating point with the most restrictive bandwidth limit is the one at which the converter delivers the most power.

🕮 **Pause.** Try Exercise 9.2 before reading on.

Figure 9.2 plots (9.9) and the natural frequency of (9.8) across the load range the application permits. Between 48 Ω and 4 Ω the zero sweeps from 585 rad/s down to 34.4 rad/s, a factor of 17, while ω_n moves only from 78.7 to 74.5 rad/s. The two curves cross at a load of 7.36 Ω. Above that load the zero lies below the resonance, the most restrictive arrangement, because the crossover must then be placed below a resonant peak the loop cannot attenuate. The default 12 Ω load lies on the favorable side of that crossing, with $z = 1.74\,\omega_n$.

The bandwidth consequence is quantitative. For a stable plant with a single right-half-plane zero at z the Poisson integral forces the sensitivity function to exceed unity over some frequency band whenever it is pushed below unity elsewhere, and the area of that penalty is weighted toward z [3, 5].

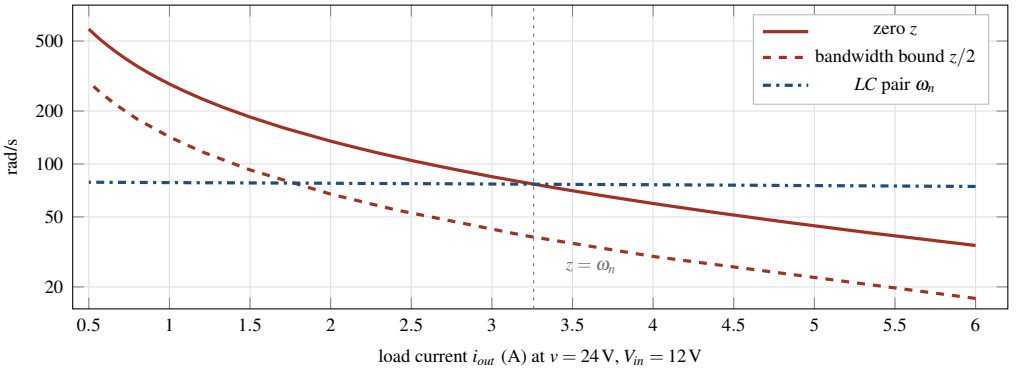


Figure 9.2 Right-half-plane zero (solid) and LC natural frequency (dash-dotted) against load current, with the bandwidth bound $z/2$ (dashed), computed from (9.5) and (9.9) over the load range the application allows. The dotted vertical line marks $z = \omega_n$. The zero sweeps by a factor of 17 between 48 Ω and 4 Ω while ω_n barely moves, and the two cross at $i_{out} = 3.26\,\mathrm{A}$, which is a 7.36 Ω load.

The practical rule that follows is that the crossover frequency of a loop closed around such a plant should stay comfortably below z, and $z/2$ is the usual working bound. Figure 9.2 draws that bound. At the default point it is $67\,\mathrm{rad/s} = 10.7\,\mathrm{Hz}$, which is already below the LC natural frequency, and at a 4 Ω load it falls to 17.2 rad/s. No choice of K_c, τ_I or controller structure changes those numbers, because they are properties of L, C, R and the operating duty.

The reason a cascade helps is that the inner loop closes around a different transfer function. Eliminating δv from (9.7) instead gives the control-to-current response in (9.10).

$$\frac{\delta i(s)}{\delta d(s)} = \frac{v_0\,(Cs + 1/R) + q_0 i_0}{LC\,s^2 + \left(\dfrac{L}{R} + r_L C\right) s + \left(\dfrac{r_L}{R} + q_0^2\right)} \tag{9.10}$$

Every coefficient in the numerator of (9.10) is positive, so the zero of the current response sits at $-(v_0/R + q_0 i_0)/(C v_0)$, in the left half plane. The physical statement is the one (9.1) makes directly, that a longer on-time raises $\mathrm{d}i_L/\mathrm{d}t$ the instant q falls with nothing competing to reverse the sign. The inner loop can therefore be closed as fast as the modulator and the sampling allow. With the current effectively an input, the output node obeys the power balance in (9.11), obtained from (9.4) with the inductor term removed.

$$C v \frac{\mathrm{d}v}{\mathrm{d}t} = i_L\,(V_{in} - r_L i_L) - \frac{v^2}{R} \tag{9.11}$$

Equation (9.11) is first order and its right-hand side is increasing in i_L for all $i_L < V_{in}/2r_L$, which at the default supply is 60 A and therefore never binding. The plant seen by the outer loop retains no inverse response.

9.5 Controller Design

Both loops are proportional-integral with the derivative time set to zero, in the velocity form of Section 1.6.2, executing every $\Delta t_c = 0.2\,\mathrm{ms}$. The outer loop converts volts of error into amps of

current demand, as shown in (9.12).

$$\begin{aligned} e_{v,k} &= v_{SP,k} - v_k \\ \Delta u^I_{v,k} &= \frac{K_{c,v}}{\tau_{I,v}} e_{v,k} \Delta t_c \\ s_{v,k} &= s_{v,k-1} + K_{c,v}\left(e_{v,k} - e_{v,k-1}\right) \\ c_{v,k} &= i_{ff,k} + s_{v,k} + I_{v,k-1} \\ i_{SP,k} &= \text{clamp}\left(c_{v,k}, 0, i_{lim}\right) \\ I_{v,k} &= I_{v,k-1} + \Delta u^I_{v,k} + \frac{\Delta t_c}{\tau_{I,v}}\left(i_{SP,k} - c_{v,k}\right) \end{aligned} \tag{9.12}$$

The inner loop converts amps of error into duty, as shown in (9.13).

$$\begin{aligned} e_{i,k} &= i_{SP,k} - i_{L,k} \\ \Delta u^I_{i,k} &= \frac{K_{c,i}}{\tau_{I,i}} e_{i,k} \Delta t_c \\ s_{i,k} &= s_{i,k-1} + K_{c,i}\left(e_{i,k} - e_{i,k-1}\right) \\ c_{i,k} &= d_{ff,k} + s_{i,k} + I_{i,k-1} \\ d_{cmd,k} &= \text{clamp}\left(c_{i,k}, 0, d_{max}\right) \\ I_{i,k} &= I_{i,k-1} + \Delta u^I_{i,k} + \frac{\Delta t_c}{\tau_{I,i}}\left(d_{cmd,k} - c_{i,k}\right) \end{aligned} \tag{9.13}$$

where I is the integral accumulator of each loop, s accumulates the proportional increments without limit and c is the output the loop demands before clipping. In the symbols of Equations (1.9) to (1.12), $s = S$, $c = v$, the feedforward term is the bias u_b and the tracking time is $T_t = \tau_I$. Anti-windup is back-calculation, so while an output is clipped the final term of each update moves the integral by $\Delta t_c/\tau_I$ of the clipped excess (1/250 per scan for the outer loop at $\tau_{I,v} = 0.05\,\text{s}$ and 1/40 for the inner loop at $\tau_{I,i} = 8\,\text{ms}$) [17]. With the tracking time equal to τ_I the error terms cancel in the update of a clipped loop, so its integral settles at the limit less the feedforward bias, $I_v = i_{lim} - i_{ff}$ for an outer loop held at i_{lim}. At a limit the tracking time shapes the release, as Section 9.6.5 shows. Both loops are reverse acting in the ISA sense, because a low output voltage calls for more current and a low current calls for more duty. Neither loop has a derivative term, and neither output is slew limited beyond what τ_{pwm} imposes on the actuator.

The alternative architecture replaces (9.12) and (9.13) by a single proportional-integral controller of identical form with gain $K_{c,dv}$ and integral time $\tau_{I,dv}$, taking e_v and producing d_{cmd} clamped to $[0, d_{max}]$. In that configuration nothing measures the inductor current for control purposes and nothing enforces i_{lim}. The inductor current and its feedforward reference remain recorded together, so the size of the unregulated current excursion is measurable.

Both feedforward terms come from the measured input voltage and load current, with the declared $r_L = 0.1\,\Omega$ in the current term. The duty feedforward is the ideal duty for the requested boost ratio, and the current feedforward is the inductor current the power balance $V_{in} i_L = v_{SP}\, i_{out} + r_L i_L^2$ demands at the present load, as shown in (9.14).

$$\begin{aligned} d_{ff} &= \text{clamp}\left(1 - \frac{V_{in}}{v_{SP}}, 0, d_{max}\right) \\ i_{ff} &= \text{clamp}\left(\frac{V_{in} - \sqrt{V_{in}^2 - 4 r_L v_{SP}\, i_{out}}}{2 r_L}, 0, i_{lim}\right) \end{aligned} \tag{9.14}$$

Table 9.3 Controller Settings and Their Adjustable Ranges

Symbol	Loop	Unit	Default	Range
$K_{c,v}$	Voltage, outer	A/V	0.12	0–1.00
$\tau_{I,v}$	Voltage, outer	s	0.05	0.01–3
$K_{c,i}$	Current, inner	%/A	50	0–200
$\tau_{I,i}$	Current, inner	ms	8	1–100
$K_{c,dv}$	Voltage only	%/V	0.1	0–3
$\tau_{I,dv}$	Voltage only	s	0.20	0.01–3
v_{SP}	Setpoint	V	24.0	13–48

The smaller root is the operating one. A negative discriminant means the demanded power cannot be pushed through r_L at any current, and i_{ff} is set to i_{lim} in that case, the largest deliverable request. The r_L in (9.14), in the integrator preload and in the bias shift described below is the declared value, because no instrument measures the inductor resistance, so a resistance changed in the simulated hardware reaches the cascade only through the measured current and voltage. Both terms enter as the bias of their loop, so the integral accumulators carry only what the feedforward misses. At the default point $d_{ff} = 0.5000$ against a true steady-state duty of 0.517263, so the inner integrator holds 0.017263 permanently and that offset is precisely the inductor drop. Both terms are static feedforward computed from a balance, as in Section 1.4.3. A dynamic compensator built as in Equation (1.4) around the duty-to-voltage path would have to invert its right-half-plane zero, which takes an unstable pole.

Three mechanisms keep transitions free of actuator jumps. In MANUAL the operator holds the duty and all three controllers track it, as Section 1.6.2 describes. In AUTO the controller that is not in use tracks the applied duty every step, so switching architecture mid-run moves nothing. Switching the feedforward on or off would step every bias by the whole feedforward amount, so each integral is shifted by the same quantity in the opposite direction at the instant of the switch, which makes the comparison in the load scenario a comparison of two responses rather than of two different startup transients. The manual duty command tracks the controller output while in AUTO, so the reverse transition is bumpless as well.

Default tuning is listed in Table 9.3. The inner gain places the current loop crossover near $K_{c,i}\, v_0/L = 600\,\mathrm{rad/s}$, using the high-frequency approximation $\delta i/\delta d \approx v_0/(Ls)$ from (9.10), which is about four times the zero and far faster than the outer loop, whose recovery to within 0.5 V of a 6 V setpoint change takes 196 ms. That separation is the inner-loop speed rule of Section 1.4.5, which Section 9.6.4 tests on this plant.

9.5.1 Explicit Model Predictive Control

A third architecture moves the duty directly in automatic operation. It is a model predictive controller computed in advance rather than at run time, where the MPC option of Section 1.6.3 solves its program online at every decision. A quadratic program whose cost gradient and constraint bounds vary linearly with a parameter vector has an optimal solution that is a continuous piecewise affine function of that vector, defined over a finite set of polyhedra, so the constrained linear quadratic regulator (LQR) reduces to a stored list of polyhedra, each paired with its own affine control law, and a decision becomes a point-location search [18]. Predictive control of the boost stage has also been formulated online, by enumerating switch sequences for direct voltage control [19]. The cascade remains the default architecture, and only scenario 13 selects the MPC.

HISTORY Predictive Control Reaches the Converter, 1980s to 2000s

Model predictive control emerged in the 1960s and over the next decade became an established strategy for constrained plants, mostly in the process industry. The power electronics community began to investigate it in the 1980s, but the meager computing resources of the time, together with new power semiconductors that allowed higher switching frequencies, limited its use. After a hiatus of about two decades, more powerful microprocessors revived the interest in the 2000s [20]. One route that followed solves the optimization offline for all possible states and stores the resulting control law in look-up tables, the explicit form this chapter's MPC takes. The same review warns that such tables suit only small problems, because their memory grows exponentially with the size of the problem [20].

The prediction model is (9.7) with the modulator lag (9.3) appended, linearized at the declared default point of Table 9.2 ($L = 20\,\text{mH}$, $C = 2\,\text{mF}$, $r_L = 0.1\,\Omega$, $\tau_{pwm} = 0.2\,\text{ms}$, $V_{in} = 12\,\text{V}$, $R = 12\,\Omega$, $v_0 = 24\,\text{V}$, $d_0 = 0.517263$, $i_0 = 4.1430\,\text{A}$). Its state $x = (i_L - i_{ss},\, v - v_{ss},\, d - d_{ss})$ and its input $\delta u = d_{cmd} - d_{ss}$ are deviations from a target (i_{ss}, v_{ss}, d_{ss}), and an exact zero-order hold over $T_s = \Delta t_c = 0.2\,\text{ms}$ gives $x_{k+1} = \Phi x_k + \Gamma\,\delta u_k$. No step test was used to identify it. Equation (9.7) is already that linearization, and a finite-difference Jacobian taken from the simulator's right-hand side matches its coefficients within a relative 1.3×10^{-9}. Its open-loop eigenvalues are $-23.3 \pm 74.1\text{j}$ and −5000/s, and duty to voltage keeps the zero of (9.9) at +134.8 rad/s, so the prediction contains the inverse response. The same matrices serve every setpoint.

Every 0.2 ms the controller plans three duty moves, one per period, and applies the first, as shown in (9.15).

$$\begin{aligned} \min_{\delta u_0, \delta u_1, \delta u_2} \; & \sum_{k=1}^{2} Q_v x_{2,k}^2 + x_3^T P x_3 + R \sum_{k=0}^{2} \delta u_k^2 \\ \text{subject to } & x_{k+1} = \Phi x_k + \Gamma\,\delta u_k, \quad x_0 = \hat{x} \\ & 0 \le d_{ss} + \delta u_k \le d_{max} \\ & i_{floor} \le i_{ss} + x_{1,k+1} \le i_{lim} \end{aligned} \tag{9.15}$$

where $x_{1,k}$ and $x_{2,k}$ are the predicted current and voltage deviations, $\hat{x}$ is the estimated state measured from the target, and P solves the discrete algebraic Riccati equation for Φ, Γ, $\text{diag}(0, Q_v, 0)$ and R. Charging the Riccati cost-to-go after the last move is the usual route to closed-loop stability in receding-horizon control [21], and no terminal set is imposed here. Volume 7, Chapter 7 shows a predictive controller stripped of that terminal cost becoming unstable at short horizons. Bryson's rule on the 0.5 V band of the voltage loop and on the duty headroom $d_{max} - d_0 = 0.3327$ gives $Q_v = 4/\text{V}^2$ and $R = 9.032$, with no weight on the current and no move penalty. With no row active the first move is the LQR's, $\delta u = -(1.2688/\text{A}\,x_1 + 0.10381/\text{V}\,x_2 + 0.24164\,x_3)$, which places the closed-loop poles at −135.9, −1434 and −4772/s. The slowest of them lies near the mirror image of the zero, at $-1.008\,z$.

All twelve rows of (9.15) are hard. The duty range and i_{lim} are read as the cascade reads them, but the current rows bound the predicted inductor current itself, where the cascade clamps only the setpoint of its inner loop. The floor $i_{floor} = \min(0.3\,\text{A}, i_{ss}/2)$ keeps the plan out of discontinuous conduction. With a floor of zero a step from 24 to 13 V reaches zero current, because the linear model's one-period current prediction errs by up to 125 mA that far from 24 V. Unlike a box on the input alone, the current rows can make the program infeasible, for example when the current already sits at the limit and the three moves cannot keep the prediction below it.

The target and the limits enter only the right-hand sides, through the parameter $\theta = (\hat{x}, -d_{ss}, d_{max} - d_{ss}, i_{lim} - i_{ss}, i_{floor} - i_{ss}, 1)$, so one table serves every setpoint, load, supply and limit. Writing (9.15) as $\min \frac{1}{2} z^T G z + z^T F \hat{x}$ subject to $C_q z \ge b(\theta)$ with z the three moves, an active set $\mathscr{A}$ gives the moves in (9.16).

$$\begin{aligned} z_u &= -G^{-1} F \hat{x}, \qquad \lambda_{\mathscr{A}} = W_{\mathscr{A}} \left(b_{\mathscr{A}}(\theta) - C_{\mathscr{A}} z_u \right) \\ z &= z_u + G^{-1} C_{\mathscr{A}}^T \lambda_{\mathscr{A}}, \qquad W_{\mathscr{A}} = \left(C_{\mathscr{A}} G^{-1} C_{\mathscr{A}}^T \right)^{-1} \end{aligned} \tag{9.16}$$

The law holds on the critical region where $\lambda_{\mathscr{A}} \geq 0$ and every row of $C_q z \geq b(\theta)$ is met, so a region that accepts θ returns the exact optimum. Every candidate set of up to three rows was tested, and a set enters the table only if its region contains a ball of nonzero radius somewhere in a stated domain: setpoints 13 to 48 V, supplies 3 to 18 V, loads 4 to 48 Ω, duty ceilings 0.5 to 0.95, current limits 3 to 20 A, outputs within 30 V of the target and currents from zero to 3 A above the limit. The table holds 109 regions. An independent enumeration that ties each target coordinate to its physical relation (a target current never above the limit, a floor never above the target current, a target duty inside its range) finds 85 reachable regions, all of them in the table. The other 24 exist only where the box lets a target coordinate pass its limit, and none was chosen in 359 976 closed-loop decisions. Outputs up to 60 V from the target would add six regions that are not stored.

Each decision scans the stored regions in a fixed order that begins with the unconstrained law. Over 135 000 test states, 15 000 of them within 10^{-6} of a region boundary and 105 000 taken from closed-loop runs of every scenario, the table's duty differs by at most 8.8×10^{-11} from that of the framework's dual active-set solver applied online to the identical program, and both agree with an independent solution of the optimality conditions to 3.2×10^{-9}. Of those states 4408 have no feasible plan, and the lookup misses every one of them. In Node.js a lookup averages 0.18 µs over 209 986 closed-loop states and 1.86 µs in the slowest region, against a median of 6.0 µs for the online solve. Played in Chrome by its animation loop, which held 60 frames per second in every case, the whole controller execution under the lookup takes 0.22 to 0.24 % of one core at the default playback speed of 0.1 times real time and 1.20 to 1.25 % at real time (5000 decisions per second), against 0.13 and 0.43 % for the cascade, whose execution includes the MPC's estimator. The online solve takes 1.06 to 1.11 % and 5.96 to 6.95 %. Every timing was taken on a host shared with other computations.

The MPC reads the measured inductor current, output voltage, input voltage and load current (the two signals the feedforward reads), the setpoint, its own commands, d_{max} and i_{lim}. A constant-gain extended Kalman filter estimates (i_L, v, d, w_C, w_L), where w_L is a voltage added to (9.1) and w_C a current drawn from (9.2), both random walks. It predicts with the declared nonlinear equations (9.1) to (9.3) at $R = 12\,\Omega$, integrated by a fourth-order Runge-Kutta method in two half-period substeps, and corrects with the steady Kalman gain of the design linearization augmented with w_C and w_L, designed for 0.01 A and 0.01 V of measurement noise and random walks of 0.05 A and 0.05 V per period, which leaves error poles at -792 ± 789j and -2548 ± 2446j 1/s. The target is the declared steady state at v_{SP} for the measured supply and the measured load resistance $\hat{R} = v / i_{out}$, with w_L passed through a first-order filter of 0.1 s to give $\bar{w}_L$, as shown in (9.17).

$$
\begin{aligned}
(V_{in} + \bar{w}_L)\, i_{ss} - r_L i_{ss}^2 &= \frac{v_{SP}^2}{\hat{R}}, \qquad v_{ss} = v_{SP} \\
d_{ss} &= 1 - \frac{v_{SP}}{\hat{R}\, i_{ss}}
\end{aligned}
\tag{9.17}
$$

The smaller root of the power balance is taken, as in (9.14). The filtered estimate shifts the target, the target-shift form of offset-free predictive control [22], so the output returns to v_{SP} with r_L or V_{in} in error. The offset-free MPC of Volume 9, Chapter 4 carries a fuel bias in its estimator to the same end. Where the setpoint cannot be held, in an overload past i_{lim} or a brownout past d_{max}, the duty and current targets become those of the nearest reachable steady state while v_{ss} stays at v_{SP}. An earlier design that estimated the load current instead of measuring it did not settle with $C = 0.5$ mF (16.82 to 45.68 V over 4 s of scenario 8), because the estimate absorbed the capacitor's model error, while the adopted design reaches the band in 18 ms on the same case.

When no stored region contains θ the decision is rejected as a miss. The duty then passes without a step to the cascade, which has tracked the applied duty at every execution (its outer loop toward the measured current), the historian logs the fallback, and the table regains the duty once 20 consecutive decisions have found a region, ramped in over another 20 decisions (4 ms). A miss forced in the middle of scenario 13's setpoint step moves the duty by 0.021 at the hand-over. With the current held at the limit in an overload the program has no solution, so a miss there leaves the cascade in charge

Table 9.4 Scenario Definitions

#	Scenario	Mode	Event
1	`steady`	MANUAL	none, $d = 0.517263$
2	`inverse`	MANUAL	d 0.5173→0.6173 at 0.2 s
3	`sp-step`	AUTO	v_{SP} 24→30 V at 0.3 s
4	`load-step`	AUTO	R 12→6→12 Ω
5	`vin-sag`	AUTO	V_{in} 12→8 V at 0.4 s
6	`duty-sat`	AUTO	$V_{in} \to 3.5$ V, $R = 24\,\Omega$
7	`current-limit`	AUTO	R 12→4→12 Ω
8	`small-cap`	AUTO	$C = 0.5$ mF, 24→28 V
9	`small-ind`	AUTO	$L = 5$ mH, 24→28 V
10	`voltage-only`	AUTO	single PI, 24→30 V
11	`cascade`	AUTO	cascade, 24→30 V
12	`aggressive`	AUTO	$K_{c,v} = 0.5$, $\tau_{I,v} = 0.02$ s
13	`explicit-mpc`	AUTO	MPC, v_{SP} 24→30 V, R 12→4→12 Ω
14	`challenge`	AUTO	R 12→7 Ω at 1 s

until the load is released. Forced at 1.2 s in scenario 13, it returns the duty to the MPC at 2.004 s, and meanwhile the cascade reaches 9.03 A where the MPC alone holds 8.84 A.

Three models stay separate. The simulated plant evaluates (9.1) to (9.3) with whatever component values the run carries, while the MPC's prediction and estimator use the declared values frozen when the application loads and never read the plant's. A changed L, C, r_L or τ_{pwm} is therefore a mismatch that the MPC absorbs through its disturbance estimate, much as the cascade absorbs a feedforward error through its integrators. The estimator executes under every architecture but feeds nothing to the proportional-integral loops, so each scenario under the cascade and the single loop reproduces its earlier run bit for bit on every previously logged column.

9.6 Results

Table 9.4 lists the fourteen scenarios the application supplies. Each scenario fixes only the quantities it requires and leaves everything else at the defaults. Summary metrics for all of them, with the shipped settings and from the scorecard of Section 1.6.4, are given in Table 9.5, read at the stated simulated time. Overshoot and settling are reported only after a setpoint change, so the disturbance cases carry a dash. In scenario 10 the single loop commands the duty directly, and in scenario 13 the MPC does. The displayed current reference is then not a commanded actuator, so its MV travel and saturation entries are dashes. The duty-cycle scorecard still reports actual actuator travel and saturation. If architectures are switched during a run, current-setpoint travel and saturation accumulate only while that setpoint commands the inner loop; saturation is normalized by that active time. Depths, peaks and crossing times quoted below are computed from the state at every 0.1 ms integration step, because the 2 ms log reads the bottom of a brief dip late and shallow (0.21 V against 0.50 V for the 5 mH inductor of scenario 9). Recovery is the first instant at which the output enters ±0.5 V of setpoint and stays inside it, which agrees with the settling column of Table 9.5 to within one 2 ms sample.

9.6.1 Open-Loop Inverse Response

Scenario 2 records the inverse response with no controller acting and is shown in Figure 9.3. The converter starts in exact equilibrium at 24.000 V with the duty held constant at 0.517263, and at $t = 0.2$ s a timed event steps the duty by ten percentage points to 0.617263. The new steady state is 29.67 V, so the eventual motion is upward by more than 5 V. The output falls 0.723 V first, bottoms out 7.5 ms after the step, crosses back through its starting value at 15.8 ms and reaches 31.46 V at the first peak of the underdamped recovery, 62 ms after the step. No tuning is implicated, because no controller is active. The output is recorded together with $V_{in}/(1-d)$, the steady-state voltage the applied duty requests, and the two signals move in opposite directions for the first 15.8 ms.

Table 9.5 Measured Voltage-Loop Performance at the Stated Simulated Time

#	t (s)	IAE (V s)	OS (%)	settling (s)	MV travel (A)	sat (%)	peak i_L (A)
1	1.03	0.000	—	—	0.00	0.0	4.14
2	0.24	0.078	—	—	3.26	0.0	7.40
3	1.08	0.305	14.8	0.194	3.49	0.0	7.04
4	4.80	0.556	—	—	13.26	1.0	9.00
5	1.97	0.208	—	—	4.04	0.0	6.84
6	2.00	7.267	—	—	11.76	74.2	6.05
7	4.78	8.376	—	—	14.01	41.6	9.47
8	1.02	0.123	10.7	0.016	1.94	0.0	5.90
9	1.04	0.115	8.8	0.028	1.90	0.0	5.88
10	1.95	0.557	30.5	0.134	—	—	7.59
11	2.00	0.305	14.8	0.194	3.49	0.0	7.04
12	2.00	0.218	177.9	0.056	8.92	0.5	9.50
13	3.00	10.319	0.0	1.706	—	—	8.85
14	3.00	0.179	—	—	5.04	0.0	7.58

TRY IT IN THE APP **2 · Inverse Response · Step the Duty by Hand**

Load the scenario and press Run. Watch the output voltage trend after the duty step at 0.2 s. The output falls first, bottoms out 7.5 ms after the step and crosses back through 24 V at 15.8 ms before it climbs toward its new steady state of 29.67 V. No controller is acting.

The stored-energy accounting makes the mechanism explicit. At the instant shown in Figure 9.3 the capacitor holds 865 mJ and the inductor 548 mJ, against 576 mJ and 172 mJ at the start. During the first 7.5 ms the capacitor energy is falling and the inductor energy is rising faster, so (9.4) is positive throughout an interval in which the measured output is negative-going. The reported zero location is $+71\,\mathrm{rad/s}$ at that instant rather than the $+135\,\mathrm{rad/s}$ reported at rest, because (9.9) moves with the operating point and the converter is delivering more current.

9.6.2 Undershoot Against Recovery Time

Scenario 3 closes the cascade around the same power stage and steps the setpoint from 24 to 30 V. Raising the outer gain does not remove the dip, it deepens it, and the exchange holds over a factor of seven in gain. At $K_{c,v}$ of 0.06, 0.12, 0.24 and 0.45 A/V the initial undershoot below 24 V is 1.80, 2.33, 3.50 and 4.62 V and the time to recover to within 0.5 V of the new setpoint is 225, 196, 146 and 114 ms. The fastest tuning produces an undershoot of 4.62 V on a 6 V setpoint change. The trade is a property of the plant rather than of the tuning, because a faster output response requires a larger immediate duty excursion, and (9.2) falls the instant the duty rises regardless of which term raised it. At every gain in the sweep the commanded duty reaches d_{max} in the first controller cycle, because the jump of both feedforward terms with the setpoint adds to the proportional kick, and the gain decides how long it stays there, from 2.7 ms at 0.06 A/V to 12.1 ms at 0.45 A/V.

Scenario 12 extends the same sweep past the useful range. With $K_{c,v} = 0.5$ and $\tau_{I,v} = 0.02\,\mathrm{s}$ on a 4 V setpoint change the undershoot reaches 4.62 V, which is larger than the step itself, the output overshoots to 35.16 V for a peak excursion of 179 % of the step, and the inductor current is driven to 9.50 A against a 9 A setpoint clamp. The undershoot is the same 4.62 V that scenario 3 produces at $K_{c,v} = 0.45\,\mathrm{A/V}$ on a step half again as large. In both runs the commanded duty sits at d_{max} from the first controller cycle until about 12 ms after the step, past the bottom of the dip, so the excursion is set by the limit rather than by the gain. At $K_{c,v} = 0.24\,\mathrm{A/V}$ in scenario 3 the duty leaves the limit after 6.9 ms. Raising the gain to 0.75 A/V produces a sustained cycle between 20.7 and 37.0 V that persists 4 s after the step. That behavior is the practical consequence of the bound in Figure 9.2. A controller

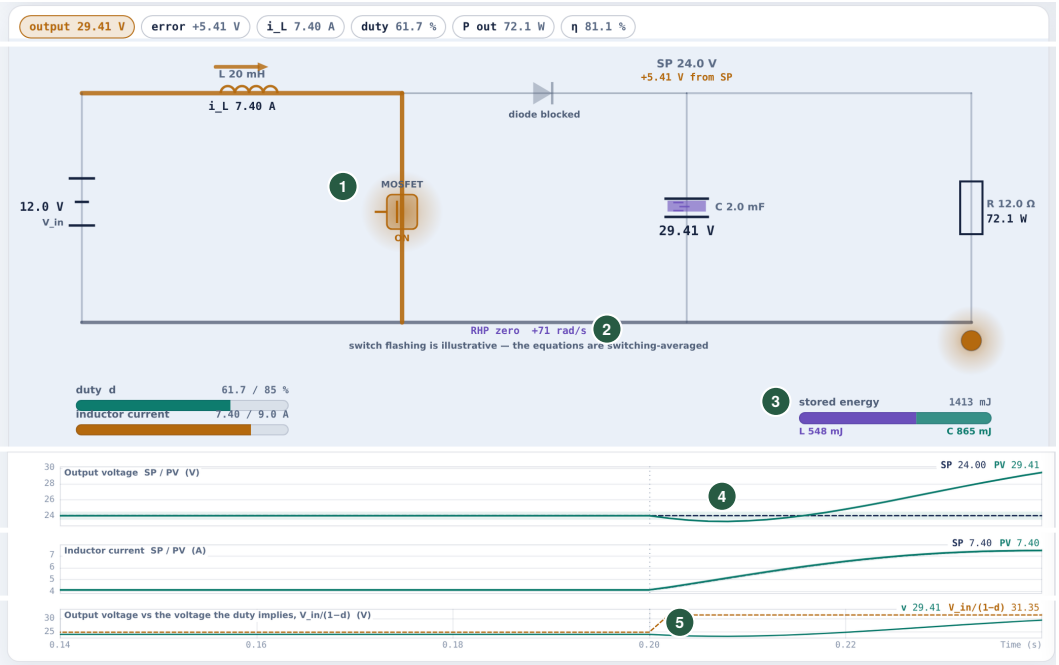


Figure 9.3 Inverse response to a duty step of ten percentage points applied in MANUAL at $t = 0.2$ s (scenario 2), shown at $t = 0.24$ s on a 0.1 s window. The picture is a composite of the application's window: the header values, the circuit view with a blank band of its canvas removed, and three of the five trend strips scaled to the circuit's width. The Control and Learn panels and the scorecard are cut. Callouts: (1) the transistor, on for the larger share of each period, (2) the right-half-plane zero at the operating point of that instant, (3) the stored energy split between L and C, (4) the output dipping below its starting value before it climbs, (5) the requested steady-state voltage $V_{in}/(1-d)$ stepping up while the output still falls. In the logged run the output falls to 23.28 V and is back through 24 V 16 ms after the step while the inductor current builds from 4.14 to 7.40 A, and the requested steady-state voltage steps to 31.35 V while the output is still falling. At the instant shown the output is 29.41 V and total stored energy has risen from 748 mJ at the start to the 1413 mJ the energy bar reads.

that inverts the plant to cross above roughly $z/2$ must generate an ever larger excursion in the wrong direction first, and the appropriate design response is to change L, C or the load rather than the gain.

The feedforward makes the same trade, as scenario 4 shows. A load step from 12 to 6 Ω with feedforward on produces a 7.82 V dip and recovery in 43 ms, and the same step with feedforward off produces a smaller dip of 5.69 V and a recovery of 804 ms, nearly nineteen times longer. The current feedforward is computed from the measured load current, so it jumps the instant the load changes and drives the duty hard immediately, which is what makes the recovery fast and what makes the dip deep. Feedforward on a nonminimum-phase plant trades undershoot for recovery speed exactly as a higher gain does. Whether 7.82 V for 43 ms is preferable to 5.69 V for 804 ms depends on what the downstream circuit cares about, and that is a specification decision rather than a tuning one.

9.6.3 Effect of Capacitance and Inductance Changes

Scenarios 8 and 9 change one component each and separate two effects that a single step response mixes together. Both apply a 24 to 28 V setpoint step with the default tuning, and the reference case at $L = 20$ mH and $C = 2$ mF gives an undershoot of 1.62 V and recovery in 127 ms.

Quartering the capacitance to 0.5 mF leaves the zero exactly where it was, at $+135\,\mathrm{rad/s}$ at 24 V and $+95\,\mathrm{rad/s}$ at the 28 V endpoint in both cases, because C does not appear in (9.9). The undershoot nevertheless grows from 1.62 to 4.27 V (3.97 V on the 2 ms log), because the same missing charge is being taken out of a quarter of the capacitance. What the smaller capacitor does move is the pole pair, from $\omega_n = 77.7$ to 155.4 rad/s, which puts the plant on the unfavorable side of the crossing in Figure 9.2, with the resonance above the bandwidth the zero permits.

> **TRY IT IN THE APP** **8 · Smaller Output Capacitor (0.5 mF)**
>
> Run the scenario and watch the RHP zero tile and the output voltage. The tile reads +135 rad/s at 24 V, as it does with the 2 mF capacitor, yet the first dip after the step to 28 V is 4.27 V, against 1.62 V with the default capacitor. The 2 ms log reads that dip as 3.97 V.

Quartering the inductance to 5 mH moves the zero by the factor of four that (9.9) predicts, from 134.8 to $+539\,\mathrm{rad/s}$ at 24 V, and the undershoot collapses from 1.62 to 0.50 V, bottoming out 1.2 ms after the step and over within 2.5 ms, too brief for the 2 ms log, which reads 0.21 V. A small inductor shortens the current rise time and the interval during which the capacitor is undersupplied. The cost appears outside this model, because peak-to-average current ripple in real hardware goes as $1/L$, so a small inductor buys loop bandwidth with switch rating, capacitor size, conduction loss and radiated emissions. The zero is a design variable that trades against all of those rather than a fixed property of the topology.

9.6.4 Single-Loop and Cascade Comparison

Scenarios 10 and 11 apply an identical 24 to 30 V setpoint step to the same power stage, differing only in whether an inner current loop exists. The single voltage controller is not the slower of the two. It recovers to within 0.5 V in 135 ms against the cascade's 196 ms, and its initial dip is the smaller of the two at 0.78 V against 2.33 V, for the same reason it is quicker to the band and for the same reason it then overshoots to 31.83 V against the cascade's 30.89 V.

The advantages of the cascade lie elsewhere. The first is that the inductor current becomes a regulated and limited variable. In scenario 7 a 12 to 4 Ω overload demands 13.5 A at steady state from a 12 V supply, the outer loop's request is clamped at 9.00 A for 99.5 % of the overload interval, the converter becomes a 9 A current source and the output settles at 19.99 V with the voltage error simply persisting. The peak current reached during that event is 9.47 A with the default inner gain and 11.32 A with $K_{c,i}$ reduced to 5 %/A, so a current limit is only as good as the loop enforcing it. The second is how far the outer loop can be pushed. At $K_{c,v} = 0.45\,\mathrm{A/V}$ the peak excursion on the same setpoint step is 58.6 % of the step with $K_{c,i} = 50\,\%/\mathrm{A}$ and 161.9 % with $K_{c,i} = 5\,\%/\mathrm{A}$, and the peak current is 9.51 against 11.09 A. The third is that load-current feedforward becomes possible at all, because there is a current loop to send it to.

Pause. Try Exercise 9.3 before reading on.

Adequate loop separation has an operational definition in this application. At $K_{c,v} = 0.12\,\mathrm{A/V}$ the recovery barely depends on the inner gain, 196 against 202 ms for inner gains of 50 and 5 %/A, although the slower inner loop moves the duty more gently and dips less (1.44 against 2.33 V). For inner gains from 15 to 200 %/A the recovery stays between 193 and 197 ms and the overshoot between 14.0 and 15.1 %, which means the inner loop has left the outer loop's dynamics. Below roughly 15 %/A the outer loop inherits the inner loop's lag, and the difference between 58.6 and 161.9 % at $K_{c,v} = 0.45$ measures the cost of that inheritance.

Raising the single-loop gain instead of changing the structure produces the failure mode this chapter quantifies. At $K_{c,dv} = 1.2\,\%/\mathrm{V}$ the voltage-only architecture rings, peaking at 37.88 V on a 30 V setpoint and drawing 10.16 A through an unregulated inductor, and it is not inside the band for good until 1.67 s after the step. At 1.5 %/V it never settles. The output cycles between 16.4 and 58.5 V with inductor current peaks of 18.25 A, past the 12 A trip. That gain is past the point where

the right-half-plane zero has consumed the available phase, and no value of $\tau_{I,dv}$ between 0.01 and 3 s recovers a settled response.

9.6.5 Saturation and Anti-Windup

Two scenarios test the actuator boundaries. In scenario 6 the supply collapses to 3.5 V into a 24 Ω load, the ideal ceiling is $V_{in}/(1-d_{max}) = 23.3$ V and the inductor drop pulls the achievable output down to 19.69 V, more than 4 V short of setpoint. The duty is pinned at d_{max} for the whole interval from the event to $t = 2.0$ s and the current setpoint is at its clamp for 92.7 % of it. No tuning fixes that, because no duty in the allowed range holds 24 V from a 3.5 V supply. Back-calculation keeps both integrals bounded while they are clipped [17], but it does not make the release free. The outer integral settles at $i_{lim} - i_{ff} = 1.96$ A, and when the nominal supply and load return at $t = 2.0$ s that stored current carries the output to 29.07 V, 5.07 V above setpoint, and keeps it outside the band for 326 ms. In a counterfactual run that zeroes both integrals at the same instant the peak is 24.11 V, and with the back-calculation term removed it is 102 V.

In scenario 7 the overload is released at $t = 2.4$ s after two seconds against the current clamp. The current feedforward is itself at the 9 A limit during the overload, so the outer integral is held at zero and the current setpoint leaves the clamp within one controller cycle of the release. The inductor current still flowing has nowhere to go but the capacitor, and the output spikes to 33.09 V 6.1 ms after the release. The default integrator then over-corrects into an undershoot to 23.12 V at 86 ms, and the output is inside the ±0.5 V band for good at 163 ms. With $\tau_{I,v} = 3$ s it is inside at 50 ms, because the slow integrator does not chase the spike. With the back-calculation term removed the outer integral would reach 19.5 A during the overload, the current setpoint would stay on its clamp for 614 ms after the release, and the output would not be inside the band for good until 1.01 s. A recursion that absorbed the whole clipped excess at every scan, instead of 1/250 of it, would settle the default case at 229 ms rather than 163 ms, so the tracking time is part of the result.

The graded challenge, scenario 14, applies a 12 to 7 Ω load step at $t = 1$ s to the cascade at its default tuning. It asks for the output to be inside ±0.5 V of 24 V for good within 30 ms, with no more than 0.3 V of overshoot above 24 V and a peak inductor current below 8 A, judged on the logged run at $t = 3$ s, on the case as given, so a changed model parameter or disturbance voids the attempt. The default tuning rides the step out with a 5.78 V dip and a 7.58 A peak, but it is not inside the band for good until 34 ms after the step and then coasts 0.43 V past 24 V, because an integral time of 0.05 s winds in the error from the dip and spends it on the other side. The suggested tuning of $K_{c,v} = 0.25$ A/V and $\tau_{I,v} = 0.15$ s passes with 24 ms, 0.21 V and 7.44 A, which leaves 6 ms, 0.09 V and 0.56 A to spare. The load-current feedforward already supplies the new 7.30 A, so the integrator has little to correct, and lengthening its time to 0.15 s at the default gain cuts the overshoot from 0.43 to 0.16 V while slowing the recovery to 38 ms. The higher gain supplies the speed. The extra speed costs 1.1 V of additional dip (6.88 against 5.78 V), which is the right-half-plane zero at work. At $K_{c,v} = 0.35$ A/V with the same integral time the recovery takes 16 ms but the current peaks at 8.22 A, and at 0.40 A/V the overshoot is 1.27 V, so the passing region is bounded on both sides. The challenge assigns the two terms separate jobs, with the proportional gain buying recovery speed that is paid for in dip depth and current and the integral time setting the overshoot.

9.6.6 Explicit MPC Against the Cascade

Table 9.6 runs scenarios 3, 4, 5 and 7 and the challenge under every architecture on one engine, with the same scenario script, plant, events and duration in every row and no noise or random draw anywhere. The cascade and the single loop run at their shipped tuning (the cascade also at the challenge's suggested tuning), and the MPC runs at scenario 13's settings with nothing retuned per case. The fourth row is the cascade retuned once by the MPC's own rule, the LQR's gains mapped onto the two loops ($K_{c,i} = 127$ %/A from the current gain, $K_{c,v} = 0.08$ A/V from the ratio of the voltage gain to the current gain and $\tau_{I,v} = 0.10$ s from the target filter), with the shipped feedforward, clamp and anti-windup. The shipped feedforward already carries the measured supply voltage and

Table 9.6 Explicit MPC Against the Shipped Architectures and the Cascade Retuned With the LQR's Gains

Structure	dip (V)	t_{band} (ms)	OS (V)	i_{pk} (A)	$t_{>9\,A}$ (ms)	IAE (V s)
Setpoint 24 to 30 V (scenario 3)						
Cascade	2.33	196	0.89	7.044	0	0.305
Voltage only	0.78	135	1.83	7.586	0	0.456
MPC	3.74	29	0.00	7.525	0	0.128
Cascade, LQR gains	2.00	56	0.42	6.812	0	0.346
Load 12 to 6 Ω (scenario 4)						
Cascade	7.82	43	0.49	9.003	5.1	0.263
Voltage only	6.95	1948	0.15	8.614	0	1.372
MPC	10.58	21	0.00	9.019	4.8	0.116
Cascade, LQR gains	7.39	54	0.24	8.808	0	0.303
Supply 12 to 8 V (scenario 5)						
Cascade	4.44	145	0.51	6.840	0	0.208
Voltage only	4.24	449	0.34	6.719	0	0.858
MPC	5.81	30	0.00	7.064	0	0.087
Cascade, LQR gains	4.15	62	0.23	6.670	0	0.237
Overload 12 to 4 Ω (scenario 7)[a]						
Cascade	12.87	—	—	9.475	145.5	8.113
Voltage only	11.10	—	—	12.281[b]	1976.9	2.710
MPC	14.26	—	—	9.005	0.7	8.245
Cascade, LQR gains	12.31	—	—	9.184	31.0	8.115
Challenge, load 12 to 7 Ω						
Cascade	5.78	35	0.43	7.578	0	0.179
Cascade, suggested	6.88	24	0.21	7.441	0	0.156
Voltage only	5.34	322	0.43	7.523	0	0.995
MPC	8.14	18	0.00	8.088	0	0.074
Cascade, LQR gains	5.46	45	0.19	7.424	0	0.205

Dip below 24 V, the time t_{band} after the event at which the output last leaves ±0.5 V of setpoint, the overshoot OS above setpoint, the peak inductor current i_{pk} and the time $t_{>9\,A}$ above the limit come from the state at every 0.1 ms step, so the challenge's 35 and 24 ms read 34 and 24 ms on the logged trajectory. IAE is the scorecard's. [a]No row holds 24 V with 9 A, and the window ends at 2.4 s. [b]Past the 12 A trip.

load current that the MPC's target reads, so both structures receive the same disturbance information. The online solve reproduces every MPC row to the printed digits, and no case produces a miss.

The MPC is inside the band for good 29, 21, 30 and 18 ms after the step, the load step, the sag and the challenge's load step, against 196, 43, 145 and 35 ms for the cascade, which makes it 6.8, 2.0, 4.8 and 1.9 times sooner (1.3 times against the suggested tuning's 24 ms), with no overshoot and with the IAE cut from 0.305 to 0.128 V s on the step. It dips deeper in every case, 3.74 against 2.33 V on the step, 10.58 against 7.82 V on the load step, 5.81 against 4.44 V on the sag, 14.26 against 12.87 V in the overload and 8.14 against 5.78 V on the challenge, and its peak current is higher on every step, 7.525 against 7.044 A on scenario 3. On the challenge the MPC meets the time and overshoot clauses with 18 of 30 ms and 0.00 V, but its 8.088 A peak breaks the 8 A clause, so the MPC as configured fails the graded challenge that the suggested cascade passes. Figure 9.4 draws the dips and the times of Table 9.6 side by side.

A counterfactual run locates the speed. The MPC's own law with its current rows removed, with the same estimator, target and gains and the duty clipped to its range, reproduces the MPC's rows on scenarios 3 and 5 and the challenge to the last digit. The faster recovery is therefore the LQR's state feedback on the estimated current, voltage and applied duty, not the look-ahead and not the constraint. The explicit MPC of Volume 7, Chapter 6 owes its advantage over a PID to state feedback in the same way. Mapping the same gains onto the two proportional-integral loops recovers only part of it, 56 ms against 29 ms on the step, with a smaller dip of 2.00 V. The current rows matter near the limit. Without them the load step peaks at 9.696 A and the overload at 10.461 A with 1989.2 ms above 9 A, while with them the MPC holds the overload to 9.005 A and 0.7 ms above the limit. The cascade, which clamps only the setpoint of its inner loop, reaches 9.475 A and spends 145.5 ms above it, and the retuned cascade 9.184 A and 31.0 ms. The rows bound a linear prediction, so the plant's current

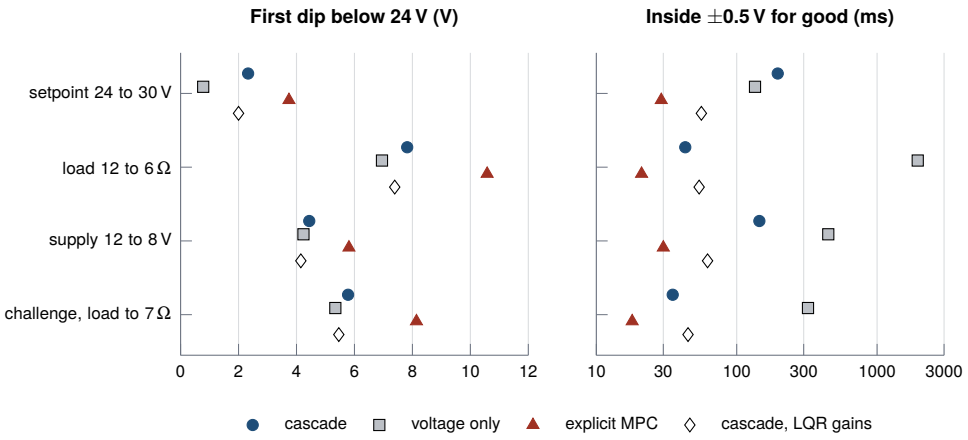


Figure 9.4 Table 9.6 redrawn for the four cases in which the output returns to its setpoint (the overload rows and the challenge's suggested tuning are left out). Each case carries four structures on the same plant: cascade (circles), single voltage loop (squares), explicit MPC (triangles) and the cascade with the LQR's gains (diamonds). The time axis is logarithmic. Notice that the triangles sit farthest left in time and farthest right in dip in every case: the fastest structure always dips deepest, which is the trade the right-half-plane zero sets.

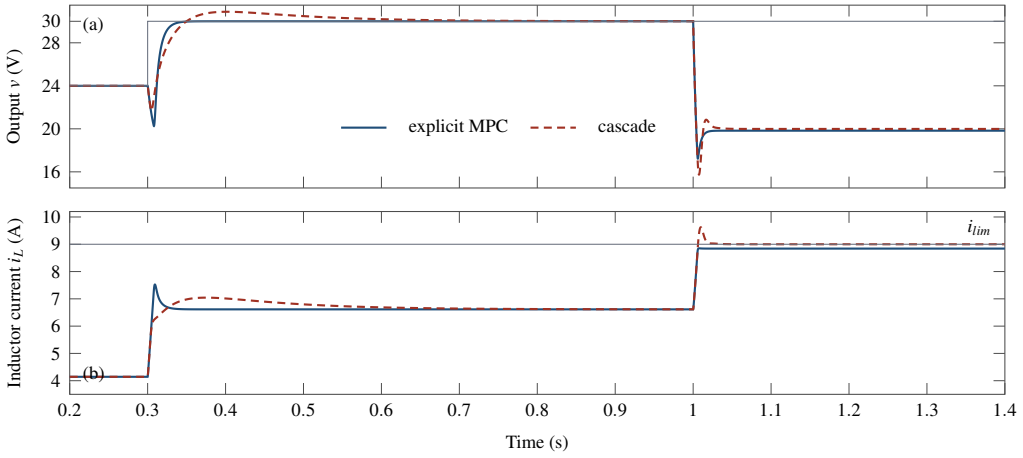


Figure 9.5 Scenario 13 run twice, first with the explicit MPC as loaded and then with the cascade selected before the step, from 0.2 to 1.40 s. MPC solid, cascade densely dashed. (a) Output voltage against the setpoint (thin gray line). (b) Inductor current against the 9 A limit (thin gray line). After the setpoint step at 0.3 s the MPC dips 3.74 V and is inside ±0.5 V in 29 ms, against 2.33 V and 196 ms. After the 4 Ω overload at 1.0 s the cascade's current peaks at 9.623 A and stays above the limit for 169.6 ms, while the MPC's peaks at 8.852 A.

still reaches 9.019 A for 4.8 ms on the load step. The price of keeping v_{SP} as the voltage reference while the limit holds the target is a held output of 19.92 V in scenario 7 against the cascade's 19.99 V.

Scenario 13 puts the step and the overload on one run, as shown in Figure 9.5. The setpoint steps from 24 to 30 V at 0.3 s, and at 1.0 s the load drops to 4 Ω, more than 9 A can supply at 30 V, until it returns at 2.0 s. The MPC's output bottoms out 8.1 ms after the step at 20.26 V and the cascade's 4.7 ms after it at 21.67 V. From 1.0063 s the cascade's current is above 9 A for 169.6 ms, with a peak

Table 9.7 Worst Case of Each Safety-Relevant Output Over 39 Cases

Output	Cascade	Voltage only	MPC
Peak i_L (A)	9.749	19.794	9.222
Time above 9 A (ms)	173.5	2385.8	4.8
Overcurrent trips	0	9	0
Discontinuous conduction	0	3	0
Highest v (V)	37.83	53.99	53.91
Lowest v (V)	7.84	10.11	5.79

Every scenario for 3 s, the challenge at its suggested tuning, every model parameter at both ends of its range on scenario 11's step, the supply and load at their extremes, the disturbance events in combination, the setpoint range and two component changes with an overload, each under the three architectures.

of 9.623 A, and it then holds 9.000 A and 19.99 V. The MPC's current peaks at 8.852 A and at 1.40 s holds 8.842 A and 19.83 V. During the hold the binding rows are the current limit and the duty ceiling on the third move only, the applied duty is 0.4394 against the target's 0.4447, and the plan predicts 8.889 A one period ahead, 47.8 mA above the current then measured. The scorecard of the whole run is row 13 of Table 9.5, whose settling of 1.706 s counts from the step because the overload keeps the output out of the band until the load is released.

The right-half-plane zero survives the change of controller. The prediction contains it, so every plan accounts for the output falling while the current builds, but three moves span 0.6 ms, a small fraction of the 7.5 ms the open-loop dip of scenario 2 takes to reach its bottom, and the terminal cost carries the rest of the response. The plan cannot remove the dip or move the zero. It places the closed loop at a different point of the trade, 3.74 V for 29 ms on the 24 to 30 V step, where the cascade sweep of Section 9.6 gives 3.50 V for 146 ms at $K_{c,v} = 0.24\,\mathrm{A/V}$ and 4.62 V for 114 ms at 0.45 A/V. Feedback on the estimated current and applied duty buys recovery time that the proportional-integral cascade does not reach, and it pays for that time in dip depth exactly as a higher outer gain does.

Table 9.7 gives the worst case of each safety-relevant output over 39 cases. The MPC's peak current is at most 9.222 A, with the 5 mH inductor and an overload, against 9.749 A for the cascade and 19.794 A for the single loop, and the MPC never trips the 12 A flag or enters discontinuous conduction. It is safer on current in 15 of the 39 cases and worse than the cascade on some output in 20. In 12 of those 20 only its peak current is higher, the price of its speed, and it stays under the limit in all of them except the 9.019 A of the load step. Seven of the other eight are model mismatch with a component changed. With $C = 0.5\,\mathrm{mF}$ the output overshoots to 33.96 V against 28.43 V, and with an overload added it falls to 5.79 V against 7.84 V. With $L = 60\,\mathrm{mH}$ it overshoots to 40.74 V against 32.16 V. With $C \le 0.3\,\mathrm{mF}$ or $L \le 3\,\mathrm{mH}$ the MPC does not settle on scenario 11's step, ending 0.660 V short at $C = 0.3\,\mathrm{mF}$ and 0.529 V short at $L = 3\,\mathrm{mH}$ after 3 s, where the cascade ends on setpoint. The one loss at the declared components is a supply sagged three times to 3 V, where the MPC's output reaches 25.77 V against 24.29 V. The highest output of 53.91 V occurs at $C = 0.2\,\mathrm{mF}$, and the application has no overvoltage limit, so no protection is crossed.

DID YOU KNOW? Nine Trips Without an Inner Loop

From the simulator: over 39 cases the single voltage loop drives the inductor to 19.794 A, trips the 12 A overcurrent flag 9 times and leaves continuous conduction 3 times. The cascade and the MPC never trip (Table 9.7). The inner loop is first of all a guard on the inductor current.

The single linearization at 24 V serves the whole setpoint range. Over steps from 24 to targets between 13 and 30 V the one-period current prediction errs by 0.9 mA or less on average and by at most 125 mA, and steps to 13 and 18 V settle in 28 and 25 ms against 180 and 128 ms for the cascade. Targets of 36 to 48 V are held by the current limit, where the target is not an equilibrium of the model and the mean error grows to 5.6 to 52 mA. No step from 24 V produced a miss. A step from 48 to 13 V leaves the table for 4.2 ms and still reaches 13 V within 0.5 V in 41 ms, where the cascade enters discontinuous conduction and takes 418 ms.

9.6.7 Overshoot Measurement Under Inverse Response

The overshoot in Table 9.5 normalizes the peak against the span from the process value before the setpoint change to the new target, and on this plant the choice of the starting value matters. The first logged sample that carries the new setpoint is taken 2 ms after the step. On a minimum-phase plant the process value has not moved by then. On this plant it has moved the wrong way. At $K_{c,v} = 0.45$ that sample already reads 22.855 V, so a normalization against it would use a span of 7.145 V instead of 6 V and report 49.0 % where the logged trajectory gives 58.3 %, and scenario 12 would read 138.3 % instead of 177.9 %. The application normalizes against the last sample before the change, so the overshoot column agrees with the logged trajectory. A metric that is exact on a minimum-phase plant can understate the overshoot by a fifth or more on a nonminimum-phase one.

9.7 Technology in Practice

Figure 9.6 puts the cascade of Section 9.5 on real hardware: a current sensor and a voltage divider feed a digital controller, whose two loops set the duty of a power transistor through its gate driver.

What Measures It

Two readings carry the cascade, and they are not equally hard to get. The output voltage reaches the controller through a divider and a conversion to digital form, whose dynamics and quantization the model omits, as its assumptions state. The inductor current is harder, because every current-sensing method trades size, bandwidth, linearity, isolation, accuracy and cost against one another [23]. The CT of Figure 9.1 stands for a perfect and instantaneous reading of the averaged current. A plain current transformer cannot pass the direct part of that current, so a real inner loop reads a shunt or a magnetic sensor that responds down to direct current [23], and the noise, offset and bandwidth of that sensor are what Section 9.8 says the model leaves out.

TECHNOLOGY Sensor: Reading the Inductor Current

A cascade's inner loop is only as good as its reading of the inductor current, and three families of sensor serve power converters. A shunt resistor turns current into a voltage by Ohm's law and can follow it from direct current up to megahertz, but it sits in the power path without isolation, adds loss and parasitic inductance, and its reading drifts as its resistance rises with temperature. Hall-effect and other magnetic sensors are non-intrusive but need temperature compensation, and the offset drift of a Hall device produces readings unrelated to the actual current. A current transformer saturates on large currents and, like a Rogowski coil, cannot detect direct current at all [23]. On a trend an offset drift reads as current the inductor does not carry, and the inner loop then holds that false value.

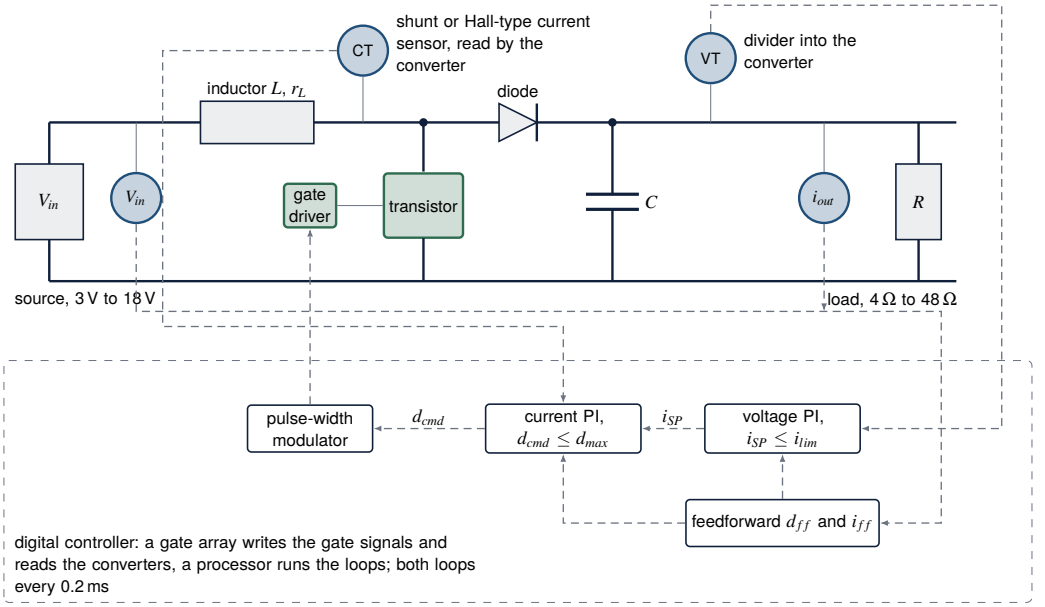


Figure 9.6 The cascade of Figure 9.1 on real hardware. The voltage loop reads the output through a divider and sets the current setpoint, the current loop reads a shunt or Hall-type sensor on the inductor path and sets the duty, and a pulse-width modulator and a gate driver turn that duty into on and off times of the transistor. The source voltage and the load current feed both feedforward terms. Solid lines are the power path and dashed gray arrows are signals. Notice that the inner loop's sensor sits in the power path beside the switch, which is why Section 9.8 calls current sensing the hardest part of building this cascade.

What Moves It

The final element is the power transistor and its gate driver, and the modulator turns the commanded duty into on and off times once per switching period. The model replaces that chain by the first-order lag $\tau_{pwm} = 0.2\,\text{ms}$ of (9.3) and by the clamp at $d_{max} = 0.85$. The clamp to $[0, d_{max}]$ is the only nonlinearity of the actuator that the chapter carries, and scenario 6 shows it binding.

> **TECHNOLOGY** **Final Element: The Power Transistor**
>
> The duty cycle acts through a power transistor, such as a metal-oxide-semiconductor field-effect transistor, that a gate driver turns on and off at the modulator's command. Silicon devices have matured to the point of pushing the fundamental limits of the material. Wide bandgap devices of silicon carbide and gallium nitride operate at higher voltages and temperatures and switch at higher frequencies with greater efficiency, with silicon carbide expected to prevail above 600 V and gallium nitride below it [8]. Faster switching brings its own trouble, because parasitic inductance and resistance in the layout cause losses and overshoots of voltage and current that raise electromagnetic interference [8]. Aging of the switches and the capacitors shows mostly in current transients and high-frequency content, so watching for it takes a fast current sensor [23].

Where the Controller Runs

A converter controller is an embedded system. Typically a field-programmable gate array manages the inputs and outputs, generating the gate signals and reading the analog-to-digital converters, while one or more microcontrollers or digital signal processors compute all or part of the control law [20]. In one arrangement the fast controller runs in the gate array while an outer loop on a processor core supplies its reference at a lower rate [20], which is the cascade of this chapter split across two layers

of hardware. Sampling intervals of 150 µs or less are common in predictive control of converters, a few tens of microseconds on modern platforms [20]. The application's control period of 0.2 ms is one period of a 5 kHz digital controller, deliberately slow so that the inverse response can be watched.

> ⏢ **IN PRACTICE** **Above Half Duty, a Peak-Current Loop Needs a Ramp**
>
> A widely used form of current-mode control compares the peak inductor current with its reference once per clock period, which suits supplies regulated over a wide range of loads at a constant frequency. Such a peak loop is unstable at a duty of 0.5 or greater: a small error grows from one cycle to the next and appears as a cycle-to-cycle oscillation of the inductor current at half the switching frequency [24]. The cure is a compensating slope added to the sensed current, and a very commonly recommended slope is half the falling slope of the inductor current, which guarantees suppression without saying how fast [24]. The default duty of 0.517263 already lies past that boundary.

What Is Modernizing It

Three developments bear on this power stage. The first is the switch itself. Wide bandgap transistors allow higher switching frequencies and therefore smaller inductors and capacitors [8], which by (9.9) also moves the right-half-plane zero away from the origin, as scenario 9 shows with a quarter of the inductance. The second is the sensor. Current sensors embedded in the converter can reduce the usual problems of sensors used for control, and monitoring the condition of switches and capacitors through their current can extend the life of a converter [23]. The third is predictive control. Platforms that combine a gate array with processor cores on one chip, and solvers that keep the computing burden down, are where research on predictive control of converters is heading [20], and the explicit table of Section 9.5.1 is one answer to that burden.

9.8 Limitations

The averaged model is the largest limitation and it is deliberate. Peak-current mode control in hardware becomes subharmonically unstable above 50 % duty unless a compensating ramp is added, and that instability lives inside a switching period and is invisible to (9.1) and (9.2) [15, 16]. The inner loop in this application is an average-current loop sampled at 5 kHz, which does not have that failure mode, so the application teaches why an inner current loop is the right architecture and not how a peak-current comparator is stabilized.

The continuous-conduction assumption is enforced rather than modeled. When the averaged inductor current reaches zero the state is clamped, a fault is raised and the run is declared out of range, so the application says nothing about the discontinuous mode that a real converter enters at light load. With the shipped disturbance ranges the equilibrium current never reaches zero, and the lightest available load of 48 Ω at 24 V still draws 1.01 A through the inductor. A full load shed from 12 to 48 Ω in AUTO does not trigger the flag either, because the inductor current bottoms out at 0.15 A.

Instantaneous efficiency is reported as the ratio of output to input power and therefore exceeds 100 % for a few milliseconds whenever stored energy is leaving the converter. That behavior is arithmetic rather than a fault, but the number invites misreading. The reported efficiency is in any case an upper bound, because switching loss, core loss and diode conduction loss are all absent and those three dominate the loss budget of a real converter at 50 W.

The measurement chain is perfect. Real current sensing is the hardest part of building the architecture this application recommends, because a shunt or a current transformer contributes noise, offset and bandwidth limits that interact directly with the inner loop gain. Nothing here penalizes a large $K_{c,i}$, so the application cannot show why real designs stop increasing it.

The explicit MPC's model leaves out the bilinear terms of (9.1) and (9.2) away from 24 V, any component a learner changes, discontinuous conduction and the trip. Its prediction errs by up to 125 mA per period at a 13 V target, which is why the current floor carries a margin, and with the capacitance or inductance a quarter of its declared value or less the plan stops settling. Its current

rows bound the predicted current, so the plant reaches 9.019 A on the load step. Because those rows are hard, a miss while an overload holds the current at the limit leaves the cascade in charge until the load is released, and a softened current row, which would change the program and its partition, is not implemented. The MPC loses the dip on every case and the challenge's current clause, and its three moves cover 0.6 ms of a response that lasts tens of milliseconds, so the look-ahead decides only near the limits.

> ⚠ **OPEN CHALLENGE** **Optimizing Within a Few Microseconds**
>
> Converter controllers act within a few tens of microseconds, so solving the optimization in real time is the main obstacle to predictive control of power electronics, and it calls for powerful platforms and efficient solvers [20]. Many algorithms also assume a linear model while many converters are not linear, so the model is linearized and loses accuracy [20], as the single linearization of this chapter does away from 24 V. Progress would be nonlinear formulations and solvers that fit inside the sampling interval. Cutting the computational burden is itself a main industrial research focus [20].

9.9 Discussion

Boost stages sit between a photovoltaic string and an inverter, between a fuel cell and a traction inverter, and between a battery pack and a motor drive, and in each of those the conversion efficiency is what decides how much of a scarce resource reaches the load [1, 2]. The design choices that set efficiency are also the choices that set the zero. A larger inductor lowers current ripple and conduction loss and moves the zero toward the origin, which (9.9) states directly and Figure 9.2 quantifies over the load range. A converter sized for the lightest ripple is a converter whose voltage loop must be the slowest, and a converter sized for loop bandwidth pays in conduction loss, switch area and radiated emissions.

The load dependence is the part that has operational consequences. Because z falls as $q_0^2 R$ falls, the zero migrates toward the origin exactly when the converter is delivering the most power and the load transients are largest. A loop tuned at light load for a crossover near $z/2$ can lie above the bound at full load, and Figure 9.2 puts numbers on the range, with the bound falling from 292 rad/s at 48 Ω to 17.2 rad/s at 4 Ω. Current-mode control provides more than a current limit and a feedforward path. It gives the outer loop a plant, (9.11), whose bandwidth does not collapse with load in the same way.

> ⚠ **OPEN CHALLENGE** **No Current Sensor Does Everything**
>
> As switches get faster, the current sensor becomes the weak link of the inner loop. Almost no current-sensing method provides the necessary size, bandwidth, linearity, isolation, accuracy and cost in one place, and almost none of the inexpensive options suits the control of high-frequency wide bandgap converters [23]. Progress would be sensors embedded in the converter, such as a Rogowski coil built into the circuit board or magnetoresistive and fluxgate elements, which can reduce or remove the usual problems of sensors used for control [23], fast enough for the inner loop and small enough to sit beside the switch.

9.10 Conclusion

This chapter studied a boost converter simulator with three states, a switching-averaged continuous-conduction model, two feedforward terms computed from measurements the hardware really has, a voltage-to-current cascade, a single-loop alternative and an explicit MPC on the same power stage, and fourteen scenarios. The control-to-output transfer function carries a right-half-plane zero at $(qv - r_L i_L)/(L i_L)$, which is +134.8 rad/s at the default operating point and sweeps by a factor of 17 across the allowed load range while the *LC* natural frequency stays near 77 rad/s. That zero fixes a trade the controller cannot alter, measured here as an undershoot growing from 1.80 to 4.62 V as the recovery time falls from 225 to 114 ms over a factor of seven in outer gain, and reproduced by the feedforward, which buys a 43 ms load recovery in place of 804 ms at the cost of a 7.82 V dip in place of 5.69 V. The inner current loop does not shrink the first dip, and this chapter measured the three

things it does instead, a current limit that holds the peak to 9.47 A instead of 11.32 A, an outer loop that can be driven to 0.45 A/V with 58.6 % rather than 161.9 % excursion, and a place to send load feedforward. The explicit MPC, a 109-region table of a three-move program with the inductor current as a hard constraint, reaches the band 1.9 to 6.8 times sooner than the cascade and holds the overload to 9.005 A against 9.475 A, while the same zero deepens its dip on every case.

Two of the scenarios are component changes rather than tuning changes, and they separate the capacitor, which deepens the dip from 1.62 to 4.27 V without moving the zero, from the inductor, which moves the zero from 135 to 539 rad/s and collapses the dip to 0.50 V. The model omits diode drop, switching loss, capacitor equivalent series resistance, current-sense dynamics and everything that happens inside a switching period, so the efficiency figures are upper bounds and the peak-current-mode instability that motivates slope compensation cannot appear. Extending the simulator with a constant-power load, a sensed-current filter and a discontinuous-conduction branch would each add a distinct failure mode to the same power stage. Within those limits the simulator turns the cost of a right-half-plane zero and the return on rewiring a loop into measured numbers.

KEY POINTS

- Raising the duty first starves the output and only later feeds it, so the output dips before it climbs.
- The zero belongs to the power stage: no gain, integral time or controller structure moves it, while the inductor and the load do.
- A faster recovery costs a deeper first dip whether the speed comes from a higher gain, from feedforward or from an MPC.
- The inner current loop does not shrink the first dip. It limits the current, tolerates a harder outer gain and gives load feedforward a place to enter.

Where else this applies. Volume 2, Chapter 12 closes an armature-current loop inside a speed loop, and the drum level of Volume 18, Chapter 4 carries a right-half-plane zero that moves with load.

References

[1] F. Blaabjerg, Z. Chen, and S. B. Kjaer, “Power electronics as efficient interface in dispersed power generation systems,” *IEEE Transactions on Power Electronics*, vol. 19, no. 5, pp. 1184–1194, 2004.

[2] A. Emadi, Y. J. Lee, and K. Rajashekara, “Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles,” *IEEE Transactions on Industrial Electronics*, vol. 55, no. 6, pp. 2237–2245, 2008.

[3] J. Freudenberg and D. Looze, “Right half plane poles and zeros and design tradeoffs in feedback systems,” *IEEE Transactions on Automatic Control*, vol. 30, no. 6, pp. 555–565, 1985.

[4] J. B. Hoagg and D. S. Bernstein, “Nonminimum-phase zeros - much to do about nothing - classical control - revisited part II,” *IEEE Control Systems*, vol. 27, no. 3, pp. 45–57, 2007.

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

[6] 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.

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

[8] U.S. Department of Energy, “Wide bandgap semiconductors for power electronics,” U.S. Department of Energy, Quadrennial Technology Review 2015, Chapter 6, Technology Assessment 6N, 2015. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/02/f29/QTR2015-6N-Wide-Bandgap-Semiconductors-for-Power-Electronics.pdf

[9] S. R. Sanders, J. M. Noworolski, X. Z. Liu, and G. C. Verghese, “Generalized averaging method for power conversion circuits,” *IEEE Transactions on Power Electronics*, vol. 6, no. 2, pp. 251–259, 1991.

[10] R. D. Middlebrook and S. Cuk, “A general unified approach to modelling switching-converter power stages,” in *1976 IEEE Power Electronics Specialists Conference*, 1976, pp. 18–34.

[11] G. W. Wester and R. D. Middlebrook, “Low-frequency characterization of switched dc-dc converters,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-9, no. 3, pp. 376–385, 1973.

[12] R. D. Middlebrook, "Small-signal modeling of pulse-width modulated switched-mode power converters," *Proceedings of the IEEE*, vol. 76, no. 4, pp. 343–354, 1988.

[13] V. Vorperian, "Simplified analysis of PWM converters using model of PWM switch. continuous conduction mode," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 3, pp. 490–496, 1990.

[14] R. W. Erickson and D. Maksimović, *Fundamentals of Power Electronics*. Cham, Switzerland: Springer International Publishing, 2020.

[15] R. D. Middlebrook, "Modeling current-programmed buck and boost regulators," *IEEE Transactions on Power Electronics*, vol. 4, no. 1, pp. 36–52, 1989.

[16] R. B. Ridley, "A new, continuous-time model for current-mode control (power convertors)," *IEEE Transactions on Power Electronics*, vol. 6, no. 2, pp. 271–280, 1991.

[17] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

[18] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[19] P. Karamanakos, T. Geyer, and S. Manias, "Direct voltage control of DC–DC boost converters using enumeration-based model predictive control," *IEEE Transactions on Power Electronics*, vol. 29, no. 2, pp. 968–978, 2014.

[20] P. Karamanakos, E. Liegmann, T. Geyer, and R. Kennel, "Model predictive control of power electronic systems: Methods, results, and challenges," *IEEE Open Journal of Industry Applications*, vol. 1, pp. 95–114, 2020.

[21] 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.

[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] A. Parsa Sirat and B. Parkhideh, "Current sensor integration issues with wide-bandgap power converters," *Sensors*, vol. 23, no. 14, p. 6481, 2023.

[24] A. Tiwary and G. A. Rincón-Mora, "Sub-harmonic stability and slope compensation boundary in peak/valley DC–DC converters," *Electronics*, vol. 15, no. 8, p. 1609, 2026.

9.11 Exercises

9.1 ⌕ Concept check · 2 min

In scenario 2 the duty steps from 0.517 to 0.617 in MANUAL, and the output voltage starts to fall although its new steady state is near 29.7 V. What sets the sign of the initial slope?

(a) The modulator lag holds the old duty, so the output follows it at first

(b) The delivered current $q\,i_L$ drops at once, while i_L cannot change in that instant

(c) The inductor drop $r_L i_L$ rises before the output can respond

(d) The load draws more current as soon as the duty rises

9.2 ⌕ Concept check · 1 min

Scenario 8 quarters the output capacitor, from 2 to 0.5 mF, and repeats the 24 to 28 V step. What happens to the right-half-plane zero and to the first dip?

(a) The zero moves four times farther out and the dip shrinks

(b) The zero moves toward the origin and the dip deepens

(c) The zero stays where it was and the dip deepens

(d) The zero stays where it was and the dip is unchanged

9.3 ⌕ Concept check · 2 min

On the 24 to 30 V step of scenarios 10 and 11, the single voltage loop dips 0.78 V and is inside the band in 135 ms, and the cascade dips 2.33 V and takes 196 ms. What does the inner current loop buy?

(a) A smaller first dip, because the inner loop removes the zero from the plant
(b) A faster recovery at the shipped tuning, because the inner loop crosses near 600 rad/s
(c) A controller pole that cancels the zero, so the outer loop can cross above z
(d) A limited inductor current, room for a harder outer gain and a path for load feedforward

9.4 Theory and calculation · 6 min

The default point is $V_{in} = 12\,\text{V}$, $R = 12\,\Omega$, $r_L = 0.1\,\Omega$ and $v = 24\,\text{V}$.

(a) The ideal relation $v = V_{in}/(1-d)$ asks for $d = 0.5$. Put that duty into Equation (9.5) and find how far short of 24 V the output settles. *Checkpoint:* 23.226 V, 0.774 V short.

(b) Use Equation (9.6) to find the duty that holds 24 V, the inductor current and the inductor drop. *Checkpoint:* $q = 0.482737$, $d = 0.517263$, $i_L = 4.1430\,\text{A}$, $r_L i_L = 0.414\,\text{V}$.

(c) Compute d_{ff} and i_{ff} of Equation (9.14) at this point. What does each integral accumulator hold at steady state, and what is the efficiency? *Checkpoint:* $d_{ff} = 0.5000$, $i_{ff} = 4.1430\,\text{A}$. Inner integral 0.017263, outer integral zero, 96.55 %.

9.5 Theory and calculation · 6 min

Use the operating point of the previous exercise and $L = 20\,\text{mH}$, $C = 2\,\text{mF}$.

(a) Evaluate the zero of Equation (9.9) exactly and with the approximation $q_0^2 R/L$. *Checkpoint:* 134.8 rad/s and 139.8 rad/s.

(b) From the denominator of Equation (9.8) find ω_n and ζ, and the ratio z/ω_n. *Checkpoint:* 77.68 rad/s, 0.300, 1.74.

(c) Repeat the exact zero for a $4\,\Omega$ load at the same 24 V, and give the bandwidth bound $z/2$ at both loads. *Checkpoint:* $q_0 = 0.44365$, $i_0 = 13.52\,\text{A}$, $z = 34.4\,\text{rad/s}$. Bounds 67 and 17.2 rad/s.

9.6 Theory and calculation · 6 min

The current feedforward of Equation (9.14) uses the measured load current at the setpoint $v_{SP} = 24\,\text{V}$, with $V_{in} = 12\,\text{V}$, $r_L = 0.1\,\Omega$ and $i_{lim} = 9\,\text{A}$.

(a) The challenge drops the load from 12 to $7\,\Omega$. Compute i_{ff} at the instant of the step, while the output is still 24 V, and the jump from its value at $12\,\Omega$. *Checkpoint:* 7.30 A, a jump of 3.16 A.

(b) Scenario 7 drops the load to $4\,\Omega$. Compute the current the power balance asks for at 24 V, with and without r_L, and what the clamp passes. *Checkpoint:* 13.52 A, 12 A lossless, clamped to 9 A.

(c) With i_L held at 9 A, use Equation (9.11) at steady state to find the output the overload holds. *Checkpoint:* 19.99 V.

9.7 App challenge · 8 min

Load scenario 2, Inverse response, step the duty by hand, and use the default seed. Change nothing, run to $t = 0.24\,\text{s}$ and export the CSV file. The scenario steps the duty from 0.517 to 0.617 at 0.2 s.

(a) Report the lowest `output_voltage_V` after 0.2 s, the row where it occurs and the first later row at or above 24 V. *Checkpoint:* 23.28 V in the 0.208 s row, back at 0.216 s.

(b) Report `output_voltage_V` and `duty_implied_V` at 0.24 s, and `stored_energy_J` at 0.2 and 0.24 s. *Checkpoint:* 29.41 and 31.35 V. 748 and 1413 mJ.

(c) Report `capacitor_energy_J` and `inductor_energy_J` at 0.2 and 0.206 s, and say where the charge withheld from the output has gone. *Checkpoint:* 576 to 543 mJ and 172 to 238 mJ.

9.8 App challenge · 8 min

Load scenario 3, Output setpoint step, 24 to 30 V, and use the default seed. The setpoint steps at 0.3 s.

(a) Change nothing, run to 0.6 s and export the CSV file. Report the lowest `output_voltage_V` after the step as a dip below 24 V, and the Settling, Overshoot and Peak i_L readings of the scorecard. *Checkpoint:* Dip 2.20 V (within 0.05 V), settling 0.194 s, overshoot 14.8 %, peak 7.04 A.

(b) Press Reset, set Voltage gain (outer) to 0.24 A/V before Run, and repeat. Then Reset, set 0.45 A/V and repeat. *Checkpoint:* 3.43 V, 0.144 s, 19.9 %, 7.50 A. 4.61 V, 0.112 s, 58.3 %, 9.51 A.

(c) From 0.12 to 0.45 A/V, how many volts of dip does each millisecond of settling cost? Why can no outer gain avoid the exchange? *Checkpoint:* About 0.03 V per millisecond.

9.9 App challenge · 9 min

Load the challenge, Ride out a load step, and use the default seed. The load falls from 12 to 7 Ω at 1 s, and each run lasts 3 s.

(a) Change nothing and run. Report the three readings of the challenge verdict and name the clauses that fail. *Checkpoint:* 34 ms, 0.43 V, 7.58 A. The time and overshoot clauses fail.

(b) Press Reset, set Voltage gain (outer) to 0.25 A/V and Voltage integral time to 0.15 s before Run, and run. Then Reset, raise the gain to 0.35 A/V and run again. *Checkpoint:* 24 ms, 0.21 V, 7.44 A, met. 16 ms, 0.28 V, 8.22 A, the current clause fails.

(c) Press the reload button, set Control architecture to Explicit MPC before Run, and run. Report the verdict. *Checkpoint:* 16 ms, 0.00 V, 8.09 A, the current clause fails.

Say what the gain and the integral time each did in (b), and why the MPC fails.

9.10 Application area · 4 min

A designer wants the bandwidth bound $z/2$ at the 4 Ω full load to equal the bound at the 12 Ω default point with $L = 20\,\text{mH}$, by changing only the inductor. The operating point at a given load does not depend on L.

(a) What inductance is needed? Use $z = 134.8\,\text{rad/s}$ at 12 Ω and 34.4 rad/s at 4 Ω. *Checkpoint:* About 5.1 mH.

(b) By what factor does the current ripple grow at the same switching frequency, and what pays for it? *Checkpoint:* About 3.9.

9.11 Application area · 3 min

The app reports 96.55 % efficiency at the default point, with $P_{in} = 49.72\,\text{W}$ and $P_{out} = 48\,\text{W}$, and its diode is ideal.

(a) A real diode drops roughly 1 V and carries the output current of 2 A on average. Estimate the efficiency with that loss added. *Checkpoint:* About 92.8 %.

(b) The efficiency badge can read above 100 % for a few milliseconds. When, and is it a fault?

9.12 Review the agent · 5 min

On scenario 4, the 12 to 6 Ω load step, the cascade with feedforward on dips 7.82 V. An engineer asks an AI assistant how to make the dip smaller. It replies: “Load-current feedforward acts before an error appears, so it can only reduce the dip. The 7.82 V must come from the outer loop. Keep feedforward on and halve the outer gain to 0.06 A/V.” Find the flaw, and say how you would check the claim in the app.

CHAPTER 10

Commanding an Attitude Instead of a Force in a Rocket Landing Simulator



The controller moves 1 engine tilt on its gimbal 2 engine thrust (throttle)

The controller reads 1 position, descent speed and tilt of the vehicle

What disturbs it 1 crosswind and gusts

The only engine sits below the vehicle, so tilting it to steer toward the pad first pushes the vehicle the wrong way while the crosswind leans it downwind. Illustration, not to scale.

IN PLAIN TERMS **Steering a Rocket Down by Tilting It**

The main idea. A rocket landing on one engine cannot push itself sideways. To reach the pad it must first lean over. A second, faster automatic controller makes it lean, and only then does the rocket start to slide.

The challenge. The engine hangs below the middle of the rocket. When it swings to start the lean, it first shoves the rocket the wrong way, and no setting can stop that. So the part that steers toward the pad must stay several times slower than the part that holds the lean.

Why it matters. A reusable rocket must land gently on a small pad, and these controllers decide if its planned path can be flown at all.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history of the throttled descent engine (Section 10.2), the fact on the misled throttle (Section 10.3), the fact on the rate term and the two try-it segments (Section 10.7) 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 10.7), Technology in Practice (Section 10.9), Limitations (Section 10.10) and the open challenge

The control question A lander on one gimballed engine cannot push itself sideways, so its position loop commands a tilt and borrows its authority from an attitude loop. How fast may that outer loop be, when the engine hung below the center of mass puts a right-half-plane zero in its path that no attitude gain can move?

Learning objectives

- Derive the attitude plant gain, the right-half-plane zero and the closed-loop frequencies of all three loops from the vehicle's constants.
- Explain the wrong-way first motion of a thrust-vectored vehicle and compute its timing from the linearized step response.
- Trace a steady crosswind back through the cascade to the standing pad offset of two proportional-derivative loops.
- Diagnose a slow actuator, a saturated gimbal and an engine short of deceleration authority, and say which of them retuning can repair.
- Repair a misbehaving cascade from the inside out.

Prerequisites Sections 1.4.5 (inner-loop speed rule), 1.4.3 (static feedforward), 1.6.2 (velocity form) and 1.6.4 (scenarios and the challenge). Chapter 9 for a cascade against a right-half-plane zero.

The app https://apmonitor.com/apps/rocket-landing/

See also Volume 7, Chapter 5, Volume 7, Chapter 7 and Volume 7, Chapter 9 for position loops that act through an angle, and Volume 20, Chapter 3 and Volume 20, Chapter 7 for tilting thrust with a wind estimate and for attitude control in orbit.

10.1 Introduction

The manipulated variable of a teaching loop is normally a force or a flow that acts on the controlled variable directly. A valve admits steam and the temperature moves, a pump adds liquid and the level moves, a motor draws current and the shaft turns. What the designer chooses is how much authority to apply and how quickly to apply it. A rocket coming down on one engine has nothing of that kind available for horizontal position. The only lateral force the vehicle can generate is the horizontal component of its own thrust, so the position loop must first ask the airframe to point somewhere else and then wait for the attitude loop to deliver the tilt. The command it issues is an angle, and that angle produces no force until a second and faster loop has delivered it.

Two consequences follow and both are geometric rather than a matter of tuning. The first is that the position loop inherits the attitude loop's speed as a ceiling, because an outer loop that revises its command before the inner loop has executed the previous one is not a cascade at all. The second is less apparent. The engine hangs below the center of mass, so the deflection that pitches the nose to the right pushes the tail to the right as well, and the vehicle accelerates the wrong way for as long as the gimbal is deflected. That transient puts a real zero in the right half plane, at a frequency set by the thrust, the pitch inertia and the distance from the center of mass to the gimbal, and no amount of inner-loop gain moves it.

Powered descent is most often presented as a trajectory problem. Optimal thrust programming for a lunar soft landing was posed in that form early on [1], the Apollo lunar module flew a polynomial guidance law that computed a commanded thrust vector for an attitude autopilot to track [2], and the modern treatment casts the same problem as a second-order cone program that solves to global optimality in bounded time [3–6]. However, a guidance law produces a desired acceleration and nothing else, and the loops that turn a desired acceleration into a gimbal angle are

ordinary proportional, integral and derivative terms arranged in a cascade, the same successive loop closure used on fixed-wing and rotary-wing vehicles [7]. Those loops decide whether the trajectory is flyable, and their limits come from the airframe.

The scenarios of the application cover cascade structure and successive loop closure, bandwidth separation between an outer and an inner loop, non-minimum-phase response and the ceiling a right-half-plane zero places on achievable bandwidth, actuator rate limits and saturation inside a feedback path, feedforward against a known nonlinearity, the offset a proportional loop leaves against a constant load, and a plant that drifts during a single run [8, 9]. The physical fundamentals behind the model are Newton's second law in two translational and one rotational degree of freedom, a moment balance taken about the center of mass with the thrust acting through a lever arm below it, and the constant exhaust-speed form of the variable-mass momentum balance [10, 11]. The behavior matters industrially because a reusable booster is graded on its last thirty seconds, and the loops that track a descent trajectory decide whether the trajectory can be flown at all.

This chapter derives, from the geometry of a thrust-vectored lander, the single frequency that bounds all three loops of an interactive planar landing simulation. The right-half-plane zero of the gimbal-to-position path and the natural frequency of the attitude loop are the same quantity $\sqrt{\ell T/I}$ apart from the square root of the attitude gain, so the bandwidth separation a cascade needs and the margin a non-minimum-phase plant allows are two readings of one number. Rigid-body flight dynamics and cascade design are both established methods, and the chapter applies them to one vehicle. Section 10.2 reviews thrust vector control and powered-descent guidance. Section 10.3 writes out a nine-state planar vehicle with its gimbal, engine, wind and touchdown models, with every equation, constant and initial condition needed to rebuild the simulator. Section 10.4 linearizes it about hover and places the three loops, and Section 10.5 documents the implemented loops and the optional predictive controller. Section 10.6 lists the scenarios, which break the cascade one mechanism at a time with a slow actuator, a saturated gimbal, a constant load and a shrinking mass, and Section 10.7 reports what they measure. Section 10.8 records defects of the application's own measurements that were found and corrected. Section 10.9 describes the real sensors, actuators and flight computer of such a vehicle. Sections 10.10, 10.11 and 10.12 state the limitations, place the results in practice and summarize them, and the exercises follow in Section 10.13.

10.2 Background

A rocket leaving a pad has no aerodynamic authority, because control surfaces need dynamic pressure and there is none at liftoff. The earliest large vehicles answered this with graphite vanes in the exhaust, which steered the vehicle at the cost of thrust, and the arrangement that replaced them mounts the whole combustion chamber on a pair of bearings and points it. Every large launch vehicle since has been steered that way, and the fact that the deflection acts through a lever arm below the center of mass is the reason the response begins in the wrong direction. The geometry, the moment it produces and the attitude dynamics it drives are standard material in space vehicle dynamics [10].

Landing under power is the same actuator run in reverse. The optimal thrust program for a soft lunar landing was set out as a minimum-propellant control problem in 1964, with the answer that the thrust magnitude should be at one bound or the other rather than in between [1]. The lunar module flew something more practical, a guidance law that shaped the acceleration as a polynomial in time-to-go and handed the resulting thrust direction to a digital autopilot, with the throttle and the attitude treated as separate channels [2]. That split between an outer law that owns the goal and an inner autopilot that owns the vehicle survives in every propulsive lander built since.

> ⧖ HISTORY **Deep Throttling for a Lunar Descent, 1963 to 1969**
>
> The descent stage of the Apollo lunar module had to carry the vehicle from lunar orbit to the surface and then hover over the last 1000 feet so that the crew could choose the exact landing site. That variable-thrust profile called for an engine that could be throttled, and the engine was also gimbaled to provide thrust vector control. Its design was first defined in 1963. Two throttling concepts were then developed in parallel for about 18 months: a fixed-area injector with helium injection, and a variable-area injector linked to cavitating flow-control valves. The variable-area design was chosen in January 1965, because the other suffered combustion instability and gave lower performance at reduced thrust. Development and qualification for the first lunar landing ran from August 1963 to April 1969 [12].

Mars entry, descent and landing renewed the interest and changed the mathematics. The powered descent problem was shown to admit a convex relaxation, which turns a nonconvex thrust-magnitude constraint into a second-order cone program with a proof that the relaxation is tight [3], and the argument was extended to pointing constraints and to the general soft-landing problem under the name lossless convexification [4]. Related work established the structure of the propellant-optimal solution directly [5], and the whole family is surveyed alongside the successive-convexification methods that handle six degrees of freedom [6]. The practical result is that a booster can compute a fresh optimal descent trajectory on board, several times a second, on the way down.

None of that removes the inner loops. A converged trajectory is a sequence of commanded thrust vectors, and a vehicle tracks a thrust vector by holding an attitude, which is a rate-limited actuator inside a feedback path. Rate limiting in a flight control loop is not a small correction, because it makes the effective loop gain depend on the size of the correction and it is a documented mechanism behind pilot-induced oscillation in piloted aircraft [13]. Successive loop closure, in which each loop is designed against a plant that already contains the closed inner loop, is the standard way to keep those interactions manageable, and its central requirement is a bandwidth separation between the levels [7]. Section 1.4.5 states the same requirement for a process cascade as an inner-loop speed rule. The application of this chapter is arranged so that the requirement can be violated deliberately and the consequence measured.

10.3 Vehicle Model

A 1200 kg single-stage lander, 8 m tall, descends in a vertical plane under one throttleable engine mounted on a gimbal 3 m below its center of mass. Horizontal position x and altitude z are measured in a pad-fixed frame with x positive to the right and z positive upward, and z is the height of the landing feet so that touchdown occurs exactly at $z = 0$. The tilt θ is measured from vertical toward $+x$, so the body axis is the unit vector $(\sin\theta, \cos\theta)$ and a positive tilt buys positive horizontal acceleration. The geometry and every symbol the equations refer to are drawn in Figure 10.1.

10.3.1 Variables and Signals

Three controlled variables are regulated by three loops, and only two of them have a setpoint an operator can type. Table 10.1 sorts every signal into its role with the value it holds at the start of the default vertical descent.

Two of the three loops are cascaded. The horizontal position loop writes θ_{SP}, the attitude loop reads it and writes δ_{cmd}, and the gimbal is the only actuator either of them ever touches. The descent-speed loop is separate and writes T_{cmd}, and it receives its setpoint from a guidance profile rather than from an operator, because a landing is graded on the speed at contact and not on the altitude history. Figure 10.2 draws the three loops with the limits and the actuator dynamics in the positions where they are applied.

🕮 **Pause.** Try Exercise 10.1 before reading on.

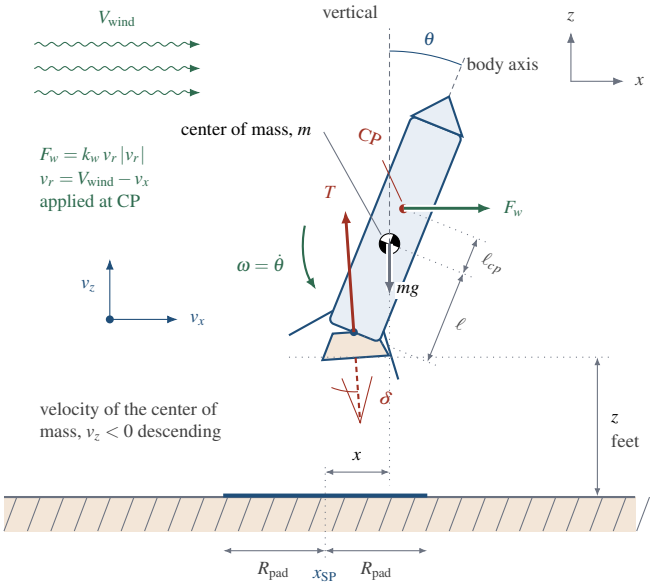


Figure 10.1 Planar landing vehicle with one gimballed engine. The body axis leans by the tilt θ from vertical toward $+x$, and both angles are drawn exaggerated against the 12° tilt limit and the 10° gimbal travel. Thrust T acts at the gimbal pivot a distance ℓ below the center of mass, so a deflection δ produces the moment $\ell T \sin\delta$ while the force itself points along $\theta - \delta$. The crosswind V_{wind} produces the side force F_w at a center of pressure ℓ_{cp} above the center of mass. Altitude z is measured to the landing feet and horizontal position x to the pad center x_{SP}, with the arrival graded against the pad radius R_{pad}.

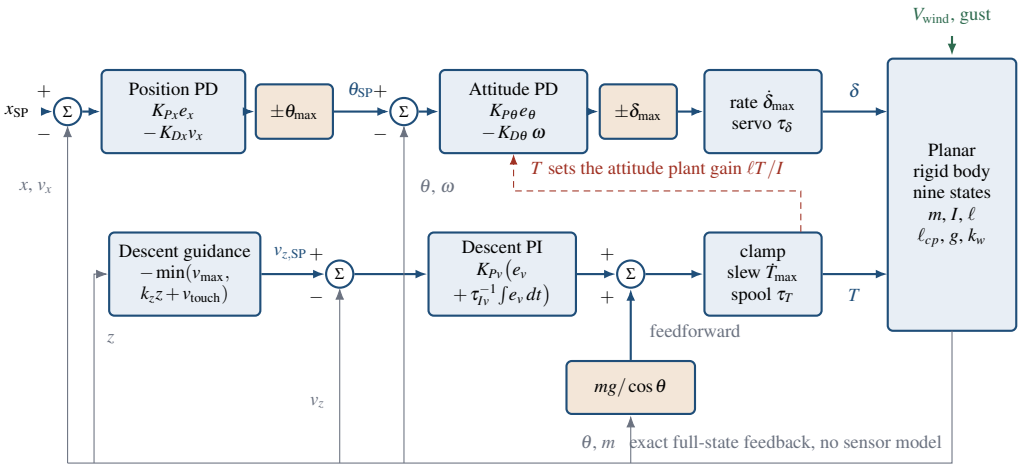


Figure 10.2 Three loops around one gimballed engine. The position loop has no actuator of its own, so its output is the attitude setpoint θ_{SP} that the attitude loop must deliver before any horizontal force exists. Both outer stages are proportional-derivative and memoryless, and only the descent-speed loop carries an integrator, biased by the tilt-corrected hover thrust $mg/\cos\theta$. The dashed path is the coupling the cascade does not remove, because the same thrust that holds altitude also sets the gain of the attitude plant.

Table 10.1 Classification of Every Signal. The Nominal Column Holds the Value at Ignition in the Vertical Descent Scenario.

Symbol	Role	Nominal	Range
x	state, horizontal position	0 m	pad ± 25 m
z	state, altitude of the feet	250 m	30 to 600 m
v_x	state, horizontal velocity	0 m/s	—
v_z	state, vertical velocity	−10 m/s	—
θ	CV, tilt from vertical	0°	—
ω	state, pitch rate	0 °/s	—
δ	MV, gimbal angle (servo state)	0°	$\pm\delta_{\max}$
T	MV, engine thrust (spool state)	11 772 N	$T_{\min}$ to $T_{\max}$
m	state, vehicle mass	1200 kg	m_{dry} to m_0
x_{SP}	setpoint, pad position	0 m	−25 to 25 m
θ_{SP}	MV of the position loop	0°	$\pm\theta_{\max}$
$v_{z,\text{SP}}$	guidance output, not typed	−10 m/s	—
δ_{cmd}	MV of the attitude loop	0°	$\pm\delta_{\max}$
T_{cmd}	MV of the descent loop	11 772 N	$T_{\min}$ to $T_{\max}$
V_{wind}	DV, steady crosswind	0 m/s	$\pm 15\,\text{m/s}$
v_{gust}	DV, gust amplitude	12 m/s	2 to 25 m/s
F_w	output, side force	0 N	—
$T\cos(\theta-\delta)$	output, vertical thrust	11 772 N	—
$mg/\cos\theta$	output, hover thrust	11 772 N	—

10.3.2 Translation and Rotation of the Planar Body

The vehicle is a rigid body with three degrees of freedom, two of translation and one of rotation. The thrust vector points along $\theta - \delta$ rather than $\theta + \delta$, for a reason derived in Section 10.3.3, and the crosswind adds a horizontal force. The translational balances are shown in Equation (10.1).

$$\begin{aligned} \dot{x} &= v_x, \qquad \dot{z} = v_z \\ m\dot{v}_x &= T\sin(\theta-\delta) + F_w \\ m\dot{v}_z &= T\cos(\theta-\delta) - mg \end{aligned} \tag{10.1}$$

where m is the vehicle mass in kg, T is the thrust in N, θ and δ are in rad, F_w is the aerodynamic side force in N and g is gravitational acceleration in m/s^2. Both lines carry units of N. The two equations share one T, which is the coupling the whole application is about, because $T\sin\theta$ spent on moving sideways is missing from $T\cos\theta$ holding the vehicle up.

> **+ DID YOU KNOW? A Throttle Misled by a Turning Vehicle**
>
> In the last 140 seconds of the first crewed lunar landing, the descent throttle swung about 15 percent peak to peak around a setting near 26 percent. Fast pitch maneuvers put centrifugal accelerations on accelerometers mounted high above the center of mass, and the guidance computer read them as a change in vertical acceleration [12].

Rotation is about the center of mass, driven by the gimbal moment and by the wind acting at a center of pressure above the center of mass, as shown in Equation (10.2).

$$\dot{\theta} = \omega, \qquad I\dot{\omega} = \ell T\sin\delta + \ell_{cp}F_w \tag{10.2}$$

where I is the pitch inertia in kg m^2, ℓ is the distance from the center of mass down to the gimbal pivot in m and ℓ_{cp} is the distance from the center of mass up to the center of pressure in m. Both

terms on the right carry units of N m. There is no aerodynamic restoring moment and no damping term of any kind, so Equation (10.2) is a pure double integrator driven by the gimbal. Placing the center of pressure above the center of mass is what makes a crosswind lean the vehicle downwind instead of merely pushing it sideways, and it is the reason the attitude loop has to hold a standing gimbal deflection in a steady wind.

10.3.3 Sign Convention for the Gimbal Deflection

The sign convention is the one statement in the model that is easy to get backward, and every claim about the wrong-way transient inverts with it. The engine sits at $\mathbf{r} = -\ell(\sin\theta, \cos\theta)$ relative to the center of mass. Evaluate the moment $\mathbf{r} \times \mathbf{F}$ at $\theta = 0$ with the thrust deflected by an angle ϕ from the body axis toward $+x$, so that $\mathbf{F} = T(\sin\phi, \cos\phi)$, as shown in Equation (10.3).

$$(\mathbf{r} \times \mathbf{F})_z = r_x F_z - r_z F_x = +\ell T \sin\phi \tag{10.3}$$

That is a counter-clockwise moment on a screen whose x axis points right, and counter-clockwise is the direction of decreasing θ. Deflecting the exhaust toward $+x$ therefore pitches the nose toward $-x$. A controller wants one number whose sign means point the nose this way, so the application defines $\delta = -\phi$, positive for a positive pitching moment. The moment is then $+\ell T \sin\delta$ as written in Equation (10.2), and the force necessarily points along $\theta - \delta$ as written in Equation (10.1). A deflection that brings the nose to the right pushes the vehicle left until the accumulated tilt is large enough for $T \sin\theta$ to overcome $-T \sin\delta$. The application carries a built-in test that evaluates the derivative function at $\theta = 0$ and $\delta = 5^\circ$ and checks that $\dot{\omega} > 0$ and $\dot{v}_x < 0$ in the same state, so the convention is pinned numerically rather than asserted.

10.3.4 Wind, Actuators and Propellant

The aerodynamic side load is quadratic in the air-relative lateral speed and is applied at the center of pressure, and Equation (10.4) gives it.

$$F_w = k_w\, v_r\, |v_r|, \qquad v_r = V_{\text{wind}} + c(t) - v_x \tag{10.4}$$

where k_w is the side-force coefficient in N/m^2 times s^2, V_{wind} is the steady crosswind in m/s and c is the decaying gust contribution in m/s. Writing $k_w = \frac{1}{2}\rho C_D A_{\text{ref}}$ makes the coefficient physical, and the shipped value of 10.6 N s^2/m^2 corresponds to $C_D A_{\text{ref}} = 17.3\,\text{m}^2$ at sea-level density. A gust event adds v_{gust} to c and the contribution then decays with time constant τ_{gust} through the exact discrete first-order hold $c \leftarrow c \exp(-\Delta t_c / \tau_{\text{gust}})$ evaluated once per control execution. There is no vertical drag, which for a base area consistent with the lateral coefficient above and an axial drag coefficient of 0.5 would be about 80 N at 10 m/s of descent against a 11 772 N weight.

Neither actuator can step. Both take a command that is rate limited first and then hand it to a first-order lag, which Equation (10.5) states.

$$\begin{aligned} \tau_\delta\, \dot{\delta} &= \delta_{\text{cmd}} - \delta, & \left|\dot{\delta}_{\text{cmd}}\right| &\le \dot{\delta}_{\max} \\ \tau_T\, \dot{T} &= T_{\text{cmd}} - T, & \left|\dot{T}_{\text{cmd}}\right| &\le \dot{T}_{\max} \end{aligned} \tag{10.5}$$

where τ_δ and τ_T are the servo and spool time constants in s. The gimbal command is clipped to $\pm\delta_{\max}$ and the thrust command to $[T_{\min}, T_{\max}]$ before the rate limit acts. The lower thrust bound matters more than the upper one, because $T_{\min}$ is 68 % of a hover: the engine cannot be shut down and relit, so descending faster than the profile asks is not something the controller can arrange by closing the throttle.

Propellant depletion is an option rather than a default and is shown in Equation (10.6).

$$\dot{m} = -\frac{T}{I_{\text{sp}}\, g_0}, \qquad m \ge m_{\text{dry}} \tag{10.6}$$

in which the specific impulse I_{sp} is in s and standard gravity g_0 in m/s^2. This is the constant exhaust-speed form of the variable-mass momentum balance [14], and at hover it removes 4.00 kg/s. The mass is clamped at m_{dry} after the integration step and a latching propellant-exhausted annunciation is raised, after which the model keeps thrusting, which is the optimistic case and is labeled as such in the application.

10.3.5 Guidance, Touchdown and Numerics

The vertical channel is given a descent-speed profile rather than an altitude setpoint, as shown in Equation (10.7).

$$v_{z,\mathrm{SP}} = -\min(v_{\max},\, k_z z + v_{\mathrm{touch}}) \tag{10.7}$$

where $v_{\max}$ is the cruise descent speed in m/s, k_z is the profile slope in 1/s and v_{touch} is the terminal descent speed in m/s. Above $z = (v_{\max} - v_{\mathrm{touch}})/k_z = 27.1$ m the command is a constant cruise and below it the descent is a first-order approach with time constant $1/k_z = 2.86$ s. The v_{touch} term is what makes contact happen in finite time, because with $v_{\mathrm{touch}} = 0$ the profile approaches the pad asymptotically and the vehicle never arrives.

Touchdown is a hard event applied after the integration step. At $z \le 0$ the arrival is classified once and the rigid body is then frozen, so a landed vehicle produces clean constant columns instead of numerical debris. The verdict is a tip-over if $|\theta| > \theta_{\mathrm{tip}}$, otherwise a hard landing if $|v_z| > v_{z,\max}$ or $|v_x| > v_{x,\max}$, otherwise a miss if the offset from the pad center exceeds R_{pad}, and otherwise a landing. Attitude is checked before speed, because landing legs care about the angle at contact and not about the average angle over the descent. The actuators keep integrating after contact so the engine can be seen spooling down, and the scenario run stops at the same instant, which means the application has no settled steady state to report for any scenario other than a pure vertical descent.

Every numerical value the model uses is listed in Tables 10.2 and 10.3. Integration is classical fourth-order Runge-Kutta over the nine states at a fixed 5 ms step, the controller executes every 10 ms, and the historian logs at 20 Hz. The fastest mode in the loop is the 80 ms gimbal servo, so the integration step is one sixteenth of it and the control period one eighth. The order of work in one step and the separation of rendering from the simulation are those of Section 1.6.1, with the touchdown correction applied after the integrator, so two machines produce identical trajectories and identical exports.

Every scenario starts trimmed on the descent profile, with $T(0) = m_0 g$ and $v_z(0) = v_{z,\mathrm{SP}}(z_0)$, so the opening seconds show a controller holding an equilibrium rather than recovering from a startup transient. Scenarios override individual initial conditions, and the disturbances available in any scenario are a crosswind slider spanning ± 15 m/s and three event buttons that inject a decaying gust, add 3 m/s of horizontal velocity, or add 8 °/s of pitch rate. Of the three, the velocity shove is the informative one. It moves neither the position nor the attitude, so at the instant it lands the only term anywhere in the cascade that responds is $-K_{Dx} v_x$.

10.3.6 Assumptions

Only mechanisms that change the tuning survive in the model. Each assumption below is given with the range in which it holds and the situation that breaks it.

Planar rigid body. Three degrees of freedom, fixed inertia, no yaw, no roll and no propellant slosh. The assumption holds for the small excursions the tilt limit permits. It breaks on a vehicle with a part-full tank, where slosh is a lightly damped mode close in frequency to the attitude loop, and it breaks for any out-of-plane disturbance.

Exact full-state feedback. The controller reads x, z, v_x, v_z, θ and ω with no noise, no bias, no quantization and no lag. This is the largest omission in the application. A real vehicle infers those states from an inertial measurement unit and a radar altimeter, and navigation error rather than tuning is what sets the achievable touchdown precision. Every result quoted here should be read as the tuning-limited case.

Table 10.2 Vehicle, Engine and Environment Constants. Plant Parameters Are Adjustable Over the Range Given, Apart From the Three Marked Hidden, Which the Parameter Panel Does Not Show.

Symbol	Meaning	Value	Unit	Range
Vehicle				
m_0	wet mass at ignition	1200	kg	1000-1600
m_{dry}	dry mass (hidden)	900	kg	600-1200
I	pitch inertia	4500	kg m^2	1500-12000
ℓ	center of mass to gimbal	3.0	m	1.0-6.0
ℓ_{cp}	center of mass to CP	1.2	m	0.0-3.0
L	body length (hidden)	8.0	m	5-14
g	surface gravity	9.81	m/s^2	8.0-11.0
Engine and gimbal				
$T_{\max}$	maximum thrust	20000	N	13000-26000
$T_{\min}$	minimum thrust	8000	N	4000-10000
τ_T	spool time constant	0.15	s	0.02-0.80
$\dot{T}_{\max}$	throttle slew limit	12000	N/s	2000-40000
I_{sp}	specific impulse	300	s	200-400
g_0	standard gravity (hidden)	9.80665	m/s^2	fixed
$\delta_{\max}$	gimbal travel limit	10.0	°	1.0-10.0
τ_δ	gimbal time constant	0.08	s	0.01-0.60
$\dot{\delta}_{\max}$	gimbal rate limit	60	°/s	5-120
Environment and mission				
k_w	side-force coefficient	10.6	N/m^2 s^2	0-30
v_{gust}	gust amplitude	12	m/s	2-25
τ_{gust}	gust decay time	2.5	s	0.5-10
z_0	ignition altitude	250	m	30-600
R_{pad}	landing pad radius	8	m	3-25
$v_{z,\max}$	touchdown speed limit	2.0	m/s	0.5-6
$v_{x,\max}$	lateral speed limit	1.0	m/s	0.2-4
θ_{tip}	tip-over angle	6	°	1-15

One thrust number with a lag and a slew limit. Combustion dynamics, throttle-dependent specific impulse, ignition transients and the possibility of an engine that will not relight are all absent. The first-order spool is a fair representation between about 40 % and 100 % throttle and understates the difficulty at the bottom of the range.

Constant inertia under propellant burn. Mass falls and I does not, which is the optimistic case. The rotational equations of an axisymmetric body losing mass through a nozzle carry inertia terms that change with the propellant distribution and a jet-damping contribution that this model omits [15]. A real vehicle emptying tanks near its base loses inertia faster than it loses mass, so its attitude loop gets faster as it descends rather than slower, and a gain that is well damped at ignition can be marginal at touchdown.

One quadratic side force at a fixed center of pressure. There is no angle-of-attack dependence, no vertical drag, no Mach dependence and no plume-ground interaction in the last few meters. Ground effect in particular is strong, unsteady and exactly where the precision requirement is tightest.

Instantaneous inelastic touchdown. Leg stroke, damping and crush cores are absent, so the boundary between a landing and a hard landing is a threshold rather than a structural calculation, and the pad is flat, level and stationary. On a ship deck or a slope both assumptions fail at the moment they matter most.

Table 10.3 Controller Settings, Values Derived at the Hover Operating Point and Numerical Constants. An (adv.) Marking Means the Control Appears Only Under the Advanced Toggle.

Symbol	Meaning	Value	Unit	Range
Controller				
K_{Px}	position gain	0.72	°/m	0-3
K_{Dx}	position rate gain	4.10	° per m/s	0-12
$K_{P\theta}$	attitude gain	1.20	°/°	0-5
$K_{D\theta}$	attitude rate gain	0.65	s	0-3
K_{Pv}	descent-speed gain	1500	N per m/s	0-6000
τ_{Iv}	descent integral time	2.5	s	0.2-60
$\theta_{\max}$	tilt limit	12.0	°	1-15 (adv.)
$v_{\max}$	cruise descent speed	10.0	m/s	2-25 (adv.)
k_z	profile slope	0.35	1/s	0.05-1.2 (adv.)
v_{touch}	terminal descent speed	0.50	m/s	0-2 (adv.)
Derived at the hover operating point				
mg	hover thrust	11772	N	—
—	hover throttle	31.43	%	—
a	attitude plant gain $\ell T/I$	7.848	$1/s^2$	—
$\sqrt{a}$	right-half-plane zero	2.801	rad/s	—
$\omega_{n\theta}$	attitude natural frequency	3.069	rad/s	—
ζ_θ	attitude damping ratio	0.831	—	—
ω_{nx}	position natural frequency	0.351	rad/s	—
ζ_x	position damping ratio	1.000	—	—
ω_{nv}	descent natural frequency	0.707	rad/s	—
ζ_v	descent damping ratio	0.884	—	—
Numerics				
Δt	integration step (RK4)	0.005	s	—
Δt_c	control period	0.010	s	—
—	historian rate	20	Hz	—

10.4 Linearized Dynamics and Loop Bandwidths

Every quantitative statement in this section comes from linearizing Equations (10.1) and (10.2) about hover, with θ and δ small, $T \approx mg$ constant and no wind. That linearization is the analysis model. It is used to place the gains and to explain the measurements, and it is used nowhere inside the application.

10.4.1 The Attitude Plant

With $\sin\delta \approx \delta$ and the wind torque removed, Equation (10.2) becomes the pure double integrator of Equation (10.8).

$$\ddot{\theta} = a\,\delta, \qquad a \equiv \frac{\ell T}{I} = 7.848/\mathrm{s}^2 \tag{10.8}$$

where a is the attitude plant gain in $1/s^2$, evaluated at $T = m_0 g = 11\,772\,\mathrm{N}$. There is no restoring moment and no damping, so proportional action supplies the entire spring and derivative action supplies the only damper the vehicle has. Substituting the attitude law $\delta = K_{P\theta}(\theta_{\mathrm{SP}} - \theta) - K_{D\theta}\omega$ gives the closed-loop pair of Equation (10.9).

$$\omega_{n\theta} = \sqrt{a\,K_{P\theta}}, \qquad \zeta_\theta = \frac{a\,K_{D\theta}}{2\,\omega_{n\theta}} \tag{10.9}$$

At $K_{P\theta} = 1.20$ and $K_{D\theta} = 0.65\,\mathrm{s}$ this places the attitude loop at 3.069 rad/s with a damping ratio of 0.831. Both gains are dimensionally clean, because $K_{P\theta}$ is an angle per angle and $K_{D\theta}$ is a time.

10.4.2 The Right-Half-Plane Zero of the Gimbal-to-Position Path

Horizontal acceleration comes from the direction of the thrust vector, which is $\theta - \delta$ and not θ. Linearizing the first line of Equation (10.1) gives $\dot{v}_x = g(\theta - \delta)$, and eliminating θ with Equation (10.8) produces the open-loop transfer from gimbal angle to horizontal position shown in Equation (10.10).

$$\frac{x(s)}{\delta(s)} = \frac{g\left(a - s^2\right)}{s^4} \tag{10.10}$$

The path is a quadruple integrator with a zero pair at $s = \pm\sqrt{a}$, one of which sits in the right half plane at $\sqrt{a} = 2.801\,\text{rad/s}$. The physical content of that zero is the wrong-way transient, the same start in the wrong direction that a cart makes under a balanced pole in Volume 7, Chapter 7. For a step deflection δ applied to the linearized model the position response is $x(t) = g\delta(at^4/24 - t^2/2)$, so the horizontal acceleration changes sign at $t = \sqrt{2/a} = 0.505\,\text{s}$, the excursion reaches its extreme of $1.5\,g\delta/a$ at $t = \sqrt{6/a} = 0.874\,\text{s}$, and the vehicle returns through its starting point at $t = \sqrt{12/a} = 1.237\,\text{s}$. At the 10° travel limit that extreme is 0.33 m in the wrong direction.

Closing the attitude loop does not help. Substituting the same attitude law into Equation (10.10) and treating θ_{SP} as the input gives Equation (10.11).

$$\frac{x(s)}{\theta_{\text{SP}}(s)} = \frac{g\,K_{P\theta}\left(a - s^2\right)}{s^2\left(s^2 + 2\zeta_\theta\omega_{n\theta}s + \omega_{n\theta}^2\right)} \tag{10.11}$$

The numerator is unchanged. Feedback moves poles and leaves zeros where they are, so the right-half-plane zero of the vehicle survives every attitude gain that could be dialed in, and the only quantities that move it are ℓ, T and I. A shorter engine bay, a heavier engine or a fatter vehicle changes the achievable position bandwidth. Retuning does not. Chapter 9 meets a right-half-plane zero of the same kind in an electrical cascade, where it belongs to the power stage and moves with load and duty cycle.

🕮 **Pause.** Try Exercise 10.2 before reading on.

10.4.3 Bandwidth Separation and the Zero Margin

A real right-half-plane zero at z limits the achievable crossover of a loop that must cross it to roughly $z/2$, because the phase the zero subtracts cannot be recovered [11]. Here that ceiling is 1.40 rad/s. The cascade imposes a second ceiling, that the position loop must be slow compared with the attitude loop, which is the inner-loop speed rule of Section 1.4.5 read in frequency. What makes this particular vehicle instructive is that the two ceilings are the same quantity in disguise, as shown in Equation (10.12).

$$\frac{\omega_{n\theta}}{\sqrt{a}} = \sqrt{K_{P\theta}} = 1.095 \tag{10.12}$$

The attitude loop sits within 10 % of the right-half-plane zero, and it does so at every attitude gain of order unity, because the ratio grows only as the square root. Making the inner loop four times faster in bandwidth costs sixteen times the attitude gain and moves the zero not at all.

At low frequency Equation (10.11) reduces to $x/\theta_{\text{SP}} \to g/s^2$, a double integrator with gain g, and the position law $\theta_{\text{SP}} = K_{Px}(x_{\text{SP}} - x) - K_{Dx}v_x$ closes it as shown in Equation (10.13).

$$\omega_{nx} = \sqrt{g\,K_{Px}}, \qquad \zeta_x = \frac{g\,K_{Dx}}{2\,\omega_{nx}} \tag{10.13}$$

At $K_{Px} = 0.72°/\text{m}$ (0.012 57 rad/m) and $K_{Dx} = 4.10°$ per m/s (0.071 56 rad s/m) the position loop lands at 0.351 rad/s with a damping ratio of 1.000, exactly critical. That is a factor of 8.7 below the attitude loop and a factor of 8.0 below the right-half-plane zero, and the two factors agree because Equation (10.12) says they must. The plant gain in Equation (10.13) is $g = 9.81\,\text{m/s}^2$ per rad of commanded tilt, which is not a controller gain and does not change with tuning, and the controller gain K_{Px} carries units of rad/m.

10.4.4 The Descent Loop and Its Feedforward

The vertical channel sees $m\dot{v}_z = T\cos(\theta-\delta) - mg$, which for small angles is $m\dot{v}_z = T - mg$, an integrator driven by the excess thrust. Proportional-integral action closes it as shown in Equation (10.14).

$$\omega_{nv} = \sqrt{\frac{K_{Pv}}{m\,\tau_{Iv}}}, \qquad \zeta_v = \frac{K_{Pv}}{2\,m\,\omega_{nv}} \tag{10.14}$$

At $K_{Pv} = 1500\,\mathrm{N}$ per m/s and $\tau_{Iv} = 2.5\,\mathrm{s}$ against 1200 kg this gives 0.707 rad/s at a damping ratio of 0.884, and the proportional term alone would give a velocity time constant of $m/K_{Pv} = 0.80\,\mathrm{s}$. Both are comfortably inside the 2.86 s guidance time constant the loop has to follow.

This is the only loop carrying an integrator and the only one that needs one, because it is the only loop with a constant load. That load is the weight, and handing it to the loop as a feedforward bias, a static feedforward from a known force in the sense of Section 1.4.3, rather than making the integrator discover it is what keeps the vertical channel quiet. Tilting steals vertical thrust as $\cos\theta$, so the bias is $mg/\cos\theta$, which is 2.23 % above mg at the 12° tilt limit, an extra 263 N out of 11 772 N. The same arithmetic explains why the tilt limit is where it is. At 30° the correction is 15.5 %, or 1821 N, and an engine limited to 14 kN can supply it only by giving up its deceleration margin, because 14 kN leaned 30° leaves 352 N of spare vertical force where 2228 N was available upright.

10.4.5 Offset in a Steady Crosswind

Neither cascaded loop has integral action, so a constant side load leaves a standing position error, and the chain that sets its size runs backward through all three levels. A 7 m/s crosswind produces $F_w = 519\,\mathrm{N}$ from Equation (10.4). Canceling that force at hover thrust requires $\sin(\theta-\delta) = -519/11772$, or −2.53°. Canceling the wind moment $\ell_{cp}F_w = 623\,\mathrm{N\,m}$ requires $\sin\delta = -623/(\ell\,mg)$, a standing gimbal deflection of −1.01°. The attitude therefore has to sit at $-2.53 - 1.01 = -3.54°$. Producing −1.01° of gimbal from a proportional gain of 1.20 requires −0.84° of attitude error, so $\theta_{\mathrm{SP}} = -4.38°$, and producing that command from $K_{Px} = 0.72°/\mathrm{m}$ requires 6.09 m of standing position error. Section 10.7 compares each link of that chain with what the application holds.

🕮 **Pause.** Try Exercise 10.3 before reading on.

10.5 Controller Design

All three loops execute in one function every $\Delta t_c = 10\,\mathrm{ms}$, which is a 100 Hz flight computer and twice as slow as the physics step. The original cascade carries no optimizer, horizon or internal vehicle model. The predictive option added to it is described in Section 10.5.4.

10.5.1 The Two Cascaded Loops

Both cascaded loops are proportional-derivative in the explicit position-plus-rate form of Equations (10.15) and (10.16), evaluated on the state vector directly.

$$\begin{aligned} \theta_{\mathrm{SP}}^{\mathrm{raw}} &= K_{Px}\,(x_{\mathrm{SP}} - x) - K_{Dx}v_x \\ \theta_{\mathrm{SP}} &= \mathrm{clamp}(\theta_{\mathrm{SP}}^{\mathrm{raw}},\, -\theta_{\max},\, \theta_{\max}) \end{aligned} \tag{10.15}$$

$$\begin{aligned} \delta_{\mathrm{cmd}}^{\mathrm{raw}} &= K_{P\theta}\,(\theta_{\mathrm{SP}} - \theta) - K_{D\theta}\,\omega \\ \delta_{\mathrm{tgt}} &= \mathrm{clamp}(\delta_{\mathrm{cmd}}^{\mathrm{raw}},\, -\delta_{\max},\, \delta_{\max}) \end{aligned} \tag{10.16}$$

Neither loop has an integral term, a derivative filter or any internal state whatsoever. The derivative terms use v_x and ω, which are states of the simulation rather than numerical derivatives of a measurement, so no filter time constant appears and no derivative kick is possible on a setpoint move. Both loops are therefore memoryless functions of the state, so neither needs initializing when control changes hands, although a transfer to automatic operation moves the gimbal to whatever the attitude law computes at that instant. Integral action is deliberately absent from the attitude loop, because a

reset term inside the fastest loop in the cascade spends the phase margin that loop has least of, and it is deliberately absent from the position loop, because an integrator in the outer loop of a cascade whose inner loop saturates is a windup trap. The consequence of the second choice is the steady offset derived above, and the application argues for a wind estimate biasing x_{SP} rather than for an outer integrator.

The gimbal command published to the actuator is the rate-limited version of δ_{tgt} shown in Equation (10.17).

$$\delta_{\mathrm{cmd}} \leftarrow \mathrm{slew}\Big(\delta_{\mathrm{tgt}},\, \delta_{\mathrm{cmd}},\, \dot{\delta}_{\max},\, \Delta t_c\Big) \tag{10.17}$$

where slew moves the previous command toward the target by at most $\dot{\delta}_{\max}\Delta t_c$. The servo lag of Equation (10.5) then acts on the result, so the rate limit and the lag are separate mechanisms applied in series and can be varied independently.

10.5.2 The Descent Loop

The vertical loop uses the velocity-form controller of Section 1.6.2 [16] with the derivative time fixed at zero, so the proportional and integral increments are the only ones active. In this loop's symbols the recursion is Equations (10.18) to (10.24), where b_k, s_k and c_k are the u_b, S_k and v_k of Equations (1.10) and (1.11), and the bias b_k changes from scan to scan.

$$e_k = v_{z,\mathrm{SP},k} - v_{z,k} \tag{10.18}$$

$$b_k = \mathrm{clamp}\left(\frac{\chi_{\mathrm{FF}}\, m_k\, g}{\max(\cos\theta_k,\, 0.25)},\, T_{\min},\, T_{\max}\right) \tag{10.19}$$

$$\Delta u_k^P = K_{Pv}\,(e_k - e_{k-1}), \qquad \Delta u_k^I = \frac{K_{Pv}}{\tau_{Iv}}\, e_k\, \Delta t_c \tag{10.20}$$

$$s_k = s_{k-1} + \Delta u_k^P \tag{10.21}$$

$$c_k = b_k + s_k + I_{k-1} \tag{10.22}$$

$$u_k = \mathrm{clamp}(c_k,\, T_{\min},\, T_{\max}) \tag{10.23}$$

$$I_k = I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{\tau_t}\,(u_k - c_k) \tag{10.24}$$

where e_k is the descent-speed error in m/s, b_k is the gravity feedforward bias in N, c_k is the thrust the algorithm demands and u_k the thrust it is allowed to command, both in N, s_k accumulates the proportional increments without limit, I_k holds the integral contribution in N, $\tau_t = \tau_{Iv}$ is the tracking time of the back-calculation in s, and the flag $\chi_{\mathrm{FF}} \in \{0, 1\}$ turns the bias on and off. The integral used at scan k was accumulated through scan $k-1$, and the update of Equation (10.24) runs only when $K_{Pv}/\tau_{Iv} \neq 0$. While the command is clipped, the term $(\Delta t_c/\tau_t)(u_k - c_k)$ drives I toward the value that reproduces the applied thrust, removing $1/250$ of the clipped excess per scan at the default $\tau_{Iv} = 2.5\,\mathrm{s}$, and the running sum s_k is never truncated, as Section 1.6.2 explains. The tracking time affects no result reported here. The only shipped case that clips the thrust command, scenario 7, stays on the upper limit until contact, and a counterfactual run with a tracking time of one scan leaves its arrival speed unchanged at 3.59 m/s. The published thrust command is the rate-limited version of u_k, $T_{\mathrm{cmd},k} = \mathrm{slew}(u_k, T_{\mathrm{cmd},k-1}, \dot{T}_{\max}, \Delta t_c)$, so the rate limit acts once and outside the controller.

The loop is reverse acting in the ISA sense, because more thrust raises v_z, so a climb rate above the commanded value has to reduce thrust. The floor of 0.25 on $\cos\theta$ caps the feedforward past 75.5° of tilt, which is well outside anything the tilt limit permits and exists only to keep the expression finite if a scenario is edited into a tumble.

Equation (10.19) clamps the bias to the engine range, so switching the feedforward off leaves $b_k = T_{\min} = 8\,\mathrm{kN}$ rather than zero and the integral has to build only the remaining 3.77 kN of hover thrust. When the switch is thrown during a run, the application shifts I_k by $mg/\cos\theta - T_{\min}$ (with $mg/\cos\theta$ clamped to the engine range), which is the difference between the two biases. That holds

the thrust command continuous across the switch and preloads the integral with the hover thrust the bias no longer supplies. A switch made before the run is a starting condition instead. The controller is built with $I = 0$ against whichever bias is selected, so a run started with the feedforward off begins with no preload at all, and Section 10.7 measures what that costs.

10.5.3 Manual Operation and the Three Models

In manual operation the throttle and gimbal commands T_{cmd} and δ_{cmd} are operator-set inputs, and a hover-trim helper computes $mg/\cos\theta$ at the present mass and tilt, applies it as the throttle and centers the gimbal. All three loops keep computing throughout. The two proportional-derivative loops are memoryless and the vertical loop is fed the operator's throttle through the tracking step of Section 1.6.2, which back-solves I_k so that c_k of Equation (10.22) reproduces the slider exactly. The throttle slider starts every run at the trim throttle $(T_0 - T_{\min})/(T_{\max} - T_{\min})$, with $T_0 = m_0 g$ clamped to the engine range, so a vehicle taken into manual operation before the run holds its trimmed start whatever engine the scenario fits. Automatic operation drives both sliders from the commands after every controller execution, so a transfer to manual moves nothing. A transfer to automatic is bumpless for the thrust, as a by-product of the tracking arithmetic, but not for the gimbal, whose target jumps from the operator's setting to the output of Equation (10.16) and is then limited only by the slew rate. After 1 s of manual flight in scenario 2 with the gimbal centered, the switch moves the gimbal command from 0° to its 10° limit within 0.2 s.

Three models are kept apart throughout this chapter. The simulation model is Equations (10.1) to (10.6), nine nonlinear states integrated at 5 ms. The analysis model is the hover linearization of Section 10.4, used to place gains and to interpret results and present nowhere in the code. The controller's internal model is two algebraic relations only, the hover thrust $mg/\cos\theta$ and the guidance profile of Equation (10.7). The controller never sees ℓ, I, k_w or ℓ_{cp}, so it knows nothing about the right-half-plane zero it is working against.

10.5.4 Joint Predictive Thrust and Gimbal Requests

An additional MPC, which solves the shared quadratic program of Section 1.6.3 on a model of its own, relinearizes the nominal planar force and torque balances at each decision. Its eight states are the six ideal kinematic readings and independently integrated gimbal and thrust. Geometry, inertia, gravity and servo constants use declared defaults. Nominal mass begins at 1200 kg and integrates estimated thrust over nominal $I_{sp}g_0$ when burn is enabled, with a 900 kg floor. Actual mass and servo states are not inputs. For example, an actual 1440 kg initial mass leaves this estimate 240 kg low. Measured velocity/rate differences correct additive horizontal, vertical and angular acceleration biases, with gain 0.1 and bounds 6, 4 m/s^2 and 2 rad/s^2. No wind or future event is read. This is disturbance correction [17], not inertia or mass identification. The nominal observer tracks applied requests under either selected law.

Every 0.1 s, the local affine model is exactly discretized for 60 steps, 6 s, in twelve move blocks. Future vertical-speed references integrate the existing guidance profile from current altitude. The shared convex QP [18] tracks $[x, v_x, v_z, \theta, \omega]$ with weights $[0.3, 2, 10, 100, 10]$ and terminal weights $[2, 4, 10, 300, 30]$. With thrust in kN and gimbal in rad, absolute input weights are $[0.001, 0.05]$ and move weights $[0.02, 2]$. Request boxes use current actuator limits. Block moves are at most 1.2 kN and 2°, further restricted by the physical slew allowances. Tilt is soft, priced quadratically at 10^5 and linearly at 100. Actual command slew and touchdown idle, centering and freeze remain. The vertical PI tracks applied post-slew thrust once per scan. A rejected plan holds both requests on handover, marks fallback and returns to Cascade until recovery.

The predictor matches nominal zero-aeroload derivatives exactly. Over one 0.1 s interval at tilts up to ± 0.2 rad, affine prediction errors are below 0.000 468 m/s and 13.4 µm against nonlinear integration. An actual two-input one-step optimizer check agrees with hand inversion within 3.5×10^{-18}. These are local checks. Mass and acceleration bias are held inside a forecast, command-slew transients are

Table 10.4 Scenario Set. Every Entry Not Listed Under Change Keeps Its Default Value From Tables 10.2 and 10.3.

#	Change	Mechanism isolated
1	none	the guidance profile and its corner
2	$x(0) = -20\,\text{m}$	three timescales and the wrong-way transient
3	$v_x(0) = 6\,\text{m/s}$	the rate term as the whole command
4	$\theta(0) = 10^\circ$	cross-coupling met by feedforward
5	$V_{\text{wind}} = 7\,\text{m/s}$	offset from proportional action
6	$z_0 = 80\,\text{m}$, 16 m/s gust at 6 s	time-to-go against loop time constant
7	$T_{\max} = 14\,\text{kN}$	deceleration authority, not hover authority
8	$\tau_\delta = 0.18\,\text{s}$, $\dot{\delta}_{\max} = 30^\circ/\text{s}$, $x(0) = -15\,\text{m}$	an inner loop faster than its actuator
9	$\delta_{\max} = 3^\circ$, $x(0) = -25\,\text{m}$, $v_x(0) = -2\,\text{m/s}$	a saturated inner loop opens the cascade
10	propellant burn on, $z_0 = 320\,\text{m}$, $x(0) = -12\,\text{m}$	three loops drifting at different rates
11	MPC, $z_0 = 100\,\text{m}$, $x(0) = -15\,\text{m}$, $v_x(0) = 2\,\text{m/s}$, $\theta(0) = 3^\circ$, wind 2 m/s, gust at 4 s	joint bounded thrust and gimbal prediction
12	graded, see Section 10.7.10	everything at once, from poor tuning

omitted, and there is no ground-contact, glide-cone or terminal landing guarantee. The method tracks guidance rather than minimizing propellant.

10.6 Scenarios

Table 10.4 lists eleven teaching scenarios and one graded challenge. Each fixes only the quantity it exists to demonstrate and inherits every other setting, so whatever a run shows is attributable to one change. All twelve open in automatic operation, and an operator is free to take manual control at any point. Each case is credited when the learner performs its listed actions and the simulation reaches its listed state, as Section 1.6.4 describes, and the challenge is judged on the case as given, so changing a model parameter or a disturbance voids the attempt.

10.7 Results

The numbers below come from the scorecard the application computes for itself and from the file it exports, with the physics, the controller and the historian running at the rates in Table 10.3. A run ends at touchdown, so the integrated measures cover the whole descent rather than a fixed window, and the descent lasts 32.45 s from 250 m under the default profile. Every setting named below is changed after the scenario is selected and before the run starts, and the application gives the same run when the setting is carried through Reset instead. The model has no random input, so each run repeats exactly. Table 10.5 collects the outcomes. Figure 10.3 draws the arrivals of that table against the pad radius and the touchdown speed limit.

> ✦ **DID YOU KNOW? Take Away the Rate Term and It Tips Over**
>
> From the simulator: scenario 3 starts the vehicle drifting at 6 m/s, and it lands 0.01 m from the pad center. With the position rate gain K_{Dx} set to zero, the same start tips over 14.10 m from the pad, and the position integral grows from 77.7 to 348.3 m s (rows 3 and 3b of Table 10.5).

10.7.1 A Joint Plan and Its Model-Mismatch Limits

Case 11 starts 100 m high, 15 m left, with 2 m/s rightward drift, 3° tilt and 2 m/s wind. A gust arrives at 4 s, without preview to either controller. Figure 10.4 and Table 10.6 use the same seed and ideal measurements. MPC improves offset and precontact thrust impulse, but contacts faster vertically and moves the gimbal more. Impulse is not a fuel-optimality result. A six-point position-PD hindsight

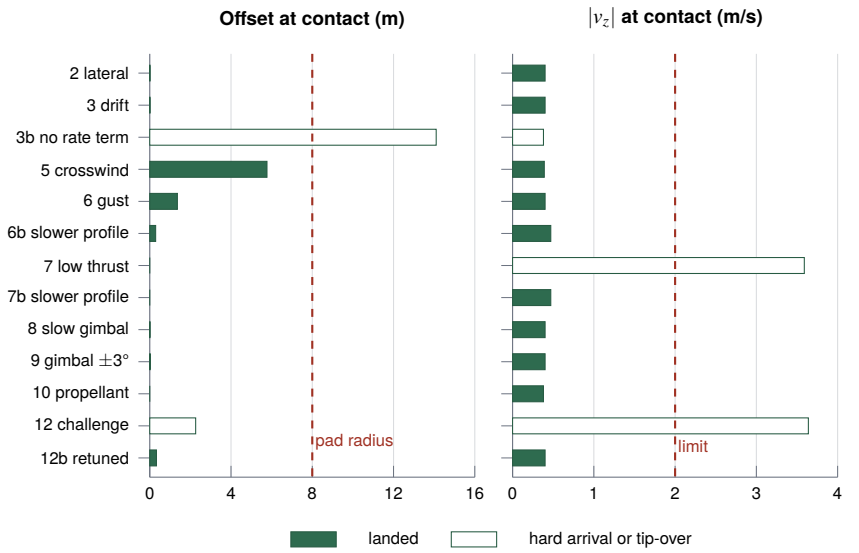


Figure 10.3 Table 10.5 redrawn for thirteen of its rows: the vertical descent, the tilted start with and without feedforward, the two gimbal variants of scenario 8 and the 1° gimbal of row 9b are left out (row 9b arrives 42.49 m away at 57.77 m/s). Filled bars are landings and open bars are hard arrivals or tip-overs. The dashed lines are the 8 m pad radius and the 2.0 m/s touchdown limit. Notice that the failures split into two kinds: row 3b tips over far from the pad at a gentle vertical speed, while rows 7 and 12 arrive too fast, and each is repaired by a different change (rows 3, 7b and 12b).

Table 10.5 Measured Outcomes of the Original Scenarios. Each IAE Is the Integral of the Absolute Error of That Loop Over the Whole Descent, Offset Is the Distance From the Pad Center at Contact, and the Verdict Is the Application's Own Classification of the Arrival.

#	Set-up	Duration (s)	x IAE (m s)	θ IAE (rad s)	v_z IAE (m)	Peak tilt (°)	Offset (m)	$\lvert v_z \rvert$ at contact (m/s)	Verdict
1	vertical descent	32.45	0	0	6.67	0.0	0.00	0.40	landed
2	20 m lateral	32.45	123.0	0.264	6.84	11.4	0.01	0.40	landed
3	6 m/s drift	32.45	77.7	0.255	6.85	12.2	0.01	0.40	landed
3b	the same with $K_{Dx} = 0$	32.46	348.3	0.612	6.94	12.0	14.10	0.38	tip-over
4	10° initial tilt	32.45	8.7	0.195	6.83	10.0	0.00	0.40	landed
4b	the same, feedforward off	31.83	7.9	0.199	12.94	10.0	0.00	0.40	landed
5	7 m/s crosswind	32.46	152.2	0.475	6.68	3.6	5.77	0.39	landed
6	gust at 20 m	15.46	21.6	0.155	6.74	7.0	1.36	0.40	landed
6b	the same with $k_z = 0.15$	22.32	29.4	0.185	3.80	7.9	0.29	0.47	landed
7	$T_{\max} = 14$ kN	26.11	0	0	9.96	0.0	0.00	3.59	hard
7b	the same with $k_z = 0.15$	39.32	0	0	3.80	0.0	0.00	0.47	landed
8	slow gimbal	32.45	89.4	0.404	7.17	14.9	0.01	0.40	landed
8a	the same, nominal gimbal	32.45	89.0	0.233	6.83	10.2	0.01	0.40	landed
8b	the same with $K_{P\theta} = 0.60$	32.45	89.9	0.406	6.73	7.9	0.01	0.40	landed
9	$\delta_{\max} = 3°$	32.45	207.6	0.315	6.79	12.0	0.02	0.40	landed
9b	the same with $\delta_{\max} = 1°$	17.39	333.0	10.57	78.37	135.4	42.49	57.77	tip-over
10	propellant depletion	39.30	70.6	0.185	5.95	8.2	0.00	0.38	landed
12	graded, as shipped	22.82	148.4	2.081	11.97	23.8	2.26	3.64	hard
12b	graded, retuned	29.45	99.6	0.246	6.76	5.9	0.33	0.40	landed

Table 10.6 Contact and Precontact Maneuver, Cascade / MPC

Case	Offset m	$\lvert v_z \rvert$ m/s	Time s	Impulse kN s
Case 11	0.726 / -0.079	0.395 / 0.519	17.45 / 15.52	217.2 / 195.0
Lateral	-0.011 / -0.000	0.395 / 0.518	32.45 / 30.58	394.0 / 372.2
Low thrust	0.000 / 0.000	3.593 / 0.518	26.11 / 30.41	315.1 / 369.4
Slow gimbal	-0.008 / -2.346	0.395 / 22.282	32.45 / 15.97	394.5 / 216.9

Last row MPC tips over. New-case gimbal travel: 13.7 / 87.2 degrees. MPC plans / mean solve time in microseconds: case 11 156 / 746, lateral 306 / 500, low thrust 305 / 435, slow gimbal 160 / 816. No rejected plans in these runs.

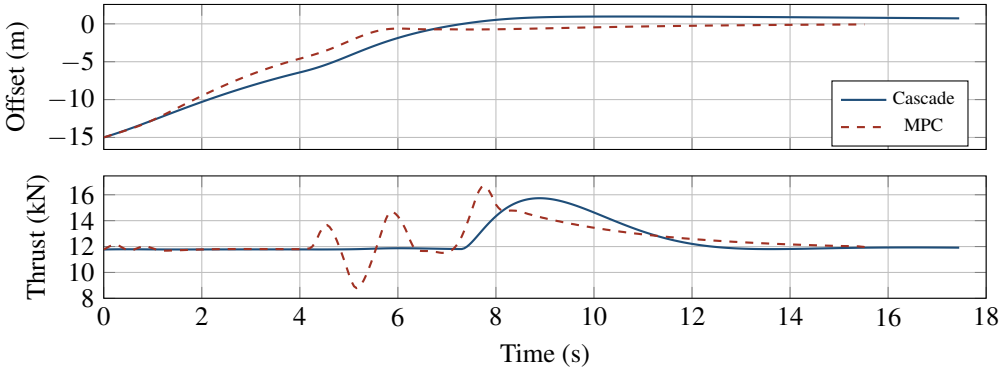


Figure 10.4 Case 11, identical descent and unknown gust at 4 s, with the Cascade as a solid line and the MPC as a dashed one. Curves end at each controller's contact. Lower impulse and offset under MPC accompany faster vertical arrival and more gimbal travel, not a universal landing gain.

grid, three simpler control changes and two isolated bias-feedforward Cascade diagnostics complete the 43-row measurement record behind this section. The finer physics step preserves this offset and impulse ranking.

Under steady crosswind, bias-feedforward Cascade reaches 0.0308 m offset, compared with MPC –0.0363 m and original Cascade 5.771 m. Thus removing offset alone does not require a horizon. With reduced thrust, MPC lands instead of arriving hard, but uses 369.36 versus 315.06 kN s impulse. Conversely, the slow-gimbal case tips over under MPC at 22.28 m/s vertical speed while Cascade lands. Doubling inertia in the new case also tips MPC over while Cascade lands. Both suggested challenge controllers pass, MPC at −0.0248 m and Cascade at 0.3341 m, with gimbal travel 86.9 versus 29.1°. Local model error can dominate nominal benefits.

In a browser, measured over six window sizes and color themes, 100 plans average 0.96–1.54 ms. The planner then takes 0.96–1.54% of one processor core at real-time playback and 3.83–6.14% at four-times speed, excluding the nominal observer and rendering.

10.7.2 Lateral Displacement and the Wrong-Way Transient

Scenario 2 starts the vehicle 20 m left of the pad and puts the whole identity of the application into one 40 s frame, which is what Figure 10.5 shows. The position error of 20 m asks for $K_{Px} \cdot 20\,\mathrm{m} = 14.4°$ of tilt, which the limiter clips to 12°, so the outer loop begins saturated and stays there for 0.98 s. The gimbal is moving by the first physics step at $t = 0.005\,\mathrm{s}$, the tilt reaches 10° at $t = 0.91\,\mathrm{s}$, and the position enters the 1 m band around the pad at $t = 14.535\,\mathrm{s}$ and stays inside it. Three timescales, in the order the cascade was designed to produce them.

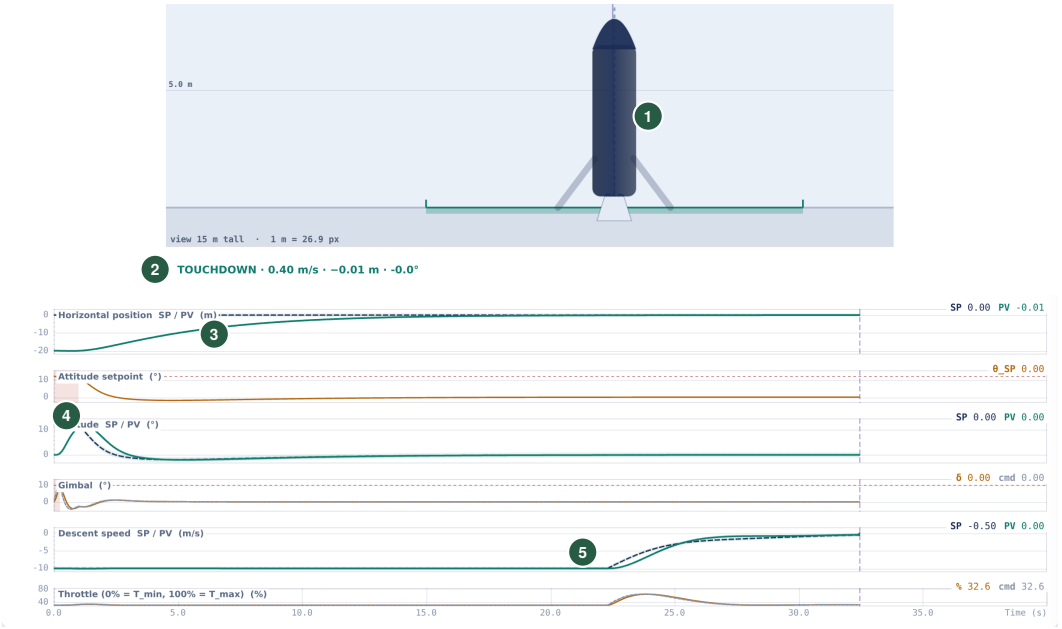


Figure 10.5 Scenario 2 at touchdown, $t = 32.45$ s, a landing from 20 m left of the pad on a 40 s trend window that holds the whole descent. The picture is a composite of the application's window: the lower part of the plant view with its touchdown line, centered above the six trend strips, which are stacked in cascade order. The header, the value chips, the Control and Learn panels and the scorecard are cut. Callouts: (1) the vehicle standing on the pad, (2) the touchdown verdict, (3) horizontal position closing on the pad over fourteen seconds, (4) its manipulated variable, the attitude setpoint, clipped at 12° for the first 0.98 s, shown as the shaded band, because K_{Px} times a 20 m error asks for 14.4°, (5) the descent strip holding −10 m/s until the profile corner at 27.1 m and then shrinking to the terminal rate. The attitude follows its command with the gimbal beneath it. In the logged run the gimbal peaks at 8.4° at $t = 0.26$ s and the attitude at 11.4° at $t = 1.25$ s. The scorecard, cut from this picture, reports a position integral of 123.0 m s, a peak tilt of 11.4° and an arrival 0.01 m from the pad center at 0.40 m/s (row 2 of Table 10.5), a peak throttle of 64 %, and saturation tiles of 2.9 % for the position loop and 0.6 % for the attitude loop, the share of the run in which each command is clipped. The position enters the 1 m band around the pad at 14.535 s.

TRY IT IN THE APP **2 · 20 m Lateral Displacement**

Load the scenario and press Run. Watch the horizontal position trend in the first second. The vehicle first drifts a further 0.118 m to the left, away from the pad, and only then turns toward it. It enters the 1 m band around the pad at 14.535 s and lands 0.01 m from the center.

The first thing the vehicle does, however, is travel the wrong way. Horizontal position drifts a further 0.118 m to the left, the horizontal acceleration is positive from the physics step ending at $t = 0.435$ s and the velocity reverses at $t = 0.71$ s, after which the vehicle closes on the pad and lands 0.01 m from center at 0.40 m/s. The linearized step response of Section 10.4 predicts an acceleration reversal at 0.505 s for an instantaneous full deflection, and the measured figure is earlier because the gimbal cannot step and the attitude loop has already begun to pull the deflection back before it saturates the travel limit. The excursion is 0.118 m rather than the 0.33 m of a sustained deflection for the same reason.

Table 10.7 Scenario 2 at Four Position Bandwidths, Each Critically Damped With K_{Dx} of 4.10, 5.90, 8.35 and 11.85 ° per m/s on the Slider Grid. The Attitude Loop Is Unchanged at 3.07 rad/s Throughout.

K_{Px} (°/m)	ω_{nx} (rad/s)	Ratio	x IAE (m s)	θ IAE (rad s)	θ_{SP} travel (rad)	Verdict
0.72	0.351	8.7	123.0	0.264	0.276	landed 0.01 m
1.50	0.507	6.1	98.5	0.337	0.407	landed 0.00 m
3.00	0.717	4.3	86.0	0.488	0.741	landed 0.00 m
6.00†	1.014	3.0	82.4	4.607	10.05	hard, 1.28 m/s lateral

†Counterfactual. The position-gain slider stops at 3 °/m, so this row was set in a script and cannot be reproduced from the interface.

10.7.3 Bandwidth Separation Sweep

The bandwidth separation is not a rule of thumb in this application, it is a number that can be swept. Table 10.7 repeats scenario 2 at four position gains, each paired with the rate gain that keeps the position loop critically damped through Equation (10.13), so the only thing changing is the separation from the attitude loop.

The table shows the outer loop improving at the inner loop's expense as the position gain rises. Within the slider's range, a four-fold increase in position gain improves the position integral by 30 %, from 123.0 to 86.0 m s, while the attitude integral grows by a factor of 1.85 and the attitude setpoint travels 2.7 times as far. Doubling the gain again, beyond the slider in the counterfactual last row, improves the position integral by 33 % in all, to 82.4 m s, while the attitude integral degrades by a factor of 17.5 and the attitude setpoint travels 36 times as far. That row is not a poorly damped landing but a failed one. At a separation of 3.0 the outer loop revises θ_{SP} before the inner loop has delivered the previous command, the tilt limit is active for 16.3 s of a 32.4 s descent against 0.98 s at the default gains, and the vehicle arrives with 1.28 m/s of lateral speed against a 1.00 m/s limit. The same separation can be reached inside the slider ranges by slowing the inner loop instead, with $K_{Px} = 3.00\,°/\text{m}$, $K_{Dx} = 8.35\,°$ per m/s, $K_{P\theta} = 0.60$ and $K_{D\theta} = 0.46$ s (2.17 rad/s at a damping ratio of 0.83). That vehicle still lands, 0.00 m from the pad center, with an attitude integral of 1.134 rad s, so the ratio alone does not decide the outcome and the failure of the last row also needs an outer loop as fast as 1.01 rad/s. A metric read on the controlled variable alone would have called the last row the best of the four. The application recognizes this and prints both natural frequencies and their ratio whenever one of the three cascade gains is moved, with a warning below a ratio of four.

10.7.4 A Slow Gimbal Actuator

Scenario 8 slows the gimbal servo from 0.08 s to 0.18 s and halves its slew limit to 30 °/s, leaving every gain alone. A lag of 0.18 s contributes $\arctan(0.18 \times 3.07) = 28.9°$ of phase at the attitude natural frequency, which is most of the margin a proportional-derivative loop of damping ratio 0.83 has to give. Restoring the nominal actuator with everything else unchanged gives the reference run 8a. Against it, the slow gimbal raises the peak tilt from 10.2° to 14.9° on a 12° command, the attitude integral from 0.233 to 0.404 rad s and the gimbal travel from 0.491 to 1.199 rad, a factor of 2.4, all of it spent on ringing through −5.5°, +2.7° and −2.3° after the peak, which the nominal actuator does not produce.

The repair is less gain rather than more. Halving $K_{P\theta}$ to 0.60 moves the attitude loop from 3.07 rad/s to 2.17 rad/s and, because $K_{D\theta}$ is left alone, from a damping ratio of 0.83 to one of 1.18. Both changes put the loop back inside what the actuator supports, and the peak tilt falls to 7.9° with the gimbal travel down to 0.214 rad, less than half the nominal-actuator figure. The position integral is unchanged at 89.9 against 89.4 m s, so the outer loop registers neither repair. The attitude integral carries no information here, because it reads 0.404 for the ringing case and 0.406 for the well-damped one. An error integral cannot separate an overshoot that comes back from a lag that never left, while the peak and the actuator travel can. Slowing the servo one step further to $\tau_\delta = 0.20$ s at the original

gains makes the attitude loop diverge, and the vehicle tumbles through 266° and meets the pad at 64 m/s. An inner loop can never be faster than its actuator, and the failure appears first as ringing and second as a tumble.

> **TRY IT IN THE APP** **8 · Slow Gimbal Actuator**
>
> Run the scenario and watch the attitude trend. On a 12° command the tilt peaks at 14.9° and rings through −5.5°, +2.7° and −2.3°. Reset, halve Attitude gain to 0.6 and run again. The peak falls to 7.9° and the ringing is gone, while the position integral barely moves.

10.7.5 Gimbal Saturation

Scenario 9 reduces the gimbal travel to 3° and starts the vehicle 25 m off the pad and still moving away from it. The attitude loop asks for 14.4° of gimbal and gets 3°, so the pitching moment is frozen at $\ell T \sin 3^\circ = 1.85\,\mathrm{kN\,m}$ no matter what the proportional-derivative law computes. While that lasts the attitude loop has no authority and the outer loop is writing commands that cannot be executed. The gimbal command is clipped for 0.91 s and the tilt command for 2.62 s, the maximum position error grows to 27.7 m against 20.1 m in scenario 2, the position integral rises from 123.0 to 207.6 m s, and the entry into the 1 m band moves from 14.535 s to 17.400 s.

The landing still succeeds, 0.02 m from the pad center, and two things save it. Neither cascaded loop has an integrator, so nothing winds up while the gimbal is pinned and the command is correct again the instant the saturation clears. The tilt limit keeps the demand bounded, so the saturation is temporary by construction. Neither protection is unlimited. At $\delta_{\max} = 1.5^\circ$ the vehicle still lands, 0.03 m from center with a peak tilt of 13.7°, and at $\delta_{\max} = 1.0^\circ$ it tumbles and arrives 42.5 m away at 57.8 m/s.

10.7.6 Steady Crosswind Offset

Scenario 5 holds a steady 7 m/s crosswind and the vehicle lands 5.77 m downwind of the pad, inside the 8 m pad radius and well away from the center. Every link of the chain derived in Section 10.4 can be read off the export, and the values below are taken at $t = 20\,\mathrm{s}$, twelve seconds before contact and with the vehicle creeping downwind at 0.04 m/s. The gimbal holds −1.00° against a predicted −1.01°, the attitude holds −3.53° against a predicted −3.54°, the attitude setpoint holds −4.37° against a predicted −4.38°, and the position sits 5.82 m downwind where the linear chain predicts 6.09 m. Matching three angles of a cascade against closed-form values is stronger evidence than matching one. Of the remaining 0.27 m of position, 0.24 m is the rate term, which is not zero because the vehicle has not stopped, and the rest is the 0.02° by which the attitude setpoint falls short of its prediction.

Doubling K_{Px} to 1.44 °/m shrinks every term in the chain in proportion and the offset falls from 5.77 m to 2.96 m, a ratio of 1.95 against the predicted 2.00. It also pushes the position loop from 0.351 rad/s to 0.496 rad/s and closer to the attitude loop, which Table 10.7 shows is not free. The textbook answer of adding integral action to the position loop is the one the application declines to take, because the same scenario file that produces this offset can produce a saturated gimbal, and an outer integrator winding up against a pinned inner loop is the classic cascade failure. Volume 20, Chapter 3 takes the other route on a multirotor that also moves sideways only by tilting, and estimates the wind with a disturbance observer.

10.7.7 Reduced Maximum Thrust

Scenario 7 lowers the maximum thrust to 14 kN against a 11.77 kN weight, a thrust-to-weight ratio of 1.19 instead of 1.70, and starts the vehicle perfectly trimmed over the pad. Hover still exists. What has been lost is deceleration authority. 2.23 kN of spare thrust is 1.86 m/s^2 of retardation, while the exponential branch of the guidance profile demands $k_z|v_z| = 0.35 \times 10 = 3.5\,\mathrm{m/s^2}$ the moment the vehicle enters it. The throttle pins at 100 %, the loop stops being a loop for 3.36 s of commanded saturation, and the vehicle arrives at 3.59 m/s against a 2.0 m/s limit. Raising $T_{\max}$ in the 250 N

steps of its slider, the smallest engine that lands on the shipped profile is 14.5 kN, at 0.46 m/s, while 14.25 kN still arrives hard at 2.06 m/s.

No tuning repairs this, because the arithmetic that separates a tuning problem from a capability problem is a comparison of required against available acceleration and it does not mention the controller. Reducing k_z to 0.15/s asks for 1.5 m/s^2 instead, which the engine has, and the same vehicle lands at 0.47 m/s with a peak throttle of 94 % and no saturation at all. The cost is 13.2 s of extra flight time, from 26.11 s to 39.32 s.

10.7.8 The Gravity Feedforward

Scenario 4 starts the vehicle 10° off vertical over the pad, so the attitude loop and the vertical loop are both disturbed at once. The tilt costs $1-\cos 10^\circ = 1.52\,\%$ of the vertical thrust, and the $1/\cos\theta$ term in the bias restores it before any descent-speed error develops. The descent integral is 6.83 m with the feedforward on against 6.67 m for a vertical descent that was never disturbed at all, so the vertical loop barely notices that the vehicle spent the first second lying over.

Switched off before the run, the feedforward leaves the controller with no preload, and the integral has to build the 3.77 kN between the engine's minimum thrust and hover by itself. On the vertical descent the descent integral then rises from 6.67 m to 12.94 m and the throttle travel from 67.0 % to 125.0 %. With the bias in place the descent-speed error is identically zero until the profile corner, and without it the error reaches 1.93 m/s at $t = 1.65$ s and accumulates 6.29 m of integrated error in the first ten seconds against zero. The thrust command falls to 8.29 kN while the integral builds, the vehicle falls faster than the profile asks, and the descent finishes 0.625 s early at 31.825 s. In scenario 4 the descent integral rises from 6.83 m to 12.94 m and the largest early descent-speed error from 0.090 m/s to 1.97 m/s (row 4b of Table 10.5), so the tilt correction is lost inside the hover thrust the integral is still building. Thrown at $t = 5$ s instead, the switch leaves the thrust command unchanged and the vertical descent keeps the descent integral of 6.67 m it has with the feedforward on. The feedforward is inexpensive here because the coupling is known exactly, being $\cos\theta$ with θ measured.

10.7.9 Propellant Depletion

Scenario 10 switches propellant depletion on and starts from 320 m. The vehicle burns 150.9 kg, from 1200 kg to 1049 kg, which is 12.6 % of its mass, and the three loops react differently because they depend on different quantities. Hover thrust falls with mass, so the peak throttle drops from 64 % to 52 %, and with it the attitude plant gain $\ell T/I$ falls from 7.85/s^2 to 6.86/s^2, moving the attitude loop from 3.07 rad/s to 2.87 rad/s and the right-half-plane zero from 2.80 rad/s to 2.62 rad/s. The horizontal authority $T/m \approx g$ barely changes, because thrust and weight fall together, so the position loop hardly notices and the vehicle lands 0.00 m from center. The vertical loop notices least of all, because the feedforward is computed from the current mass rather than the initial one.

The direction of that drift is a consequence of holding the inertia fixed. A real vehicle emptying tanks near its base loses inertia faster than it loses mass, so its attitude loop and its right-half-plane zero both move to higher frequency as the descent proceeds, and the jet damping that accompanies the mass flow is a further term this model does not carry [15]. The lesson survives the sign either way, which is that the plant a controller lands with is not the plant it was tuned for.

10.7.10 The Graded Challenge

The graded scenario combines a 20 m offset, 4 m/s of drift, 3° of initial tilt, a light crosswind, a 12 m/s gust at $t = 12$ s and an engine limited to 16 kN, and ships with tuning that is wrong in three separate ways. Equations (10.9) and (10.13) put the attitude loop at 1.40 rad/s with a damping ratio of 0.34 and the position loop at 0.57 rad/s with a damping ratio of 0.18, a separation of only 2.5, and $\tau_{I_v} = 20$ s leaves the vertical loop with almost no integral action. The challenge is passed by a landing within 1 m of the pad center at less than 1.5 m/s of vertical speed and less than 5° of tilt, judged at contact on the case as given. As shipped the vehicle reaches 23.8° of tilt, spends 10.96 s against the tilt limit and arrives hard at 3.64 m/s, 2.26 m off the pad and 5.22° over, missing all three limits.

Applying the tuning derived in Section 10.4, which is also the application's suggested tuning, lands it 0.33 m from the pad center at 0.40 m/s and 0.02° of tilt, inside the three limits by 0.67 m, 1.10 m/s and 4.98°, with the peak throttle at 97 % and no margin to spare. The position integral falls from 148.4 to 99.6 m s and the attitude integral from 2.081 to 0.246 rad s. Every repair made a loop slower and better damped rather than stronger, which is the usual direction of travel when a cascade misbehaves and the opposite of the first instinct most operators bring to it.

10.7.11 Manual Operation

Two manual experiments are worth running before any loop is closed. Trimming the engine to hover with the helper, from the vertical descent scenario, holds the thrust exactly equal to the weight and therefore holds the descent rate exactly constant. The vehicle arrives at 10.00 m/s after 25.0 s, which is a hard landing produced by a perfectly correct trim. A trim is not a landing, and the vertical loop exists to change the speed rather than to hold it.

The second experiment is to leave 0.5° of gimbal in from the same trimmed state. Equation (10.8) says the tilt then grows as $a\sin(0.5°)t^2/2$, which reaches the 6° tip-over angle at $t = 1.75$ s, and the simulation crosses it at $t = 1.83$ s, the difference being the servo lag. The vehicle continues through 345° of rotation and meets the ground at a vertical speed of 61.7 m/s, 71.4 m downrange, at $t = 11.535$ s. Half a degree of a ten-degree actuator, left alone for under two seconds, is enough to lose the vehicle, which is what a plant with no restoring moment means in practice.

10.8 Defects in the Application Measurements

The saturation tile of each loop reports the fraction of the run during which the manipulated variable sat against a limit. When this chapter was first written, the limits handed to the metric accumulator for the two cascaded loops were constants fixed at 10° for the gimbal and 15° for the attitude setpoint, so the position loop reported 0 % saturated even in the graded scenario, where the tilt command is clipped for 10.96 s of a 22.82 s descent. The application now takes both limits from $\theta_{\max}$ and $\delta_{\max}$ as they stand, draws its limit lines there, and the position-loop tile reads 2.9 % in scenario 2 and 48.0 % in the graded scenario.

The attitude and descent tiles then still under-reported, because they scored the achieved gimbal angle and thrust, which approach a stop only through the rate limit and the servo or spool lag. Scenario 9 read 0.0 % while the gimbal command was clipped for 0.91 s, and scenario 7 read 6.3 % against 12.8 % of commanded saturation. Both tiles now count the boolean saturation flag each loop publishes on the command against the limit in force, the column the trend strips shade from, and read 2.8 % and 12.8 %. Saturated times quoted in Section 10.7 are counted from the same flags over the physics steps of the run.

The gravity feedforward switch once shifted the integral by the whole of $mg/\cos\theta$, overshooting by the 8 kN floor of Equation (10.19), so that switching it off in flight commanded a climb. The application now shifts the integral by the difference between the two biases and the transfer is continuous in both directions. A switch made before the run once preloaded the integral while the same switch carried through Reset did not, and the briefings described only the first. The application now gives the run without preload on both paths, and the briefings teach it.

10.9 Technology in Practice

Figure 10.6 puts the three loops of Figure 10.2 on flight hardware: navigation sensors feed one flight computer, and its outputs drive a gimbal actuator and a throttle actuator on the same engine.

What Measures It

The controller of this chapter reads x, z, v_x, v_z, θ and ω exactly, and Section 10.3.6 names that as the largest omission of the model, since a real vehicle infers those states from an inertial measurement unit and a radar altimeter. The lunar module's landing radar handed velocity and range to the guidance

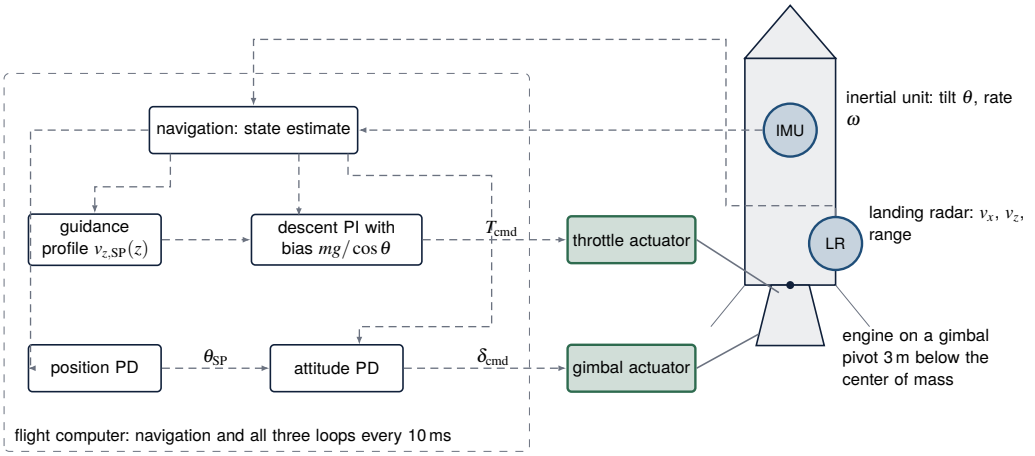


Figure 10.6 The three loops of Figure 10.2 on flight hardware. An inertial unit and a landing radar feed a navigation estimate in the flight computer, the position loop writes the attitude setpoint for the attitude loop, and the attitude and descent loops drive a gimbal actuator and a throttle actuator on the one engine. The application replaces the sensors and the navigation block by the exact state. Solid lines are the vehicle and its devices, and dashed gray arrows are signals. Notice that both actuators act on the same engine below the center of mass, which is why the thrust that holds the vehicle up also sets the gain of the attitude loop.

computer in binary form [19]. Each of those readings carries a lag, a noise and a failure mode that the exact state vector of Section 10.5 leaves out, which is why every result in Table 10.5 is the tuning-limited case.

> **TECHNOLOGY** **Sensor: Doppler Landing Radar**
>
> The lunar module sensed its velocity and slant range relative to the surface with a landing radar that joined a three-beam Doppler velocity sensor to a frequency-modulated radar altimeter. The Doppler shift of each beam is directly proportional to the velocity along that beam, and the altimeter reads range from the frequency difference between the transmitted and the received sweep. Frequency trackers follow each return and signal lock, and the beam velocities are resolved into components for the guidance computer [19]. Ground tests found false lock-on to Doppler returns from the vehicle's own vibrating legs, engine skirt and lower structure. Less sensitivity to low-frequency vibration, an antenna turned away from the legs and a baffle against the engine bell cured it [19]. A falsely locked tracker reports the motion of the structure, not of the ground.

What Moves It

The gimbal servo of Equation (10.5) stands for the gimbal actuator, with its 80 ms lag, its 60 °/s slew limit and its 10° travel, and the spool lag stands for the throttle actuator and the engine. The lunar descent engine was throttled over a range of ten to one, and its gimbal was a slow trim drive backed by attitude jets [12, 20]. The simulated vehicle has no jets, so its gimbal is the only source of pitching moment, and its engine spans only $T_{\min} = 8\,\mathrm{kN}$ to $T_{\max} = 20\,\mathrm{kN}$, a floor that the descent loop meets in scenario 7.

TECHNOLOGY Final Element: Throttled Engine on a Trim Gimbal

The lunar descent engine held both of the chapter's final elements. Its basic requirements were a throttling ratio of 10 to 1 and gimbaling of 6° either way in two axes. Its lowest throttle point was 10 % of rated thrust, and thrust between 65 and 92.5 % was treated as a region not to operate in. A triple-redundant electric actuator positioned the throttle [12]. The gimbal was a trim drive that moved at a constant angular rate for as long as the computer held its command on. During the descent burn the autopilot steered attitude with it to save jet propellant, and when the rather slow drive could not keep the attitude error within bounds, the reaction control jets took over [20].

Where the Controller Runs

All three loops run in one flight computer. On the lunar module the guidance computer issued engine on and off commands, the descent-engine thrust and the gimbal commands under program control, and its digital autopilot chose which attitude jets to fire [20]. The guidance law above that autopilot and the autopilot below it are the outer and inner layers of this chapter's cascade. The reports read for this section give no execution period for the autopilot. The application executes every loop each 10 ms, which is a 100 Hz flight computer, one eighth of the gimbal servo's time constant.

IN PRACTICE Margins to Start From, Gains That Must Move

A design criteria monograph of 1970 collects what launch vehicle autopilots learned in flight. Gain margins of 6 dB and phase margins of 40° proved suitable values with which to begin a design. Because the vehicle's dynamics change rapidly in flight, gains chosen for liftoff may be inadequate later, and a preprogrammed change of gain is recommended. Propellant slosh can be a major contributor to instability, and baffles are the most common remedy. A rate gyro mounted where the structure bends senses that bending and can feed it back as control action, so where the sensor sits is itself a design choice [21].

What Is Modernizing It

Three developments bear on this cascade. The first replaces microwave radar with light. A navigation Doppler lidar uses a frequency-modulated continuous wave to measure velocity and range together, in three fixed directions, from altitudes above 5 km, with velocity resolved to centimeters per second. Vibration of its fiber network shows up as frequency drift and costs accuracy, so the instrument's mounting is part of its design [22]. The second is terrain-relative navigation. A flash lidar records a full three-dimensional image with a single laser pulse, and a landing that must detect hazards of 30 cm needs a ground sample distance of about 15 cm [23]. Both address the navigation error that the exact state vector of this chapter hides. The third is adaptive control. An adaptive augmenting layer added to a launch vehicle autopilot extends the failures and flight anomalies the control system can accommodate while it keeps a direct link to classical gain and phase margins [24], the margins the in-practice segment above starts from.

10.10 Limitations

The simulator teaches a cascade and does not design a lander. The controller is given a perfect state vector, so the navigation problem that dominates real touchdown precision is absent entirely, and the numbers in Table 10.5 are tuning-limited rather than sensor-limited. Moving horizon estimation or Kalman filtering against the same model would supply that state vector, and would estimate the crosswind as a disturbance state so the standing offset of Section 10.7 could be removed by biasing x_{SP} instead of by an integrator in a loop whose inner loop saturates. Planar motion excludes roll, yaw and the coupling between them that a real thrust-vector control system spends most of its authority on, and the fixed inertia excludes slosh, which is the mode most likely to sit near the attitude loop on a part-full vehicle. The controller is fixed in structure, so the convex trajectory generation that defines the modern powered-descent literature [3–6] cannot be demonstrated inside the application, although the right-half-plane zero of Equation (10.10) is exactly the property that makes a guidance

law worth having, because a trajectory that respects the vehicle's own bandwidth asks less of the loops that track it. Every disturbance is operator-triggered or constant rather than stochastic, and a fixed integration step with a fixed control period gives bit-identical repeats that no field controller delivers. A run terminates at contact, so no scenario except the vertical descent has a settled steady state to report, and every integrated measure in this chapter covers a whole descent rather than a common window. Where every modeling assumption stops holding is given in Section 10.3.6, and the three costliest are the exact state feedback, the constant inertia under propellant burn and the missing ground effect in the last few meters.

The predictive option fixes nominal geometry and lag, dead-reckons mass and corrects acceleration bias. Its soft tilt band and finite local forecast do not ensure contact feasibility, as the slow-servo and doubled-inertia failures show.

> ⚠ **OPEN CHALLENGE** **Landing in Its Own Exhaust**
>
> Every propulsive lander meets the interaction of its hot exhaust with the ground beneath it. The plume obscures cameras and sensors, threatens the vehicle's stability and tilt, raises local heating and pressure and throws debris. Apollo flight experience cannot be extrapolated to the larger landers now planned, and generalized models remain a challenge because the physics joins rarefied gas flow to granular soil, with validation work remaining for all methods and tools. Progress would be models checked against flight data gathered by instruments carried on the landers themselves [25]. The model of this chapter has no ground effect at all, as Section 10.3.6 states.

10.11 Discussion

An expendable booster is graded on where it puts its payload. A reusable one is graded twice, and the second grade is awarded in the last thirty seconds, on a pad a few tens of meters across, with an engine that cannot be shut off and relit and a propellant margin measured in seconds. The economics follow directly. Every kilogram of propellant reserved for the landing is a kilogram not lifted, so the descent is flown as close to the minimum as the guidance and the loops beneath it allow, which is why the powered-descent problem was worth convexifying [3, 4] and why solving it on board in bounded time is a practical rather than academic achievement [6].

What this application adds to that picture is the part the optimization does not touch. The trajectory can be optimal and the vehicle can still miss, because the loops that track it are bounded by $\sqrt{\ell T/I}$, a number fixed when the engine bay was laid out. Scenario 6 shows the same gust costing 1.36 m of pad offset on a fast approach and 0.29 m on a slow one, a factor of 4.7 for 6.86 s of extra hang time and the propellant that goes with it. The gust arrives at the same instant on both runs but not at the same altitude, 20.1 m on the shipped profile and 30.2 m at $k_z = 0.15/\mathrm{s}$, which is part of what the slower profile buys. A gust modeled as a step of side force decaying with one time constant is not a wind field, but it arrives at the same instant with the same amplitude on every run, so the comparison is a controlled experiment. That trade between how fast a vehicle comes down and how accurately it arrives is the terminal guidance design problem, and in this application it is one adjustable profile slope. The same structure, an outer loop that carries the objective and no actuator, an inner loop that carries the authority and no knowledge of the objective, and a bandwidth separation that keeps the two from contesting the same actuator, appears in quadrotors, missiles, gimballed camera mounts and ship autopilots [7], and on the bench in the ball and beam of Volume 7, Chapter 5, whose position loop asks an inner drive for an angle.

10.12 Conclusion

A vehicle that can only move sideways by pointing itself gives the outer loop of a cascade an angle to command instead of a force, and that single change reorders what the design has to respect. The simulation of this chapter integrates nine states under planar rigid-body dynamics with a gimbal servo, an engine spool, a quadratic side force applied above the center of mass and a graded touchdown

event, and closes two proportional-derivative loops and one velocity-form proportional-integral loop around them at 100 Hz.

Linearizing about hover puts a right-half-plane zero at $\sqrt{\ell T/I} = 2.80\,\mathrm{rad/s}$ in the path from gimbal to position, and closing the attitude loop leaves it there, so the separation the cascade needs and the margin the zero allows are the same number. The shipped tuning places the attitude loop at 3.07 rad/s and the position loop at 0.351 rad/s, a separation of 8.7. A sweep that pushes the position gain past its slider range to a separation of 3.0 improves the position integral by 33 % while degrading the attitude integral by a factor of 17.5 and turning a landing into a hard arrival.

Twelve scenarios, each with a stated completion requirement, break the cascade one mechanism at a time, with a slow gimbal costing 14.9° of peak tilt against 7.9° after the inner gain is halved, a 3° gimbal stretching the entry into the 1 m band from 14.5 s to 17.4 s, and a 7 m/s crosswind leaving a 5.77 m standing offset whose gimbal, attitude and attitude-setpoint links each match an algebraic chain to within 0.02°. The saturation tiles and the feedforward transfer found defective while this chapter was written have been corrected. The added nominal MPC improves selected arrivals but spends more gimbal motion and fails with wrong servo or inertia assumptions. The lander gives a course an accessible cascade whose outer loop must borrow its authority.

KEY POINTS

- The lander moves sideways only by tilting, so the position loop asks for a tilt and waits for the attitude loop to produce it.
- The engine below the center of mass puts a right-half-plane zero at $\sqrt{\ell T/I} = 2.80\,\mathrm{rad/s}$ that no attitude gain moves.
- The shipped loops are separated by a factor of 8.7, and at 3.0 the landing turns into a hard arrival.
- A steady 7 m/s crosswind leaves a 5.77 m offset, because neither cascaded loop integrates.
- A slow gimbal is repaired by less attitude gain, an engine short of deceleration only by a slower descent.

Where else this applies. Volume 20, Chapter 3 moves a multirotor sideways by tilting it, and Chapter 9 meets a right-half-plane zero in an electrical cascade.

References

[1] J. Meditch, "On the problem of optimal thrust programming for a lunar soft landing," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 477–484, 1964.

[2] A. R. Klumpp, "Apollo lunar descent guidance," *Automatica*, vol. 10, no. 2, pp. 133–146, 1974.

[3] B. Acikmese and S. R. Ploen, "Convex programming approach to powered descent guidance for Mars landing," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 5, pp. 1353–1366, 2007.

[4] B. Acikmese, J. M. Carson, and L. Blackmore, "Lossless convexification of nonconvex control bound and pointing constraints of the soft landing optimal control problem," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2104–2113, 2013.

[5] P. Lu, "Propellant-optimal powered descent guidance," *Journal of Guidance, Control, and Dynamics*, vol. 41, no. 4, pp. 813–826, 2018.

[6] D. Malyuta, T. P. Reynolds, M. Szmuk, T. Lew, R. Bonalli, M. Pavone, and B. Acikmese, "Convex optimization for trajectory generation: A tutorial on generating dynamically feasible trajectories reliably and efficiently," *IEEE Control Systems*, vol. 42, no. 5, pp. 40–113, 2022.

[7] R. W. Beard and T. W. McLain, *Small Unmanned Aircraft: Theory and Practice*. Princeton, NJ: Princeton University Press, 2012.

[8] 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.

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

[10] B. Wie, *Space Vehicle Dynamics and Control, Second Edition*. Reston, VA: American Institute of Aeronautics and Astronautics, 2008.

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

[12] W. R. Hammock, Jr., E. C. Currie, and A. E. Fisher, "Apollo experience report: Descent propulsion system," NASA Manned Spacecraft Center, Houston, TX, NASA Technical Note TN D-7143, 1973, nASA Technical Reports Server 19730011150.

[13] D. H. Klyde, D. T. McRuer, and T. T. Myers, "Pilot-induced oscillation analysis and prediction with actuator rate limiting," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 1, pp. 81–89, 1997.

[14] F. O. Eke and T. C. Mao, "On the dynamics of variable mass systems," *International Journal of Mechanical Engineering Education*, vol. 30, no. 2, pp. 123–137, 2002.

[15] S.-M. Wang and F. O. Eke, "Rotational dynamics of axisymmetric variable mass systems," *Journal of Applied Mechanics*, vol. 62, no. 4, pp. 970–974, 1995.

[16] K. J. Åström and T. Hägglund, *Advanced PID Control*. Research Triangle Park, NC: ISA – The Instrumentation, Systems, and Automation Society, 2006.

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

[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] P. Rozas and A. R. Cunningham, "Apollo experience report: Lunar module landing radar and rendezvous radar," NASA Manned Spacecraft Center, Houston, TX, NASA Technical Note TN D-6849, 1972, nASA Technical Reports Server 19720016521.

[20] D. H. Shelton, "Apollo experience report: Guidance and control systems: Lunar module stabilization and control system," NASA Lyndon B. Johnson Space Center, Houston, TX, NASA Technical Note TN D-8086, 1975, nASA Technical Reports Server 19760004104.

[21] National Aeronautics and Space Administration, "Effects of structural flexibility on launch vehicle control systems," NASA, Washington, DC, Space Vehicle Design Criteria (Guidance and Control) SP-8036, 1970, nASA Technical Reports Server 19700030458.

[22] A. Gragossian, B. Barnes, J. Heppler, J. Follman, N. Dostart, R. Amundsen, A. Ben Shabat, G. Hines, and F. Amzajerdian, "Navigation doppler lidar for lunar landers," in *Optica Advanced Photonics Congress 2022*. Optica Publishing Group, 2022, p. LsM3C.4.

[23] F. Amzajerdian, A. Gragossian, P. F. Brewster, J. M. Heppler, F. G. Wilson, G. D. Hines, S. A. Laughter, D. K. Litton, and A. Bulyshev, "Development of a compact lidar sensor for terrain relative navigation and terrain hazard avoidance," in *AIAA SciTech 2025 Forum*, 2025.

[24] J. H. Wall, J. S. Orr, and T. S. VanZwieten, "Space launch system implementation of adaptive augmenting control," in *AAS Guidance, Navigation, and Control Conference*, no. AAS 14-051, 2014, nASA Technical Reports Server 20140008747.

[25] A. M. Korzun and W. A. Chambers, "Development efforts and status of plume-surface interaction capabilities for propulsive landing systems," Presentation, AIAA SciTech Forum, NASA Technical Reports Server 20230018215, 2024.

10.13 Exercises

10.1 ○ Concept check · 2 min

In the cascade of Figure 10.2, what does the horizontal position loop manipulate?

(a) The attitude setpoint θ_{SP}, which the attitude loop must deliver first

(b) A side thruster that pushes the vehicle toward the pad

(c) The gimbal angle δ, which turns the thrust toward the pad at once

(d) The thrust magnitude T, whose horizontal share grows with the tilt

10.2 ○ Concept check · 2 min

An engineer raises the attitude gain $K_{P\theta}$ from 1.20 to 4.80 at the hover operating point. What happens to the right-half-plane zero of the path from θ_{SP} to x?

(a) It doubles in frequency, together with the attitude loop
(b) It stays at $\sqrt{\ell T/I} = 2.80\,\mathrm{rad/s}$, since feedback moves only poles
(c) It cancels against the faster attitude poles and disappears
(d) It crosses into the left half plane once the attitude loop is faster

10.3 Concept check · 1 min

In scenario 5 the vehicle lands 5.77 m downwind of the pad center in a steady 7 m/s crosswind. What causes the offset?

(a) The 12° tilt limit stops the vehicle leaning far enough into the wind
(b) The descent integrator winds up and takes thrust from the horizontal channel
(c) Neither cascaded loop integrates, so the standing lean needs a standing position error
(d) Without a wind sensor the vehicle cannot hold any lean against a steady wind

10.4 Theory and calculation · 6 min

Use the hover operating point of Table 10.2: $m = 1200\,\mathrm{kg}$, $g = 9.81\,\mathrm{m/s^2}$, $\ell = 3.0\,\mathrm{m}$ and $I = 4500\,\mathrm{kg\,m^2}$.

(a) Compute the attitude plant gain a, the zero $\sqrt{a}$, and the attitude natural frequency and damping ratio at the shipped $K_{P\theta} = 1.20$ and $K_{D\theta} = 0.65\,\mathrm{s}$. *Checkpoint:* $a = 7.848$ per $\mathrm{s^2}$, 2.801 rad/s, 3.069 rad/s and 0.831.

(b) The graded challenge ships with $K_{P\theta} = 0.25$, $K_{D\theta} = 0.12\,\mathrm{s}$, $K_{Px} = 1.90^\circ/\mathrm{m}$ and $K_{Dx} = 1.20^\circ$ per m/s. Place both loops and compute their separation. *Checkpoint:* Attitude 1.40 rad/s at 0.34, position 0.570 rad/s at 0.18, separation 2.46.

(c) Find the attitude gain that would put the attitude loop at twice the zero frequency. *Checkpoint:* $K_{P\theta} = 4$.

10.5 Theory and calculation · 6 min

For a step deflection δ of the gimbal at hover, the linearized model gives $x(t) = g\delta\left(at^4/24 - t^2/2\right)$ with $a = 7.848\,\mathrm{s^{-2}}$.

(a) Find the times at which the horizontal acceleration changes sign, the wrong-way excursion is largest, and the vehicle passes back through its starting point. *Checkpoint:* 0.505 s, 0.874 s and 1.237 s.

(b) Compute the largest wrong-way excursion for $\delta = 5^\circ$. *Checkpoint:* 0.164 m.

(c) Scenario 10 ends its descent with $a = 6.86\,\mathrm{s^{-2}}$. Recompute the time of the acceleration reversal, and say whether a change of the attitude gains would move any of the three times. *Checkpoint:* 0.540 s.

10.6 Theory and calculation · 7 min

Use Equation (10.14) and the descent-loop settings of Table 10.3, with $m = 1200\,\mathrm{kg}$ and hover thrust $mg = 11\,772\,\mathrm{N}$.

(a) Compute the descent loop's natural frequency, its damping ratio and the velocity time constant of the proportional term alone. *Checkpoint:* 0.707 rad/s, 0.884 and 0.80 s.

(b) Compute the gravity feedforward $mg/\cos\theta$ at 12° and at 30° of tilt, and the vertical force a 14 kN engine has to spare at 30°. *Checkpoint:* 12 035 N and 13 593 N, and 352 N against 2228 N upright.

(c) Scenario 7 limits the engine to 14 kN. Compare the deceleration the profile of Equation (10.7) demands at its corner, $k_z v_{\max}$, with what the engine can add above hover, and find the largest

profile slope on the 0.01 s^{-1} slider grid whose corner demand the engine can meet. *Checkpoint:* 3.5 m/s^2 against 1.86 m/s^2, so $k_z = 0.18\,s^{-1}$.

10.7 App challenge · 8 min

Load scenario 1, Vertical descent (start here), in Auto at the default seed. Read the clock when the run stops at touchdown and the Touchdown speed and Peak throttle tiles.

(a) Run to touchdown as loaded. Read the touchdown time and speed, and compute the altitude of the profile corner from Equation (10.7). *Checkpoint:* 32.45 s at 0.40 m/s, corner at 27.1 m.

(b) Reset, open Advanced, set Descent profile slope to 0.15 and run to touchdown. Reset again, set it to 0.80 and run once more. *Checkpoint:* 39.32 s at 0.47 m/s, then 25.38 s at 3.29 m/s with the throttle peaking at 96 %.

(c) Reload the scenario, switch to Manual before pressing Run, press Trim to hover and run to touchdown without touching anything else. *Checkpoint:* 10.00 m/s at 25.0 s.

Accept a time within 0.05 s and a speed within 0.01 m/s.

10.8 App challenge · 8 min

Load scenario 8, Slow gimbal actuator, in Auto at the default seed. Each run goes to touchdown. Read the Peak tilt tile and the verdict.

(a) Run as loaded and read the peak tilt. Compute the phase that the 0.18 s servo lag contributes at the attitude natural frequency of 3.07 rad/s. *Checkpoint:* 14.9°, and 28.9° of phase.

(b) Reset, set Attitude gain to 0.60 and run again. Read the peak tilt, and compute the new attitude natural frequency and damping ratio with $K_{D\theta}$ left at 0.65 s. *Checkpoint:* 7.9°, 2.17 rad/s and 1.18.

(c) Reload the scenario, set Gimbal time constant to 0.20 s in the Parameters tab and run to touchdown. *Checkpoint:* A tip-over, peak tilt about 266°, at 64.2 m/s.

Accept a tilt within 0.2° in (a) and (b).

10.9 App challenge · 8 min

Load scenario 5, Steady 7 m/s crosswind, in Auto at the default seed, and run each case to touchdown.

(a) Run as loaded and read the Pad offset tile. *Checkpoint:* 5.77 m.

(b) Export the run and read the row at `time_s` = 20: `gimbal_deg`, `theta_deg`, `theta_sp_deg`, `x_m` and `vx_mps`. Compare them with the predictions of Section 10.4, and account for the gap in position with the rate term of the position law. *Checkpoint:* −1.00, −3.53, −4.37, 5.82 m and 0.043 m/s. The rate term is worth 0.24 m.

(c) Reset, set Position gain to 1.44 °/m and run again. Compare the offset ratio with the factor of two that the chain predicts. *Checkpoint:* 2.96 m, a ratio of 1.95.

Accept an offset within 0.02 m and an angle within 0.01°.

10.10 Application area · 4 min

Section 10.3.6 calls exact full-state feedback the largest omission of the model.

(a) Name the instruments a real lander infers its states from, and say what then sets its touchdown precision.

(b) A wind estimate is used to bias x_{SP} upwind by the offset the chain predicts. The estimate reads 6 m/s while 7 m/s blows. Using the chain of Section 10.4, estimate the offset that remains. *Checkpoint:* About 1.6 m.

10.11 Application area · 4 min

The model holds the pitch inertia fixed while propellant burns.

(a) In scenario 10 the vehicle burns 150.9 kg and the model's a falls from 7.85 to 6.86 s^{-2}. Suppose the real vehicle's inertia fell by 20 % over the same descent. Compute a, the attitude natural frequency at $K_{P\theta} = 1.20$ and the zero at touchdown. *Checkpoint:* 8.58 s^{-2}, 3.21 rad/s and 2.93 rad/s.

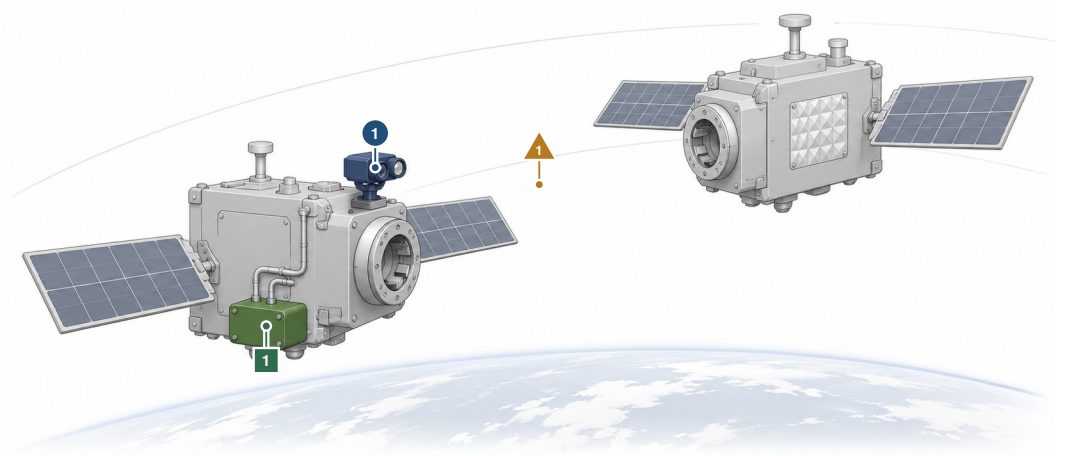
(b) Name the assumption that hides propellant slosh, and say why slosh matters to this loop in particular.

10.12 Review the agent · 5 min

An AI assistant is asked how to land scenario 2 faster and replies as follows. "The ceiling on the position loop is the speed of the attitude loop beneath it. Quadruple the attitude gain to 4.80 and set the attitude rate gain to 1.30 s. The attitude loop doubles to 6.1 rad/s at the same damping of 0.83, so the ceiling doubles with it. Then run the position loop at its 3 °/m maximum with $K_{Dx} = 8.35°$ per m/s, a separation of 8.6, as comfortable as the shipped 8.7." Find the flaw and say how to check it.

CHAPTER 11

Canceling Coriolis Coupling by Feedforward in an Orbital Rendezvous Simulator



The controller moves 1 thrusters on three axes

The controller reads 1 relative position and velocity sensor

What disturbs it 1 differential drag along the track

The approaching craft reads its position relative to the target and fires its thrusters, but because both circle the Earth, a push aimed straight at the target also lifts it or lowers it. Illustration, not to scale.

IN PLAIN TERMS Steering Up Close to a Craft in Orbit

The main idea. The view from the target turns once each orbit, so near another craft a push in one direction also moves the arriving craft in another. The controller already knows these side effects and adds small extra pushes that cancel them before any error appears.

The challenge. A faster controller hides the side effects but never removes them. The extra pushes also burn fuel, so there is a best speed that uses the least.

Why it matters. Every supply flight to a space station ends with this slow approach, flown on fuel that cannot be refilled.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history of proximity operations and the fact on manual braking (Section 11.2), the first try-it segment (Section 11.3.3), the second try-it segment and the fact on station keeping (Section 11.6) 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 11.6), Technology in Practice (Section 11.7), Limitations (Section 11.8) and the open challenge

The control question A chaser near a target in orbit is flown in axes that turn once per orbit, and the gravity-gradient and Coriolis terms make three thrusters on three axes a coupled plant. What does canceling that known coupling by feedforward buy that more feedback gain cannot, and what does it cost in propellant?

Learning objectives

- Derive the free-drift paths of the Clohessy-Wiltshire equations and explain why a hold behind the target is free.
- Show that the model feedforward leaves three damped second-order loops, and compute the stiffness an uncompensated loop loses.
- Compute the propellant-optimal bandwidth and the bandwidth the thrusters can support.
- Measure what a wrong mean motion, a thrust limit and a degraded thruster cost at two bandwidths.
- Compare a propellant-pricing MPC with the compensated loop on propellant and docking time.

Prerequisites Sections 1.4.2 (where a disturbance enters), 1.4.3 (static feedforward and the ideal compensator) and 1.6.4 (scenarios and exported files). Section 1.6.3 for the MPC only. Chapter 10 for a gravity feedforward on another vehicle.

The app https://apmonitor.com/apps/orbital-rendezvous/

See also Volume 6, Chapter 2 for a static feedforward that removes loop interaction. Volume 20, Chapter 7 for the attitude loop this chapter leaves out, Volume 20, Chapter 8 for orbit maintenance under drag and oblateness, Volume 20, Chapter 6 for thruster allocation and a failed thruster.

11.1 Introduction

Three thrusters on three orthogonal axes present as three independent actuators, and the natural pairing puts each one with the position it appears to move [1]. That pairing is correct almost everywhere a spacecraft flies. It stops being correct inside a few kilometers of another vehicle, because the frame in which the relative position is measured is attached to the target and the target is following a curved path. The axes rotate once per orbit, at 1.131 mrad/s in a 400 km circular orbit, and every consequence of writing Newton's second law in rotating axes appears as a term in the relative equations of motion.

Two of those terms make the pairing false. A Coriolis term of magnitude $2n$ times a velocity appears with the axes exchanged, so an along-track velocity produces a radial acceleration and a radial velocity produces an along-track acceleration. A gravity-gradient term of magnitude $3n^2$ pushes a radial displacement further out rather than pulling it back. Neither is a force acting on the chaser. Both are exact once the target's mean motion is known, which is the property the simulator of this chapter is built around. A proportional-derivative loop on each axis, tuned as though the axes were independent, is fighting a positive feedback path on the radial axis and is being pushed sideways by a path it cannot see. Adding four algebraic terms that the flight computer can already evaluate removes both effects and turns the plant into three uncoupled second-order systems.

Relative motion of this kind has a settled analytical description. The linearized equations for a satellite near a circular reference orbit were published as a terminal guidance system for satellite rendezvous [2], and they are the same equations that appear in the lunar theory of the previous century [3]. Rendezvous guidance built on them ranges from piloted braking gates and glideslope profiles [4] to trajectory families chosen so that a failed burn leaves a passively safe relative orbit [5], and to model predictive formulations that carry the thrust bound and the approach corridor as explicit constraints [6, 7]. However, the step that makes any of those methods work is elementary and is usually left implicit. The coupling terms are known, so they are subtracted rather than rejected, and what remains for feedback is only the unmodeled part of the plant.

The plant carries a specific set of process dynamics and control principles: the difference between a pairing that is wrong and a plant that is coupled, an unstable open-loop mode and a marginally stable one on the same vehicle, feedforward from a model rather than from a measured disturbance, the sensitivity of a feedforward term to the parameter it is evaluated with, actuator saturation as a

limit on achievable bandwidth, and proportional offset under a constant load [1, 8, 9]. The physical fundamentals are Newton's second law written in a rotating frame, a first-order expansion of the gravitational difference between two nearby bodies, and the separation between a bounded periodic mode and a secular one. The behavior matters industrially because station resupply, crew rotation, on-orbit servicing and debris removal all reduce to the same terminal approach, flown against a propellant budget that cannot be replenished in flight.

This chapter measures what one control decision is worth on a plant whose coupling is known exactly. It works with an interactive rendezvous simulator, built on the framework of Section 1.6, in which the Clohessy-Wiltshire terms are canceled or left in place by a single switch. The central result is a propellant argument. Because the feedforward term $3n^2x$ must itself be produced by the thrusters, the cumulative velocity change of a compensated regulator has an interior minimum in closed-loop bandwidth at $\omega_n = n\sqrt{3\mathrm{e}} = 3.23\,\mathrm{mrad/s}$, which a fine sweep of the simulation places at 3.14 mrad/s. The six-state simulator's thruster bound, docking test, disturbance set and bumpless-transfer arrangement are set down completely enough that another implementation reproduces the same trajectories. Its set-ups present the plant before the controller and break the coupling, the actuator bound and the model error one at a time. The Clohessy-Wiltshire model and the algebraic cancellation of known terms are both established methods, and the chapter uses them as they stand.

Section 11.2 places the plant in rendezvous practice and guidance. Section 11.3 gives the relative motion model, its free-drift solutions, its constants and its assumptions, with the equations and initial states a reimplementation needs. Section 11.4 derives the model feedforward, its sensitivity to the assumed altitude, the bandwidth the thrusters allow and the predictive controller. Section 11.5 lists the set-ups, and Section 11.6 measures the compensation at fixed gains, the bandwidth sweep, saturation, model error, station keeping, a degraded thruster, the graded challenge and the predictive controller. Section 11.7 sets the loops on flight hardware, Sections 11.8 and 11.9 state the limits of the model and what transfers to other plants, and the exercises of Section 11.11 close the chapter.

11.2 Background

Rendezvous was the enabling technique of the lunar program and it was flown by hand first. The Gemini program established that two crewed vehicles could be brought together and docked, and the piloted procedures worked out there, with braking gates, a defined approach corridor and a station-keeping phase before contact, carried forward into Apollo and then into the Space Shuttle [10]. What the crews were taught is the lesson the simulator of this chapter reproduces. Pointing the nose at the target and thrusting does not close the range, and the correct action at long range is frequently a burn along the velocity vector rather than toward the object being approached.

> ⧖ HISTORY **Proximity Operations Gets a Name, 1973 to 1977**
>
> Gemini and Apollo flew an inertial final approach and braked at the end. By mid-1973 the reusable orbiter raised a new concern: its primary reaction control jets, of 870 pounds thrust, could contaminate a small target with their exhaust, induce attitude rates on it or even push it away. Piloted simulations in 1976 showed that the inertial approach did not remove the concern, while approaches along the target's velocity vector or along its radius vector were acceptable. By March 1977 the term proximity operations had been coined for flight within 2000 feet of the target, characterized by nearly continuous trajectory control, whereas rendezvous maneuvers come at intervals of hours or tens of minutes [11].

Automated rendezvous developed on a separate track and for a different reason. A station in permanent occupation has to be resupplied whether or not a pilot is aboard the arriving vehicle, and Progress freighters, the European Automated Transfer Vehicle and the commercial cargo and crew vehicles that followed all perform the terminal approach without a hand on a controller [12]. The transition from piloted to automated operation is not a matter of replacing the pilot with a controller, because the pilot was also supplying the navigation, the fault detection and the decision to abort.

A review of what automation has to take over, and of the flight demonstrations that attempted it, is given in [13].

> ✦ DID YOU KNOW? **Braking Left to the Crew**
>
> A 1966 memo on Apollo terminal guidance deleted automatic braking from the guidance computer. Based on Gemini experience, crews could brake as well given range and range rate, and the program would have used 50 to 100 words of scarce storage. The first automated rendezvous and docking was flown in October 1967 by two uncrewed spacecraft [11].

The guidance literature splits along the same line. Impulsive targeting solves a two-point boundary value problem for the two burns that move a vehicle from one relative state to another, which is a purely open-loop construction and requires that the relative motion between the burns is predicted correctly. Glideslope approaches replace it with a closed-loop rule in which the closing speed is commanded as a linear function of the remaining range, executed as a sequence of small burns, and the relative navigation needed to support it is treated together with the guidance in [4]. Both families raise a question no amount of tuning answers, which is what the vehicle does if the next burn does not happen. Trajectories chosen so that the uncorrected relative motion never intersects the target, known as safety ellipses and passive-abort trajectories, are designed by adding that requirement as a constraint on the free-drift solution [5]. Model predictive control is the natural home for the whole set, because the thrust bound, the approach cone, the plume-impingement keep-out region and the terminal speed limit are all constraints, and formulations that carry them explicitly for proximity operations are given in [6, 7].

The simulator of this chapter sits underneath that literature rather than beside it. An optimizer that plans a constrained approach still has to be compared against the simplest thing that could work, and the simplest thing that could work on this plant is a proportional-derivative loop with the known terms fed forward. The simulator also carries such a planner as an option, and Section 11.6.9 weighs it against that loop on propellant and on time to dock.

11.3 Relative Motion Model

Two vehicles fly near one another in a circular orbit of radius $R = R_E + h_{orb}$ about a central body of gravitational parameter μ. The target occupies the origin of a right-handed frame whose axes are carried around with it, and the chaser is described by its position and velocity in that frame. Figure 11.1 defines the frame, the sense of each axis and every symbol the equations use.

11.3.1 Classification of Variables

Three loops are declared, one per axis, and each one pairs a relative position with the thruster acceleration on the same axis. Table 11.1 names every signal, the part it plays in the loop and the range over which it can be set.

Nothing in the vehicle is measured imperfectly. The controller reads the true position and velocity on all three axes with no noise, no bias, no lag and no navigation filter, which is the largest single idealization in the model and is revisited in Section 11.3.5. The only imperfection available is in the actuator, because a de-rated thruster still consumes the command it was given and does not produce the acceleration.

11.3.2 The Clohessy-Wiltshire Equations as Implemented

Both vehicles are point masses in the gravity field of a spherical central body. Expanding the difference of their gravitational accelerations to first order in the separation, and writing the result in axes that

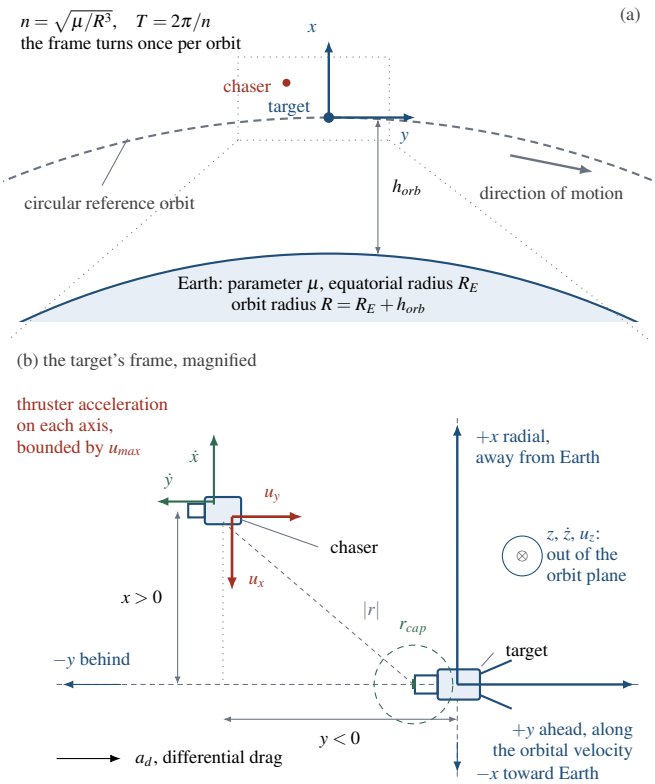


Figure 11.1 Relative motion geometry and the local-vertical local-horizontal frame. The frame origin rides on the target, x points radially away from the central body, y points along the orbital velocity, and $z = \hat{x} \times \hat{y}$ completes the right-handed triad along the orbit angular momentum, which is into the page in this view. A chaser behind and above the target therefore has $y < 0$ and $x > 0$. The whole frame turns at the mean motion n, and the chaser is driven only by its own bounded thruster acceleration (u_x, u_y, u_z) and by the optional along-track differential drag a_d. Panel (b) is magnified by roughly four orders of magnitude relative to panel (a).

rotate at the constant rate n about the orbit normal, gives the system in Equation (11.1).

$$\begin{aligned} \ddot{x} &= 3n^2 x + 2n\dot{y} + u_x \\ \ddot{y} &= \quad -2n\dot{x} + u_y + a_d \\ \ddot{z} &= -n^2 z \quad\quad + u_z \end{aligned} \tag{11.1}$$

where x, y and z are the relative position components in m, u_x, u_y and u_z are the commanded thruster accelerations in m/s^2, a_d is the along-track differential drag in m/s^2 and n is the mean motion of the reference orbit in rad/s. Every line carries units of m/s^2. The terms $3n^2x$ and $-n^2z$ are the net of the gravity gradient and the centrifugal term of the rotating frame, and the two terms $\pm 2n$ times a velocity are Coriolis. None of the four is a force on the chaser, and all four vanish if the same motion is written in an inertial frame. All four are consequences of choosing axes in which the target is stationary.

The mean motion follows from the circular-orbit condition and is given by Equation (11.2).

$$n = \sqrt{\frac{\mu}{R^3}}, \qquad R = R_E + h_{orb}, \qquad T = \frac{2\pi}{n} \tag{11.2}$$

Table 11.1 Classification of Every Signal. Entries Marked (adv.) Are Available Only in the Advanced Set.

Symbol	Role	Default	Range
x	CV, radial position	0 m	—
y	CV, along-track position	0 m	—
z	CV, cross-track position	0 m	—
$\dot{x},\dot{y},\dot{z}$	states, relative velocity	0 m/s	—
x_{SP}	setpoint, radial	0 m	± 300 m
y_{SP}	setpoint, along-track	0 m	± 500 m
z_{SP}	setpoint, cross-track (adv.)	0 m	± 200 m
u_x	MV, radial acceleration	0	$\pm u_{\max}$
u_y	MV, along-track acceleration	0	$\pm u_{\max}$
u_z	MV, cross-track acceleration	0	$\pm u_{\max}$
a_d	DV, differential drag	0	$\pm 20\,\mu\mathrm{m/s^2}$
—	DV, degraded thruster axis	none	x, y or z
$1-\eta$	DV, thrust lost on that axis	0 %	0 to 100 %
ω_n	tuning, in-plane bandwidth	10.0	1 to 40 mrad/s
ζ	tuning, damping ratio	1.00	0.20 to 2.00
$\omega_{n,z}$	tuning, cross-track (adv.)	8.0	1 to 40 mrad/s
$\hat{h}_{orb}$	tuning, assumed altitude (adv.)	400 km	200 to 1000 km
—	option, control architecture	PD + comp.	three, see text
T_{arr}	tuning, MPC arrival time (adv.)	26 min	5 to 40 min
$\lvert r\rvert$	output, range to the port	0 m	—
$\lvert v\rvert$	output, relative speed	0 m/s	—
$\dot{r}$	output, closing speed	0 m/s	—
Δv	output, cumulative burn	0 m/s	—
n	output, true mean motion	1.131	mrad/s
T	output, orbital period	5553.6	s

where μ is the gravitational parameter in $\mathrm{m^3/s^2}$, R_E is the equatorial radius of the central body in m and h_{orb} is the orbital altitude in m. At the default 400 km this evaluates to $n = 1.131\,367\,\mathrm{mrad/s}$ and $T = 5553.6\,\mathrm{s}$, which is 92.56 min. Every coupling term in Equation (11.1) scales as n or n^2, so the coupling weakens with altitude. Between 200 km and 1000 km the mean motion falls by 15.8 % and the gravity-gradient coefficient $3n^2$ by 29.1 %.

Equation (11.1) is linear and time invariant, and the only nonlinearities anywhere in the simulation are the thruster clamp, the de-rating of a faulted thruster, the washout of the bumpless-transfer bias and the three latched proximity events. The state vector is exactly the six relative coordinates, and the propellant total is accumulated after the integration step rather than carried as a seventh state.

11.3.3 Free-Drift Behavior of the Three Axes

Setting $u_x = u_y = u_z = a_d = 0$ in Equation (11.1) leaves the free-drift solution, and the three axes behave completely differently. For a chaser released from rest at a pure radial offset x_0 the in-plane solution is Equation (11.3).

$$x(t) = (4 - 3\cos nt)\,x_0, \qquad y(t) = 6\,(\sin nt - nt)\,x_0 \tag{11.3}$$

The radial component is bounded and periodic and reaches $7x_0$ at half a period. The along-track component is not bounded, because the secular term $-6ntx_0$ never stops growing, and its slope $6nx_0$ is 0.679 m/s for $x_0 = 100\,\mathrm{m}$ at the default altitude, or 40.7 m of drift per minute from an initial condition in which nothing was moving. A radial offset is an unstable mode of this plant.

An along-track offset is the opposite case. Substituting $x = z = 0$ and $\dot{x} = \dot{y} = \dot{z} = 0$ into Equation (11.1) makes every right-hand side vanish, because every term contains x, z or a velocity. A

vehicle parked anywhere along the velocity vector with zero relative velocity is at an exact equilibrium of the plant and stays there at no propellant cost. This asymmetry is why hold points on a real approach are placed along the velocity vector rather than above or below the target.

The cross-track axis is an undamped harmonic oscillator at exactly the orbital frequency, and its solution for a pure out-of-plane rate $\dot{z}_0$ is Equation (11.4).

$$z(t) = \frac{\dot{z}_0}{n} \sin nt \tag{11.4}$$

The amplitude is $\dot{z}_0/n$, which is 88.4 m for 0.10 m/s, and the error returns to zero once per orbit whatever the controller does. A plane error is bounded and self-correcting, so the cross-track loop is given its own and slower bandwidth.

The case that matters most for the control problem is a pure closing velocity. For a chaser starting at y_0 with an along-track rate $\dot{y}_0$ and nothing else, the radial solution is Equation (11.5).

$$x(t) = \frac{2\dot{y}_0}{n} (1 - \cos nt) \tag{11.5}$$

The Coriolis term has converted a velocity that pointed exactly at the target into a radial excursion whose amplitude is $2\dot{y}_0/n$, and for 0.5 m/s that amplitude is 883.9 m. Nothing in a coast removes it. Figure 11.2 draws all three cases in the plane of the orbit, computed from Equations (11.3) and (11.5) with the same mean motion the simulator uses.

Pause. Try Exercise 11.1 before reading on.

> **TRY IT IN THE APP** **3 · Aimed Straight at the Target**
>
> Load the scenario, which starts in Manual with every thruster at zero, and press Run. Watch the range to the port. It falls to a closest approach of 115.22 m at 363.5 s, with the chaser 73.7 m above the port, and then grows as the chaser climbs away, although its velocity pointed straight at the port.

11.3.4 Numerical Constants, Limits and Latched Events

Table 11.2 collects every constant in the model. A fixed 0.5 s step of fourth-order Runge-Kutta advances the six states, the control law is evaluated on the same 0.5 s period, and one sample is logged per simulated second. Both rates are set against the fastest thing in the problem, which is the closed loop at a few mrad/s rather than the orbit itself. A control period of 0.5 s is 8×10^{-4} of one period of a closed loop at 10 mrad/s. A step does its work in the order of Section 1.6.1, with the propellant accounting and the proximity tests placed after the integration. The right-hand side of Equation (11.1) writes nothing back into the state, so the four Runge-Kutta stages cannot disagree.

Three events are latched after the integration step and are the only discontinuities in the run. Inside the capture radius r_{cap} a relative speed at or below v_{cap} declares a dock and the run pauses once so the propellant total can be read, except in the graded challenge, which runs on to its verdict at 1500 s. A relative speed above v_{cap} declares a collision. Beyond $r_{\max}$ a loss-of-proximity flag is raised, with 5 % hysteresis, marking the range at which the linearization of Equation (11.1) stops being a fair description. Propellant is accumulated as the time integral of the commanded acceleration magnitude given by Equation (11.6).

$$\Delta v = \int_0^t \sqrt{u_x^2 + u_y^2 + u_z^2}\, \mathrm{d}\tau \tag{11.6}$$

A second total is accumulated the same way from the accelerations actually delivered, so a de-rated thruster makes the two diverge. Neither is an impulsive budget, because the propulsion model is a continuous bounded acceleration.

Two continuous disturbances and two one-shot disturbances are available in every set-up. A differential-drag bias of $\pm 20\,\mu\mathrm{m/s^2}$ enters the along-track equation directly. A selected axis can

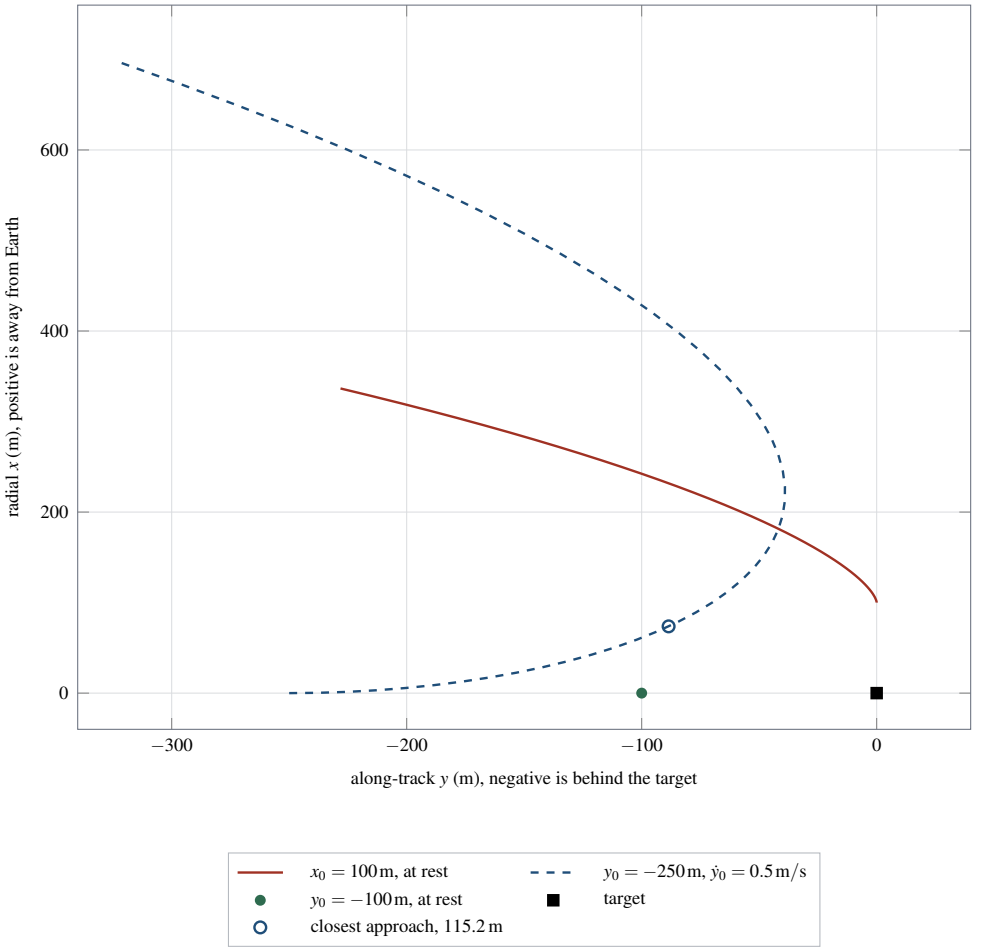


Figure 11.2 Twenty minutes of free drift from three initial conditions, evaluated from the closed-form solutions of Section 11.3.3 at $n = 1.131\,367\,\mathrm{mrad/s}$. The radial offset (solid) curls away and falls behind, the along-track offset (filled circle) does not move at all, and the coast whose relative velocity points exactly at the target (dashed) rises above it and passes over the top with a miss distance of 115.2 m at $t = 364\,\mathrm{s}$ (open circle). The target is the filled square. Every thruster is off in all three cases.

have between 0 % and 100 % of its thrust removed, and the commanded value stays the manipulated variable so the failure appears as a gap between the command and the applied trace. A venting impulse adds 0.03 m/s to $\dot{y}$ and 0.01 m/s to $\dot{x}$, and a plume-impingement event adds 0.02 m/s to $\dot{x}$ away from the target. In manual operation a 0.05 m/s translation pulse is available in each of the six directions as an instantaneous change in the corresponding velocity state, charged to the propellant budget.

11.3.5 Assumptions

First-order expansion about a circular orbit. Equation (11.1) drops every term of second order and higher in the separation divided by the orbit radius, and it assumes the reference orbit is exactly circular. Eccentricity, the J_2 oblateness term and third-body effects are all absent. The model declares its own validity boundary at $r_{\max} = 2000\,\mathrm{m}$ against an orbit radius of 6778 km, a ratio of 3×10^{-4}. The oblateness term is the first thing to matter over long horizons, because it makes the relative

Table 11.2 Plant, Controller and Numerical Constants. An Adjustable Parameter Shows the Range Over Which It Can Be Set. A Hidden Row Is a Constant the Model Uses and Does Not Report.

Symbol	Meaning	Value	Unit	Range
Orbit				
μ	gravitational parameter	3.986004418	10^{14} m^3/s^2	fixed
R_E	equatorial radius (hidden)	6378137	m	fixed
h_{orb}	true orbital altitude	400	km	200-1000
n	mean motion, derived	1.131367	mrad/s	—
T	orbital period, derived	5553.6	s	—
$3n^2$	gravity-gradient term	3.840×10^{-6}	1/s^2	—
$2n$	Coriolis coefficient	2.263×10^{-3}	1/s	—
Propulsion and proximity				
$u_{\max}$	thrust acceleration limit	0.050	m/s^2	0.002-0.050
Δv_p	translation pulse (hidden)	0.05	m/s	fixed
—	duty threshold (hidden)	1	% of $u_{\max}$	fixed
r_{cap}	capture radius	3.0	m	1.0-10.0
v_{cap}	maximum contact speed	0.10	m/s	0.01-0.50
$r_{\max}$	loss-of-proximity range	2000	m	500-10000
Controller				
ω_n	in-plane bandwidth	10.0	mrad/s	1-40
ζ	damping ratio	1.00	—	0.20-2.00
$\omega_{n,z}$	cross-track bandwidth	8.0	mrad/s	1-40
$\hat{h}_{orb}$	assumed orbital altitude	400	km	200-1000
K_P	in-plane position gain	1.000×10^{-4}	1/s^2	—
K_D	in-plane rate gain	2.000×10^{-2}	1/s	—
τ_b	bias washout (hidden)	60	s	fixed
T_{arr}	MPC arrival time	26	min	5-40
Metrics and numerics				
—	settled band, each loop	1.0	m	—
—	equilibrium tolerance	0.01	m	—
Δt	integration step (RK4)	0.5	s	—
Δt_c	control period	0.5	s	—
—	historian rate	1	Hz	—

motion secular in a way the circular model does not capture [14], and the drift it produces is a formation flying problem rather than a rendezvous problem only because a rendezvous lasts an hour. Volume 20, Chapter 8 flies a single satellite with the oblateness term and drag in the model, over days rather than an hour.

Point masses with no attitude. Neither vehicle has an orientation, so thrust is available instantly along any axis, the docking port is always pointing the right way and there is no attitude loop competing for the same thrusters. A real chaser holds an attitude to aim both its sensors and its docking mechanism, and the attitude loop is the faster and more delicate of the two. Volume 20, Chapter 7 treats such a loop on a small satellite, with reaction wheels and magnetic torquers.

Continuous bounded acceleration. Reaction control thrusters are on and off devices with a minimum impulse bit, arranged in a geometry that rarely delivers a pure single-axis force, and the vehicle loses mass as they fire. The model has none of that, so a command of 0.001 mm/s^2 is executed exactly.

Exact relative navigation. The controller reads the true state, so the rate feedback that supplies all of the damping costs nothing here and is never free on orbit. Close-range relative navigation is a filtered combination of lidar, camera and radio ranging, and its error rather than its tuning is what sets achievable docking precision [13].

One environmental acceleration. Differential drag is the only non-gravitational effect besides thrust, it acts along-track only, and the adjustable range of $\pm 20\,\mu\mathrm{m/s^2}$ is a teaching stress range, not a representative drag prediction at either altitude. Solar radiation pressure and plume impingement as a continuous effect are absent.

Geometric docking. Contact is declared by a radius and a speed. There is no soft-capture mechanism, no lateral or angular misalignment tolerance and no contact dynamics, so a dock and a collision are separated by a number rather than by a structural calculation.

Two deliberately unphysical set-ups. Mean motion is one of the best-known numbers on any real mission, since it falls out of the tracking solution to many digits, so the 15.8 % ephemeris error of set-up 8 is not a credible failure. It stands in for the quantities that genuinely are uncertain, such as a mass property, a thrust coefficient or an inertia, and the arithmetic of Equation (11.12) is identical for all of them. A thruster that delivers a smooth 30 % of its command is not how reaction control thrusters fail either, but it isolates the one question set-up 10 exists to ask, which is what a feedforward term does when the actuator it assumes to be exact is not.

11.4 Controller Design

One switch selects among three architectures, the proportional-derivative loops alone, the same loops with the model feedforward described below, which is the default, and the model predictive controller of Section 11.4.6. The default controller has two parts. The feedback part is a proportional-derivative law on each axis, written as a gain on the error minus a gain on the measured rate and evaluated on the state vector directly, as shown in Equation (11.7).

$$\begin{aligned} a_x &= K_P\,(x_{\mathrm{SP}} - x) - K_D\,\dot{x} \\ a_y &= K_P\,(y_{\mathrm{SP}} - y) - K_D\,\dot{y} \\ a_z &= K_{P,z}\,(z_{\mathrm{SP}} - z) - K_{D,z}\,\dot{z} \end{aligned} \tag{11.7}$$

where a_x, a_y and a_z are the feedback demands in $\mathrm{m/s^2}$, K_P and $K_{P,z}$ are position gains in $1/\mathrm{s^2}$ and K_D and $K_{D,z}$ are rate gains in 1/s. There is no integral term anywhere, no derivative filter and no internal state, so the feedback law is a memoryless function of the state. It is therefore not the velocity-form PID of Section 1.6.2, and its bumpless transfer is arranged separately, as Section 11.4.4 describes.

11.4.1 Gain Parameterization

Six raw gains obscure the quantity a designer actually chooses, so the loops are parameterized the way a second-order closed loop is specified, by a natural frequency and a damping ratio [1], as shown in Equation (11.8).

$$K_P = \omega_n^2, \quad K_D = 2\zeta\,\omega_n, \quad K_{P,z} = \omega_{n,z}^2, \quad K_{D,z} = 2\zeta\,\omega_{n,z} \tag{11.8}$$

The radial and along-track axes share one ω_n because they are dynamically coupled through the Coriolis terms, and giving them different bandwidths only means that whatever coupling survives arrives at a frequency the other loop is not prepared for. The cross-track axis has no cross terms at all and is given its own and usually slower ω_n, because Equation (11.4) says its error is bounded and returns to zero once per orbit without help. The damping ratio is shared across all three, because it sets the shape of the response rather than its speed. At the reference tuning $\omega_n = 10\,\mathrm{mrad/s}$ the in-plane loop is 8.84 times faster than the orbit itself.

11.4.2 Model Feedforward and Exact Cancellation

The feedforward block is an algebraic copy of the coupling terms of Equation (11.1), evaluated with the mean motion the controller believes it is flying at, and the total command is Equation (11.9).

$$\begin{aligned} u_x &= -3\hat{n}^2 x - 2\hat{n}\dot{y} + a_x \\ u_y &= \quad\quad + 2\hat{n}\dot{x} + a_y \\ u_z &= \hat{n}^2 z \quad\quad + a_z \end{aligned} \tag{11.9}$$

where $\hat{n} = \sqrt{\mu/(R_E + \hat{h}_{orb})^3}$ is computed from the controller's own assumed altitude, which is a separate control from the plant altitude so that the mismatched case can be set up deliberately. Substituting Equation (11.9) into Equation (11.1) with $\hat{n} = n$ leaves Equation (11.10).

$$\ddot{e} + K_D\,\dot{e} + K_P\,e = 0 \tag{11.10}$$

on each of the two in-plane axes, with $K_{P,z}$ and $K_{D,z}$ in place of K_P and K_D on the cross-track axis, where e is the position error on the axis concerned. Three coupled axes, one of them unstable, have become three independent damped second-order systems with their poles exactly where ω_n and ζ place them. The cancellation requires only the state the controller already measures and one number it already knows, with no prediction horizon and no online solve. The coupling terms enter each axis at the same point as the thrust, so the ideal compensator of Equation (1.4) reduces to a pure gain on measured states, with no lead–lag and no delay. The app's own check on this identity drives the compensated loop from the start of set-up 6, $x_0 = 100\,\text{m}$ at rest, at $\omega_n = 12\,\text{mrad/s}$ and $\zeta = 0.7$ for 600 s and compares it against the analytic second-order response. The worst deviation is 0.148 m against a tolerance of 0.165 m, and that tolerance is not arbitrary. It is the half-sample of travel that a zero-order hold at Δt_c introduces against a continuous law, and the deviation halves whenever Δt_c halves, to 0.074 m at 0.25 s.

Pause. Try Exercise 11.2 before reading on.

Switching the compensation off sets all three feedforward terms to zero and leaves Equation (11.7) alone against Equation (11.1). The radial axis then closes as Equation (11.11).

$$\ddot{x} + K_D\dot{x} + \left(K_P - 3n^2\right)x = 2n\dot{y} \tag{11.11}$$

The gravity-gradient term subtracts directly from the position gain. At $\omega_n = 3\,\text{mrad/s}$ the loss is $3n^2/K_P = 42.7\,\%$, at $\omega_n = 10\,\text{mrad/s}$ it is 3.84 %, and the whole of the radial stiffness is gone at $\omega_n = \sqrt{3}\,n = 1.9596\,\text{mrad/s}$. The right-hand side is the term that has no analogue in a single-axis problem at all, because it drives an axis whose own error is zero.

11.4.3 Sensitivity to an Error in the Assumed Altitude

Feedforward cancels the difference between the plant terms and the model terms, so a mismatch in $\hat{n}$ leaves the residuals of Equation (11.12).

$$\ddot{x} + K_D\dot{x} + \left[K_P - 3\left(n^2 - \hat{n}^2\right)\right]x = 2\left(n - \hat{n}\right)\dot{y} \tag{11.12}$$

The useful comparison is against the gains rather than against n itself. A target at 200 km flown with an ephemeris for 1000 km gives $n = 1.1834\,\text{mrad/s}$ and $\hat{n} = 0.9962\,\text{mrad/s}$, an error of 15.8 %. The surviving radial term is $3(n^2 - \hat{n}^2) = 1.224 \times 10^{-6}\,/\text{s}^2$, which is 30.6 % of K_P at $\omega_n = 2\,\text{mrad/s}$ and only 7.6 % of it at 4 mrad/s. The surviving Coriolis term $2(n - \hat{n}) = 3.743 \times 10^{-4}\,/\text{s}$ is 9.4 % of K_D at the lower bandwidth and 4.7 % at the higher one. The same model error is therefore worth four times as much to a slow loop as to a fast one, which is the trade every feedforward design turns on. As in Section 1.4.3, the feedback kept beside the feedforward removes what a wrong model term leaves behind.

11.4.4 Actuator Bound and Bumpless Transfer

Each axis is clamped independently at $\pm u_{\max}$ and there is no rate limit, no minimum impulse bit and no mass depletion. Because Equation (11.7) carries no integrator there is nothing to wind up while the clamp is active, so recovery is immediate once the error shrinks. The demand at the start of a maneuver is K_P times the initial error, so the bandwidth an actuator can support is bounded by Equation (11.13).

$$\omega_n \lesssim \sqrt{\frac{u_{\max}}{e_0}} \tag{11.13}$$

where e_0 is the initial error on the axis in m. A degraded thruster multiplies the clamped command by its efficiency η on the selected axis only, and the commanded value remains the manipulated variable that the trends and the export report.

A proportional-derivative controller has no integrator for the tracking of Section 1.6.2 to write into, so bumpless transfer is arranged explicitly. While the operator holds the actuator, each axis stores the bias $b_i = u_{i,\mathrm{man}} - (ff_i + a_i)$, which is the difference between what the operator is commanding and what the law would command. On transfer to automatic operation the command is $ff_i + a_i + b_i$, which reproduces the operator's value exactly, and the bias then decays as $b_i \leftarrow b_i \exp(-\Delta t_c/\tau_b)$ with $\tau_b = 60\,\mathrm{s}$. Authority moves to the control law over about a minute rather than in one step, and only after the bias has washed out is the loop the pure system of Equation (11.10). In the other direction the three manual commands follow the controller output, so returning to manual operation moves nothing either. What they follow is the law's request before the clamp, which the clamp then limits exactly as it limited the automatic command, so in set-up 7 the stations read −20.23 and 19.15 mm/s^2 at 10 s while the thrusters still receive $\pm 4\,\mathrm{mm/s^2}$.

11.4.5 Three Models Kept Apart

The simulation model is Equation (11.1) with the true n from Equation (11.2) at the true altitude, integrated over six states. The analysis model is Equations (11.3) to (11.5) and (11.10) to (11.12), which this chapter uses to choose gains and to read the measurements and which appears nowhere in the app's code. What the proportional-derivative controller itself holds is the four algebraic terms of Equation (11.9) and the single number $\hat{n}$, and nothing further. That controller does not know μ, R_E, $u_{\max}$ or the docking criteria, and it never predicts forward. The predictive controller of Section 11.4.6 holds more, namely Equation (11.1) at $\hat{n}$ without the drag term, the declared μ and R_E, the limit $u_{\max}$ and the contact speed v_{cap}, and it predicts five minutes ahead. The instruments are the same for both, the six true relative states read without error.

11.4.6 Model Predictive Control on the Exact Model

The third architecture is a model predictive controller (MPC), an option that only set-up 11 selects. Every $T_s = 2\,\mathrm{s}$, once in four executions of the control law, it solves one quadratic program for the thrust over the next $N = 150$ steps and applies only its opening block, which it keeps until the following decision replans from fresh measurements. This is the receding-horizon law reviewed in [15] and applied to proximity operations in [6, 7]. It differs from the framework's MPC option of Section 1.6.3 in its objective, which prices propellant rather than tracking error. The first plan is made when the run is initialized, so the MPC flies from $t = 0$.

Its model is not fitted. Equation (11.1) is linear, so the plan predicts with the equations the plant integrates, without the drag term and at the $\hat{n}$ of Equation (11.9), which comes from $\hat{h}_{orb}$ on the declared R_E and μ and never from the true altitude. Under an exact zero-order hold it is Equation (11.14).

$$\xi_{k+1} = \Phi\xi_k + \Gamma\left(u_k^+ - u_k^-\right),\ \Phi = \mathrm{e}^{AT_s},\ \Gamma = \int_0^{T_s} \mathrm{e}^{A\tau}\mathrm{d}\tau B \tag{11.14}$$

where $\xi = (x, \dot{x}, y, \dot{y}, z, \dot{z})^T$, A and B are the matrices of Equation (11.1) at $\hat{n}$, and $u_k^{\pm} \geq 0$ are the three positive and three negative thrust accelerations held over step k. With $\hat{h}_{orb} = h_{orb}$ the free prediction

equals the closed-form Clohessy-Wiltshire solution to better than 10^{-7} m over the horizon, so there is no residual to report and the lesson is propellant rather than model error, except in set-up 8, where $\hat{n}$ is 15.8 % low for this controller as for the feedforward. Splitting each thrust into two nonnegative parts makes propellant linear, $T_s \sum_i (u_{i,k}^{+} + u_{i,k}^{-})$ per step, which is what the opposed thruster pairs spend. The six thrusts are held over blocks of 1, 2, 4, 8, 16, 32 and 87 steps, 42 decisions.

The 300 s horizon is 5.4 % of the orbit and a fifth to a sixth of an approach, so the rest of a transfer is charged as a terminal cost. Each transfer, begun at a new setpoint or when the MPC takes over outside its box, has an arrival time T_{arr}, and the plan's end state is charged with the propellant of the two-burn Clohessy-Wiltshire transfer from that state to rest on the setpoint at T_{arr}. For the coast τ left after the last judged step, 298 s ahead, both burns are fixed linear maps of the predicted position error $\tilde{r}$ and velocity $\tilde{v}$ there, as shown in Equation (11.15).

$$\begin{gathered} \Delta v_1 = -M\tilde{r} - \tilde{v}, \qquad \Delta v_2 = -\left(\Phi_{vr} - \Phi_{vv} M\right)\tilde{r}, \\ M = \Phi_{rv}^{-1}\Phi_{rr} \end{gathered} \tag{11.15}$$

where Φ_{rr}, Φ_{rv}, Φ_{vr} and Φ_{vv} are the position and velocity blocks of $e^{A\tau}$. The six components are soft rows pinned to zero, each slack $\varepsilon_j \geq |\Delta v_j|$ priced like propellant, and the objective is Equation (11.16).

$$J = \sum_{k=0}^{N-1} \sum_{i} \left[T_s \left(u_{i,k}^{+} + u_{i,k}^{-} \right) + R \left(u_{i,k}^{+2} + u_{i,k}^{-2} \right) \right] + \sum_{j} \left(\sigma_j \varepsilon_j + \rho_j \varepsilon_j^2 \right) \tag{11.16}$$

where $R = 10^{-6}$ per $(\mathrm{mm/s^2})^2$ keeps the program strictly convex, and the slack prices in m/s of propellant are $\sigma = 1$, $\rho = 10^{-4}$ on the terminal rows and $\sigma = 10$, $\rho = 100$ on the envelope and the box. No position error is weighted during a transfer. On 16 plans of set-ups 3, 7, 10 and 11 every terminal slack equals the hand-solved two-burn $|\Delta v|$ to 1.5×10^{-13} m/s, and the applied command is the program's first block exactly, so nothing outside the solve schedules a burn. Once less than 10 s of coast remains beyond the horizon, the terminal rows give way to a box of ± 0.5 m per axis from the arrival step, or a minute ahead if later, with the final velocity priced as the propellant to stop it, and after T_{arr} the plan holds that box with a centering weight of 10^{-3} per $\mathrm{m^2}$ s. Because an end state at rest on the target costs nothing, T_{arr} is a deadline, and the plan arrives early whenever the coast allows.

The thrusts are bounded hard, $0 \leq u_{i,k}^{\pm} \leq u_{\max}$, with $u_{\max}$ read as the proportional-derivative clamp reads it. With no corridor or keep-out zone in the app, the docking rule supplies a soft envelope on the closing speed of each axis, Equation (11.17).

$$\left| \dot{r}_{i,k} + \left(r_{i,k} - r_{\mathrm{SP},i} \right) / 200\,\mathrm{s} \right| \leq v_{cap}/2 + \varepsilon_{e,i} \tag{11.17}$$

where r_i is the position on axis i. Only the side the chaser occupies is applied, both sides within 1 m, so the closing speed allowed is half the contact speed plus 5 mm/s per meter still to go, judged every second step. Each plan has 42 thrusts, at most nine slacks and 84 bound rows, and is solved by the dual active-set method [16] with a ridge of $10^{-9}(1 + 2RN)$.

The arrival time was chosen once, on set-up 11 alone, from 5 to 40 min, as the least propellant among settings that dock no later than set-up 6 (1785.0 s), enter the 1 m box no later (2212.5 s) and reject no plan. The 25 and 26 min settings lie within 1 %, at 0.3552 and 0.3567 m/s, and the longer is shipped. The usual tracking form, a position weight at every step with a propellant price and a terminal weight from the discrete Riccati solution, was flown on the same start and solver at 12 settings (prices 3×10^5 to 10^7, terminal bandwidths 1, 3 and 10 mrad/s). Each spent more than 0.4191 m/s, docked after 1785 s, collided, or stalled 21 to 63 m off, because a linear propellant price outweighs a quadratic incentive that vanishes near the target.

Table 11.3 Set-Up List. Every Entry Not Named Under Change Keeps Its Default From Table 11.2. The First Four Run in Manual Operation With Every Thruster Command at Zero.

#	Change	Mechanism isolated
1	$x(0) = 100\,\mathrm{m}$	an unstable free-drift mode
2	$y(0) = -100\,\mathrm{m}$	an exact equilibrium of the plant
3	$y(0) = -250\,\mathrm{m}$, $\dot{y}(0) = 0.5\,\mathrm{m/s}$	Coriolis turns closing speed into altitude
4	$y(0) = -60\,\mathrm{m}$, $\dot{z}(0) = 0.10\,\mathrm{m/s}$	a bounded, uncoupled third axis
5	slow set, compensation off, $x(0) = 100\,\mathrm{m}$	diagonal feedback against a coupled plant
6	slow set, compensation on, $x(0) = 100\,\mathrm{m}$	exact cancellation of the same coupling
7	$u_{\max} = 4\,\mathrm{mm/s^2}$, $x(0) = 200\,\mathrm{m}$, $y(0) = -200\,\mathrm{m}$	bandwidth the actuator cannot support
8	$h_{orb} = 200\,\mathrm{km}$, $\hat{h}_{orb} = 1000\,\mathrm{km}$, $\omega_n = 2\,\mathrm{mrad/s}$	feedforward on a wrong parameter
9	$y_{\mathrm{SP}} = y(0) = -20\,\mathrm{m}$	station keeping and offset under load
10	along-track thruster at 30 %, $x(0) = 80\,\mathrm{m}$, $y(0) = -80\,\mathrm{m}$	feedforward through a lying actuator
11	slow set, MPC, $T_{arr} = 26\,\mathrm{min}$, $x(0) = 100\,\mathrm{m}$	a plan that pays only for propellant
12	graded, see Section 11.6.7	all of the above, from poor tuning

Offset-free predictive controllers add an integrating disturbance estimated by a Kalman filter [17], as about half of the framework's MPC options do (Section 1.6.3). This one departs from that default and starts each plan from the six measured states the proportional-derivative loops read. Such a filter leaves set-ups 3, 7 and 11 unchanged, but it reads set-up 10's degraded thruster as an outside force proportional to the command and chases it, spending 11.54, 1.92, 1.14 and 1.00 m/s at filter time constants of 106, 1004, 9992 and 150 246 s, against 0.99 m/s without one. Without it, drag of 10 and 20 µm/s^2 leaves a steady hold error of 0.024 and 0.031 m, against the proportional a_d/K_P of 0.100 and 0.200 m. None of the three architectures integrates.

Any plan whose program ends short of optimal, or whose thrust is undefined, infinite, beyond $\pm u_{\max}$ or a negative split, is rejected, and the compensated law of Equation (11.9) takes over without a bump, because under the MPC its bias b_i tracks the applied command at every execution. The record is marked, and the next accepted plan takes control back. A rejection forced at 602 s of set-up 11 moves the command by one execution of the compensated law to within 10^{-12} m/s^2, and no run in this chapter rejects a plan. Under the MPC the saturation counter counts the command at the limit. The program minimizes the pairs' propellant, $\int (|u_x| + |u_y| + |u_z|)\,\mathrm{d}t$, while Equation (11.6) counts the vector's length, so both are reported.

11.5 Scenario Definitions

Table 11.3 lists eleven teaching set-ups and one graded challenge. The app numbers them as its scenarios 1 to 11 and the challenge, each with the briefing and completion rule of Section 1.6.4. The first four run with every thruster off, so the plant is met before the controller is, and set-ups 5 and 6 differ by exactly one boolean. Set-up 11 is the start and tuning of set-up 6 with the model predictive controller selected, so set-ups 5, 6 and 11 fly one transfer under three architectures. Two tunings recur. The reference set is $\omega_n = 10\,\mathrm{mrad/s}$, $\zeta = 1.00$ and $\omega_{n,z} = 8\,\mathrm{mrad/s}$ with compensation on. The slow set used for the central comparison is $\omega_n = \omega_{n,z} = 3\,\mathrm{mrad/s}$ and $\zeta = 1.00$, which puts the loop only 2.65 times faster than the orbit and is where the coupling terms stop being a correction and become the dominant dynamics.

11.6 Results

All of the values below come from the app's own scorecard, tiles and exported samples (Section 1.6.4), taken with the physics step, the control period and the log rate set as in Table 11.2. The exported file,

Table 11.4 Free Drift With Every Thruster Off. The Closed-Form Column Is Evaluated From Section 11.3.3 at the Same Mean Motion.

#	Quantity	Simulated	Closed form
1	x at 1200 s (m)	336.536	336.536
1	y at 1200 s (m)	−228.163	−228.163
1	x at 2778 s (m)	700.000	700.000
1	y at 2778 s (m)	−1886.568	−1886.568
2	y at 3600 s (m)	−100.000	−100
2	Δv over 3600 s (m/s)	0.0000	0
3	closest approach (m)	115.22	115.22
3	time of closest approach (s)	363.5	363.6
3	x at closest approach (m)	73.7	73.75
3	x at 1200 s (m)	696.904	696.904
4	peak $\lvert z\rvert$ (m)	88.389	88.389
4	z at 5556 s (m)	0.238	0.238

at 1 Hz, carries the feedforward and feedback contributions separately on each axis, both propellant totals and the true and assumed mean motions, from which every tabulated result can be recomputed. An integrated measure depends on how long the run lasted, so every one is quoted beside its horizon. A setting that departs from a set-up is chosen before the run starts, and the app gives the same run whether it is chosen then or carried through a Reset. Docking times and closest approaches are read at the 0.5 s physics step, and a run continued past its docking pause is resumed unchanged. The saturation fraction is measured against the $u_{\max}$ in force on each axis, while overshoot and settling time are reported only after a setpoint move, so for these transfers from an initial offset the results quote instead the time from which the range to the port stays below 1 m. Table 11.4 collects the four free-drift set-ups and Table 11.5 the closed-loop ones.

11.6.1 Effect of the Compensation at Fixed Gains

Set-ups 5 and 6 differ by one boolean and are the controlled experiment the simulator exists to run. Both start 100 m above the target with the loops at $\omega_n = 3\,\mathrm{mrad/s}$ and $\zeta = 1.00$. The compensated run docks at 1785 s having spent 0.4191 m/s, and its along-track axis never moves more than 0.0074 m from the setpoint it started on. The uncompensated run with the same gains takes 3516 s to dock, which is 1.97 times as long, spends 0.5402 m/s over the full 6000 s against 0.4195 m/s for the compensated run on the same horizon, which is 28.8 % more, and drags the along-track axis 11.475 m away from a setpoint nothing asked it to leave. That last measurement carries the argument, and it is the trace shown in Figure 11.3.

> **TRY IT IN THE APP** **5 · PD Only, No Orbital Compensation**
>
> Run the scenario and watch the along-track trend, which starts on its setpoint. While the radial loop brings the chaser down, the along-track axis is pulled 11.475 m away. Load 6 · PD + orbital compensation and run again. The along-track error stays within 0.0074 m, and the chaser docks at 1785.0 s.

Two mechanisms produce it and Equation (11.11) names both. The gravity-gradient term removes 42.7 % of the radial position gain at this bandwidth, which is why the radial approach is slow, and the Coriolis term on the right-hand side drives the along-track axis with the radial rate the controller is deliberately generating. Neither is a disturbance in the usual sense, because both are consequences of the controller's own action on a plant whose known terms it is not using.

The uncompensated loop degrades gracefully into instability rather than abruptly. At $\omega_n = 2\,\mathrm{mrad/s}$ the net radial stiffness $K_P - 3n^2$ is $1.6 \times 10^{-7}\,\mathrm{/s^2}$, four percent of K_P, and the radial error

Table 11.5 Closed-Loop Outcomes. Each Error Integral Accumulates the Absolute Error of Its Own Loop Across the Horizon Named in the Third Column, Δv Is the Commanded Total of Equation (11.6), the Docking Time Is the Physics Step on Which the Range First Fell Inside $r_{cap} = 3\,\mathrm{m}$ at or Below $v_{cap} = 0.10\,\mathrm{m/s}$, and "Inside 1 m" Gives the Time From Which the Range Stayed Below 1 m. Row 5e Uses a Bandwidth Between the 0.5 mrad/s Settings of the App and Is a Model Run.

#	Set-up	Horizon	Δv	IAE		Peak	Docked	Note
				radial	along-track	$\lvert y - y_{SP}\rvert$	at	
		(s)	(m/s)	(m s)	(m s)	(m)	(s)	
5	slow set, compensation off	3000	0.5294	113888	21896	11.475	—	not inside 1 m
6	slow set, compensation on	3000	0.4191	66536	8.17	0.0074	1785.0	inside 1 m at 2213 s
5b	set-up 5 run to 6000 s	6000	0.5402	116506	25145	11.475	3516.0	inside 1 m at 4282 s
6b	set-up 6 run to 6000 s	6000	0.4195	66581	8.45	0.0074	1785.0	—
5c	compensation off, $\omega_n = 8$	3000	0.6547	26573	3537	6.080	789.0	—
6c	compensation on, $\omega_n = 8$	3000	0.6607	24924	3.2	0.0074	669.5	—
5d	compensation off, $\omega_n = 2$	6000	2.0917	542598	9310	1.979	—	radial still at 80.19 m
5e	compensation off, $\omega_n = 1.96$	6000	2.3017	599372	108	0.024	—	radial still at 99.78 m
6d	compensation on, $\omega_n = 2$	6000	0.4775	99897	12.6	0.0074	2677.0	—
7	$u_{\max} = 4\,\mathrm{mm/s^2}$, $\omega_n = 10$	2400	2.2313	61475	51357	200.0	766.0	233.5 s clipped
7b	the same at $\omega_n = 4.5$	2400	1.4082	89518	88729	200.0	1462.5	38.5 s clipped
7c	the same at $\omega_n = 4$	2400	1.3504	99799	99810	200.0	1643.0	never clipped
7d	the same at $\omega_n = 3$	2400	1.3022	132707	132723	200.0	2190.0	never clipped
7e	$u_{\max} = 50\,\mathrm{mm/s^2}$, $\omega_n = 10$	2400	2.3015	39848	39850	200.0	657.5	never clipped
8	$\hat{n}$ 15.8 % low, $\omega_n = 2$	3000	0.9130	143325	92568	100.0	—	radial still at 12.48 m
8b	the same with $\hat{n} = n$	3000	0.7444	98914	98931	100.0	2882.0	—
8c	$\hat{n}$ 15.8 % low, $\omega_n = 4$	3000	0.7188	56572	47713	100.0	1602.5	—
8d	the same with $\hat{n} = n$	3000	0.6995	49916	49923	100.0	1441.0	—
9	V-bar hold at –20 m	3600	0.0000	0.0	0.0	0.000	—	exact equilibrium
9b	the same with $a_d = 10\,\mu\mathrm{m/s^2}$	3600	0.0360	0.0	340	0.100	—	offset a_d/K_P
9c	the same with $a_d = 20\,\mu\mathrm{m/s^2}$	3600	0.0720	0.0	680	0.200	—	offset a_d/K_P
9d	the same hold at $x = +20\,\mathrm{m}$	3600	0.2765	0.0	0.0	0.000	—	radial offset, not free
10	along-track thruster at 30 %	1800	2.1882	15939	24554	80.0	986.5	delivered 0.8888 m/s
10b	the same with compensation off	1800	2.3175	19045	25415	80.0	1014.0	delivered 0.9125 m/s
10c	the same with a healthy thruster	1800	0.9206	15939	15940	80.0	550.5	—
12	graded, as shipped	1500	0.4646	91918	42159	60.0	—	radial grew to 87.86 m
12b	graded, retuned	1500	0.5750	7969	11756	60.0	494.5	inside 1 m at 624 s

is still 80.19 m after 6000 s with 2.0917 m/s already spent. A model run at 1.96 mrad/s, between the 0.5 mrad/s settings the app offers and just above $\sqrt{3}\,n = 1.9596\,\mathrm{mrad/s}$, has a net stiffness of

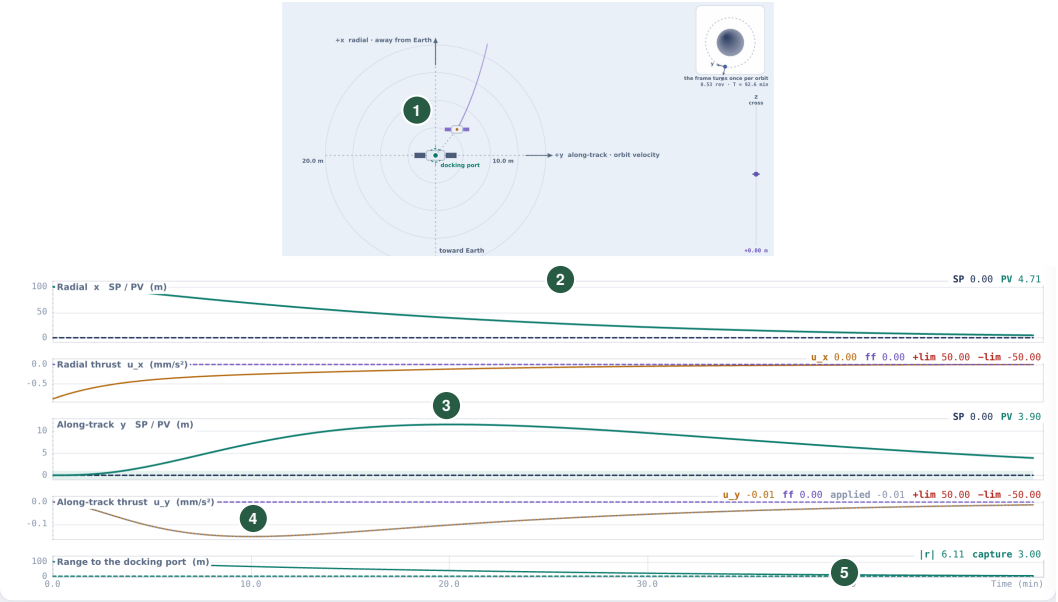


Figure 11.3 A 100 m radial transfer with the orbital compensation switched off. Set-up 5 is shown at $t = 2968$ s, read on the app's clock, which the crop leaves out, flown by three independent proportional-derivative loops at $\omega_n = 3$ mrad/s. The picture is a composite of two parts of one screenshot: the plant view (top, reduced) and the trend panel (bottom). (1) The chaser just above the docking port. (2) The radial measurement, fallen from 100 m to 4.71 m in fifty minutes. (3) The along-track measurement, the diagnostic. It began on its setpoint, nothing disturbed it, and the Coriolis term $-2n\dot{x}$ produced by the radial correction has carried it to 11.5 m and back to 3.90 m, for an along-track error integral of 21 773 m s on the scorecard, which the crop leaves out. (4) The along-track thrust strip. Each thrust strip is scaled to its own command, at most 0.90 mm/s^2 radially and 0.16 mm/s^2 along-track, and the ± 50 mm/s^2 limits lie far off both axes. (5) The range to the port, still 6.11 m.

1.6×10^{-9}/s^2 and leaves the radial error at 99.78 m, essentially where it started, after 2.3017 m/s of propellant spent holding it there. The compensated loop at the same $\omega_n = 2$ mrad/s docks at 2677.0 s for 0.4775 m/s.

11.6.2 Bandwidth Sweep and the Propellant Minimum

A fair comparison requires sweeping the bandwidth, because enough feedback authority masks any plant. Table 11.6 repeats the 100 m radial transfer at thirteen bandwidths with and without the compensation, each critically damped, over a 9000 s horizon. Figure 11.4 draws the same columns on logarithmic axes.

The compensated column has an interior minimum, and it is there for a reason worth deriving. With the cancellation exact, the critically damped error from rest is $x(t) = x_0(1 + \omega_n t)\mathrm{e}^{-\omega_n t}$, the feedback part of the command is $\ddot{x}$, and its absolute integral is $2\omega_n x_0/\mathrm{e} = 0.736\,\omega_n x_0$, which grows with bandwidth. The feedforward term $-3\hat{n}^2 x$ is a thrust command in its own right, and $\int_0^\infty x\,\mathrm{d}t = 2x_0/\omega_n$, so it contributes $6n^2 x_0/\omega_n$, which shrinks with bandwidth. Setting the two equal gives Equation (11.18).

$$\omega_n^\star \approx n\sqrt{3\mathrm{e}} = 2.856\,n = 3.23\,\text{mrad/s} \tag{11.18}$$

A model sweep in steps of 0.01 mrad/s puts the measured minimum at 3.14 mrad/s with 0.4189 m/s, within 3 % of the prediction. The bandwidth setting of the app moves in 0.5 mrad/s steps, and its

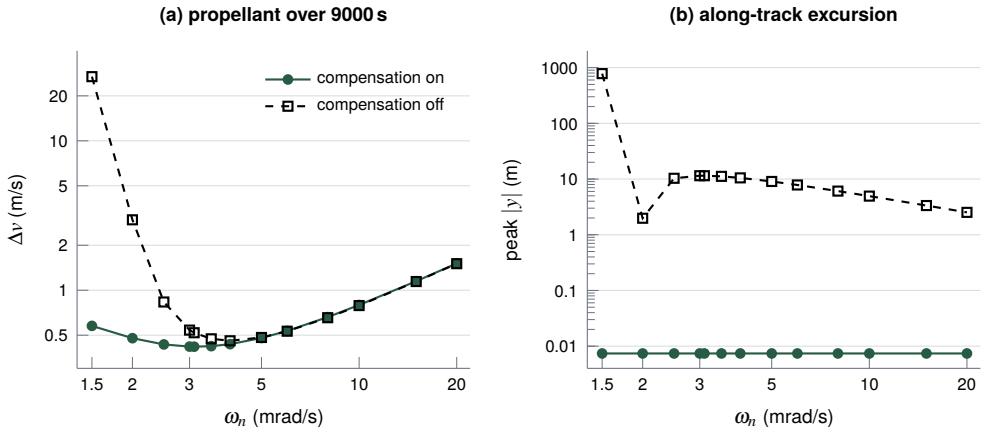


Figure 11.4 Table 11.6 drawn on logarithmic axes, the 100 m radial transfer at thirteen bandwidths with the compensation on (filled circles, solid) and off (open squares, dashed). The row at 3.1 mrad/s is the model run of the table. Notice in (a) that the two propellant curves merge above about 6 mrad/s, the compensated one lowest near 3 mrad/s and the other at 4 mrad/s, and in (b) that the gap in excursion never closes, from 0.0074 m against 2.523 m even at 20 mrad/s.

Table 11.6 The 100 m Radial Transfer at Thirteen Bandwidths, Each Critically Damped, Over a 9000 s Horizon. Peak $|y|$ Is the Largest Along-Track Excursion From a Setpoint the Vehicle Started On. A Dash in a Docked Column Means the Range Never Fell Inside r_{cap} Within the Horizon. The Row Marked † Lies Between the 0.5 mrad/s Settings of the App and Is a Model Run.

ω_n	compensation on			compensation off		
(mrad/s)	Δv (m/s)	docked (s)	peak $\|y\|$ (m)	Δv (m/s)	docked (s)	peak $\|y\|$ (m)
1.5	0.5772	3568.5	0.0074	26.8822	—	780.995
2.0	0.4775	2677.0	0.0074	2.9605	—	1.979
2.5	0.4337	2142.0	0.0074	0.8341	6239.5	10.319
3.0	0.4195	1785.0	0.0074	0.5409	3516.0	11.475
3.1†	**0.4189**	1727.5	0.0074	0.5193	3264.5	11.476
3.5	0.4223	1530.0	0.0074	0.4730	2569.5	11.175
4.0	0.4354	1339.0	0.0074	**0.4589**	2050.5	10.499
5.0	0.4788	1071.0	0.0074	0.4827	1468.5	9.050
6.0	0.5343	892.5	0.0074	0.5318	1143.0	7.828
8.0	0.6607	669.5	0.0074	0.6547	789.0	6.080
10.0	0.7962	535.5	0.0074	0.7896	601.0	4.941
15.0	1.1495	357.0	0.0074	1.1433	378.0	3.345
20.0	1.5110	268.0	0.0074	1.5055	277.0	2.523

cheapest setting is 3.0 mrad/s at 0.4195 m/s. The reference tuning of 10 mrad/s costs 0.7962 m/s, which is 90 % more propellant than the minimum for a docking at 535.5 s instead of 1727.5 s. Rendezvous is flown slowly because that trade is close to linear at the top of the range.

The uncompensated column carries three separate results. It has a floor that the compensated column does not, because Equation (11.11) is unstable whenever $K_P < 3n^2$, and no damping ratio repairs that. At 1.5 mrad/s the radial error grows to 3111.9 m in 9000 s with 26.8822 m/s spent, far outside the 2 km range at which the model is declared valid, while the compensated loop at the same bandwidth docks at 3568.5 s for 0.5772 m/s. Its own propellant minimum is then at 4.0 mrad/s rather than 3.1 mrad/s, displaced upward because the gravity gradient eats part of the nominal gain

and a larger ω_n is needed to achieve the same closed-loop speed. At its own best point it spends 0.4589 m/s against 0.4189 m/s, a penalty of only 9.6 %, and above 6 mrad/s the two propellant columns are indistinguishable to within 1 %. What never converges is the along-track excursion. Even at 20 mrad/s, eighteen times the orbital rate, the uncompensated loop still pulls the along-track axis 2.523 m off a setpoint it started on, against 0.0074 m compensated, a factor of 340. The value of canceling known coupling is not primarily propellant. It is the removal of the cross-axis interference, and that benefit does not diminish with bandwidth. The general rule is that raising loop gain attenuates a disturbance but cannot remove a structural coupling, because the coupling is inside the loop rather than outside it [1].

Pause. Try Exercise 11.3 before reading on.

11.6.3 Thruster Saturation

Set-up 7 starts 200 m above and 200 m behind with the thruster limit cut to 4 mm/s^2, about one twelfth of the default, and leaves the reference tuning in place. The initial demand on each axis is $\omega_n^2 e_0 = 20\,\mathrm{mm/s^2}$, five times what is available, so both in-plane axes pin at the clamp. The command is clipped for 233.5 s of the 2400 s run, and the saturation fraction reads 9.71 % on the radial loop and 5.46 % on the along-track loop. The vehicle still docks at 766.0 s, and the price is 2.2313 m/s of propellant against 2.3015 m/s for the same tuning with the full 50 mm/s^2 available. While saturated the loop is not a proportional-derivative controller at all. It is a constant-acceleration burn, and the poles the gains were chosen to place do not exist.

The remedy is a slower loop rather than a stronger one. Equation (11.13) gives $\sqrt{u_{\max}/e_0} = 4.47\,\mathrm{mrad/s}$, and the measurements bracket it closely. At 4.5 mrad/s the clamp is still active for 38.5 s, and at 4.0 mrad/s it never acts. Slower is also cheaper here, because 4.0 mrad/s spends 1.3504 m/s against 2.2313 m/s at 10 mrad/s, a saving of 39.5 % bought with 877.0 s of extra approach time. Pushing further down to 3.0 mrad/s saves only another 3.6 % while adding 547.0 s, which locates the knee of the trade at about 4 mrad/s for this maneuver.

11.6.4 Feedforward With an Incorrect Mean Motion

Set-up 8 puts the target in a 200 km orbit and loads the flight computer with an ephemeris for 1000 km, so $\hat{n}$ is 15.8 % low, and runs the loop deliberately slowly at $\omega_n = 2\,\mathrm{mrad/s}$. The residuals of Equation (11.12) hand back 30.6 % of the position gain and 9.4 % of the rate gain. Measured against the same case with $\hat{n} = n$, the propellant rises from 0.7444 m/s to 0.9130 m/s, a penalty of 22.6 %, and the chaser is still 12.48 m out after fifty minutes where the matched model has it docked at 2882 s.

The second half of the experiment is the one that generalizes. Leaving the model wrong and raising ω_n to 4 mrad/s drops the propellant penalty from 22.6 % to 2.7 %, from 0.7188 m/s against 0.6995 m/s, and both cases dock. A quantity that is wrong by a fixed fraction stops mattering once the loop closed around it is fast enough, so the question a feedforward model has to answer is not whether it is right but whether it is right enough at the bandwidth being paid for.

The error integrals give a cleaner diagnostic than the propellant does. Exact cancellation makes the two in-plane axes behave identically, and they start from identical 100 m errors in this set-up, so their error integrals should match. With $\hat{n} = n$ they read 98914 and 98 931 m s, equal to within 0.02 %. With $\hat{n}$ wrong they read 143325 and 92 568 m s, a difference of 55 %. Asymmetry between two axes that were set up symmetrically is the signature of residual coupling, and it appears in the reported metrics without any need to plot a trajectory.

11.6.5 Station Keeping on the Two In-Plane Axes

Set-up 9 parks the chaser 20 m behind the target on the velocity vector, on the setpoint, at rest. Over a full hour the propellant counter does not move from 0.0000 m/s and both error integrals stay at zero, because the state is an exact equilibrium of Equation (11.1) and a proportional-derivative controller sitting on an equilibrium commands nothing. Moving the radial setpoint to +20 m before the run

holds the chaser 20 m above that point instead, and it costs 0.2765 m/s per hour, because $3n^2x$ must be canceled continuously while the along-track offset still costs nothing. The 20 m offset and the tuning are the same in both cases, and the choice of axis is the whole difference between free and 6.6 m/s per day.

> **+ DID YOU KNOW? Free Behind the Target, Paid For Above It**
>
> From the simulator: holding 20 m behind the target on the velocity vector spends 0.0000 m/s of propellant in an hour, while holding 20 m above it with the same tuning spends 0.2765 m/s (rows 9 and 9d of Table 11.5). The gravity-gradient term must be canceled for as long as the hold lasts.

Adding differential drag converts the hold into the textbook demonstration of proportional offset. Under a constant along-track acceleration a_d the loop settles where $K_P e = a_d$, and at $a_d = 10\,\mu\text{m/s}^2$ against $K_P = 10^{-4}/\text{s}^2$ the measured offset is 0.100 m, matching a_d/K_P to the resolution of the export. Doubling the drag doubles the offset to 0.200 m. On a real vehicle that offset is either accepted, removed with an integral term, or removed by feeding the estimated drag forward, which is the same mechanism as the orbital compensation one level up.

11.6.6 Feedforward Through a Degraded Thruster

Set-up 10 removes 70 % of the along-track thrust without telling the controller. The commanded propellant total rises from 0.9206 m/s to 2.1882 m/s while only 0.8888 m/s is delivered, and the docking time stretches from 550.5 s to 986.5 s. The mechanism is the feedforward itself. The radial channel is healthy and generates $\dot{x}$ exactly as designed, and every part of that $\dot{x}$ demands $2\hat{n}\dot{x}$ of along-track thrust to cancel the Coriolis term it produces. Only 30 % of that arrives, and the remainder appears as an along-track error that the proportional-derivative loop must then chase with the same crippled thruster. The along-track error integral rises from 15940 to 24 554 m s, a 54 % degradation.

The radial axis is untouched, and that is worth stating precisely because it is easy to expect otherwise. Its error integral reads 15 939 m s in both the healthy and the faulted case, and its error stays inside the 1 m band from 638 s in both. The cancellation $-3\hat{n}^2x - 2\hat{n}\dot{y}$ is exact whatever value $\dot{y}$ takes, so the radial axis stays decoupled from an along-track channel that is failing. Turning the compensation off makes both axes worse rather than better, with the commanded total rising to 2.3175 m/s and the docking time to 1014.0 s, so a degraded actuator is not a reason to abandon the model term. Sweeping the fault from 0 to 95 % of thrust lost over a 3000 s horizon finds no crossover at all. The compensated loop is cheaper at every point, by 1.1 % of commanded propellant with a healthy thruster, 6.0 % at 70 % lost and 2.1 % at 95 % lost, where neither architecture reaches the port. Exact cancellation degrades gracefully with actuator authority, because the term it cancels is proportional to the state rather than to the command. It is a reason to detect the fault, because the mismatch between the commanded acceleration and the acceleration a navigation filter observes is the signal that identifies a failed thruster, and control allocation across a redundant set is what recovers from it. Neither is modeled here. Volume 20, Chapter 6 allocates the demands of its position loops over a set of thrusters and runs a case with one of them failed.

11.6.7 Graded Challenge

The graded set-up starts the chaser 40 m above and 60 m behind the port with 0.02 m/s of residual closing rate and a thruster limit of 25 mm/s^2, and hands over tuning that is wrong in three ways at once. The compensation is off, $\omega_n = 2\,\text{mrad/s}$ puts $K_P = 4 \times 10^{-6}/\text{s}^2$ within 4 % of $3n^2 = 3.84 \times 10^{-6}/\text{s}^2$, and $\zeta = 0.4$ would produce overshoot at a docking port if the loop reached one. At the delivered settings the closest approach is 60.43 m at 518.5 s, after which the range only increases, and the radial error grows from 40 m to 87.86 m by 1500 s with 0.4646 m/s spent. The run is judged at 1500 s. It passes when the chaser has docked inside 3 m below 0.10 m/s without a collision, both in-plane errors have stayed within ± 1 m over the last third of the run, from 990 s, and no more than 0.9 m/s of

propellant has been spent, with the model parameters and disturbances as given. Changing a model parameter or a disturbance voids the attempt.

One successful configuration makes three changes. Compensation is not required for a pass. PD only at 10 mrad/s and damping ratio 1.00 docks near 549 s with a worst scored error of 0.13 m. Turning compensation on removes the terms eating the radial stiffness and costs nothing. Setting $\omega_n = 10\,\mathrm{mrad/s}$ demands $\omega_n^2|e_0| = 7.21\,\mathrm{mm/s^2}$ at the start from $|e_0| = 72.1\,\mathrm{m}$, against a $25\,\mathrm{mm/s^2}$ limit, and the largest thrust magnitude measured over the whole run is $7.00\,\mathrm{mm/s^2}$. Setting $\zeta = 1.00$ removes the overshoot. The retuned run docks at 494.5 s for 0.5750 m/s, inside the budget, its range stays below 1 m from 624 s, and its worst in-plane error over the judged window is 0.032 m. That leaves 0.968 m inside the band and 0.325 m/s under the budget, and the same settings are the suggested tuning, with $\omega_{n,z} = 8\,\mathrm{mrad/s}$ added. The radial error integral falls from 91918 to 7969 m s, a factor of 11.5. The division of the work between the two changes is the result worth recording. The gains moved the response from divergent to a slower pass, and the model term reduced the remaining error within the judged window.

11.6.8 Free Drift in Manual Operation

The first four set-ups are flown by hand with every thruster at zero, and two of them contradict the expected behavior. Releasing the chaser 100 m directly above the target produces 40.7 m per minute of backward drift and a radial excursion to 700 m at half an orbit, from an initial condition in which nothing was moving. Releasing it 100 m behind the target produces nothing at all, to the resolution of the export, for as long as the run is left going. The third carries the Coriolis lesson of the chapter's title. A coast with 0.5 m/s of pure closing speed from 250 m behind, with the relative velocity vector pointing exactly at the docking port, reaches a closest approach of 115.22 m at 363.5 s and is 73.7 m above the port when it gets there, after which it climbs away. Aiming at a target is correct in a fixed frame and incorrect in this one.

11.6.9 Propellant Against Time Under Predictive Control

Table 11.7 sets the MPC beside both proportional-derivative architectures on five approaches, each from the same state over the same horizon, the loops at their set-up's tuning and every MPC row at the 26 min arrival of set-up 11 with no retuning. No plan was rejected, and Table 11.8 reduces each case to a saving and a delay.

On set-up 11, the start of set-up 6 at 3 mrad/s, the MPC docks at 1322.5 s at 0.045 m/s and uses 0.3567 m/s by 3000 s, against 1785.0 s and 0.4191 m/s for the compensated loop, 14.9 % less and 462.5 s sooner. It spends 47 % of its total in the first 10 s. Counted over the pairs it is 4 % above the two-burn transfer arriving at the same 1322.5 s, 0.401 against 0.387 m/s, and on the tile 26 % above, 0.357 against 0.284 m/s, because a diagonal burn costs the pairs more than its length. The price is the along-track axis, which starts on its setpoint and swings 20.85 m behind the target at $t = 626\,\mathrm{s}$, where the compensated loop holds it within 0.007 m, as shown in Figure 11.5. Released above the target, the chaser drifts backward as Equation (11.3) describes, and the planned transfer rides that drift, whereas the compensated loop pays for $3n^2x$ and the Coriolis terms at every step to fly a straight radial line.

The 26 min arrival was tuned on set-up 11 alone, and against loops at 10 mrad/s the trade turns over. On set-ups 7, 3 and 10 the MPC spends 33 to 55 % less than the compensated loop and docks 530 to 905 s later (33 to 57 % and 502 to 820 s against the loop alone), because it keeps its own schedule. On set-up 7 its $4\,\mathrm{mm/s^2}$ thrusters cannot brake the cheaper coast in time, so it passes 23.62 m beyond the target with an envelope slack of 0.116 m/s. In every set-up in which the MPC acts there is no collision, and every contact lies between 0.045 and 0.067 m/s. On set-up 8, the one mismatch case, the MPC docks at 1516.0 s on 0.5894 m/s while neither loop docks within 3000 s, because replanning every 2 s from the measured state absorbs the error the feedforward carries.

Moving the arrival time on set-up 11 traces the MPC's own trade. Over 3000 s, 10 min costs 0.645 m/s for a docking at 562.5 s, 20 min costs 0.439 at 1144.5 s and 40 min costs 0.382 at 2294.0 s, against 0.3567 m/s at 1322.5 s for the shipped 26 min. Set beside the compensated column of Ta-

Table 11.7 The Three Architectures on Five Approaches. Δv Is the Tile of Equation (11.6) at the End of the Horizon, Pairs Is the Propellant of the Opposed Thruster Pairs That the MPC Minimizes, Contact Is the Relative Speed at Docking, Past Port Is the Largest Distance Beyond the Setpoint on an Axis That Started More Than 1 m Off, and Excursion Is the Largest Error of an Axis That Started on Its Setpoint. Set-Ups 5 and 6 Supply the Proportional-Derivative Rows of Set-Up 11.

					Docked		Inside	Past		IAE		
#	Architecture	Horizon (s)	Δv (m/s)	Pairs (m/s)	at (s)	Contact (m/s)	1 m (s)	port (m)	Excursion (m)	radial (m s)	along-track (m s)	Plans
11	PD only (set-up 5)	3000	0.5294	0.6723	—	—	—	0.00	11.475	113888	21896	—
11	PD + comp. (set-up 6)	3000	0.4191	0.5103	1785.0	0.008	2212.5	0.00	0.007	66536	8	—
11	MPC	3000	**0.3567**	0.4014	1322.5	0.045	1367.0	0.00	20.847	44839	18908	1500
7	PD only	2400	2.2408	2.9446	863.5	0.021	1029.5	1.95	—	64809	48875	—
7	PD + comp.	2400	2.2313	3.0036	766.0	0.026	892.0	0.00	—	61475	51357	—
7	MPC	2400	**1.4901**	1.7837	1545.0	0.064	1585.5	23.62	—	82401	254033	1200
3	PD only	1800	1.5891	2.0137	706.0	0.020	880.0	0.13	12.935	5883	44939	—
3	PD + comp.	1800	1.5508	1.9581	621.5	0.026	747.0	0.00	0.010	5	44838	—
3	MPC	1800	**0.9134**	1.0648	1526.0	0.057	1586.5	0.00	62.858	61467	191652	900
8	PD only	3000	1.9458	2.0930	—	—	—	0.00	—	412324	128933	—
8	PD + comp.	3000	0.9130	1.1528	—	—	—	0.32	—	143325	92568	—
8	MPC	3000	**0.5894**	0.7281	1516.0	0.059	1580.5	14.08	—	39814	118641	1500
10	PD only	1800	2.3175	2.6513	1014.0	0.035	1601.0	18.36	—	19045	25415	—
10	PD + comp.	1800	2.1882	2.5354	986.5	0.035	1551.5	15.91	—	15939	24554	—
10	MPC	1800	**0.9875**	1.1434	1516.0	0.067	1572.0	14.95	—	32770	99709	900

Table 11.8 The Saving and the Delay of the MPC at the 26 min Arrival Time, Against Each Proportional-Derivative Architecture. A Negative Delay Is an Earlier Docking, and a Dash Means That Architecture Never Docked.

	MPC		against PD + comp.		against PD only	
#	Δv (m/s)	docked (s)	less Δv (%)	delay (s)	less Δv (%)	delay (s)
11	0.3567	1322.5	14.9	−462.5	32.6	—
7	1.4901	1545.0	33.2	779.0	33.5	681.5
3	0.9134	1526.0	41.1	904.5	42.5	820.0
10	0.9875	1516.0	54.9	529.5	57.4	502.0
8	0.5894	1516.0	35.4	—	69.7	—

ble 11.6, these points compare the two controllers at equal docking time, and each lies below that column. Industrial predictive controllers usually carry their economics as a linear objective in a steady-state target calculation [18], while this one prices propellant linearly inside the dynamic plan itself.

On every hold of Table 11.9 the MPC spends what the compensated loop spends, and under drag it holds 4.2 and 6.5 times closer, with no integral action in either. The graded challenge is unchanged, and the MPC fails it at the shipped arrival with either tuning, because it reads neither. At the 1500 s verdict it has not docked, its worst in-plane error from 990 s is 35.35 m and it has spent 0.217 m/s, since 26 min lies at 1560 s. With $T_{arr} = 10$ min, not shipped, it docks at 542.5 s on 0.424 m/s and passes.

Halving the physics step to 0.25 s leaves the Δv of every architecture on set-ups 3, 7, 8, 10 and 11

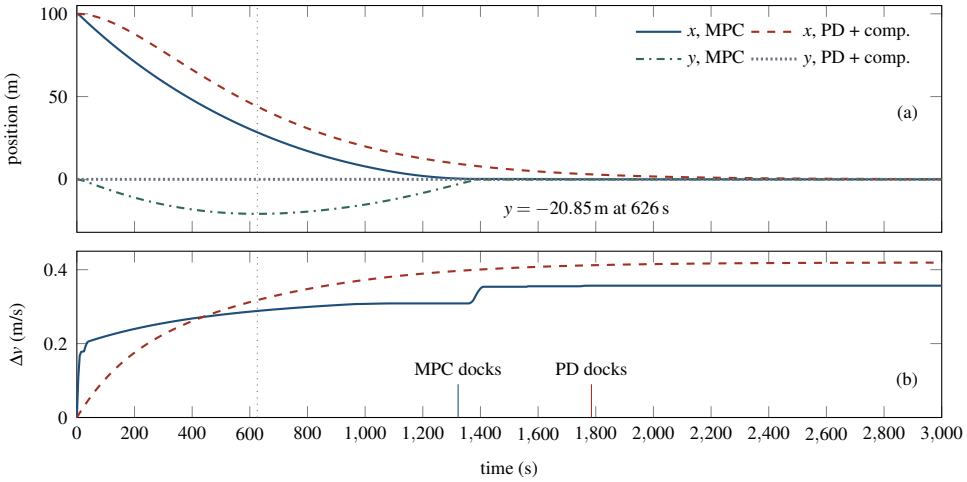


Figure 11.5 Set-up 11 under the MPC and the same start under PD with compensation, both at 3 mrad/s. (a) Radial position x under the MPC (solid) and PD (dashed), and along-track position y under the MPC (dash-dotted) and PD (heavy dotted). (b) Propellant used, the tile of Equation (11.6), under the MPC (solid) and PD (dashed). At 626 s (thin vertical dotted line) the MPC has spent 0.288 m/s against 0.318 m/s and its along-track axis is 20.85 m behind the target, where the compensated loop holds it within 0.007 m. The MPC docks at 1322.5 s and the loop at 1785.0 s, and by 3000 s they have used 0.3567 and 0.4191 m/s.

Table 11.9 Station Keeping Over 3600 s and the Graded Challenge at Its 1500 s Verdict. Each Hold Entry Gives Δv in m/s and the Largest In-Plane Error in m. The Challenge's Worst Error Is Judged From 990 s, and Its 10 min Arrival Is Not Shipped.

Hold	PD only	PD + comp.	MPC
9, at $y = -20$ m	0.0000, 0.000	0.0000, 0.000	0.0000, 0.000
9b, $a_d = 10\,\mu\text{m/s}^2$	0.0360, 0.100	0.0360, 0.100	0.0360, 0.024
9c, $a_d = 20\,\mu\text{m/s}^2$	0.0720, 0.200	0.0720, 0.200	0.0720, 0.031
9d, $x_{SP} = +20$ m	0.2869, 0.799	0.2765, 0.000	0.2765, 0.001

Graded challenge	Δv	docked (s)	worst (m)
delivered, PD, fails	0.4646	—	87.86
retuned, comp., meets	0.5750	494.5	0.03
MPC, 26 min, fails	0.2170	—	35.35
MPC, 10 min, meets	0.4243	542.5	0.00

unchanged to four decimals and every docking time within 0.2 s. In Chrome on the built app a plan takes 2.8 to 6.2 ms on average and at most 30.5 ms across set-ups 7, 9, 11 and the graded challenge, which is 0.14 to 0.31 % of a core at real time and 4.3 to 9.2 % at the top playback speed of 30 times, measured in two runs at load averages of 1.3 to 2.8 on the host. At that speed the page keeps a mean frame of 16.7 ms, with a 99th percentile of 16.8 ms, and the graded challenge takes 50.0 s of wall time under both the MPC and the loop alone.

11.7 Technology in Practice

Figure 11.6 puts the compensated loops of Equation (11.9) on flight hardware: relative navigation sensors feed a filter in the flight computer, the feedforward and feedback demands are summed there, and the sum is turned into firings of a set of reaction control thrusters.

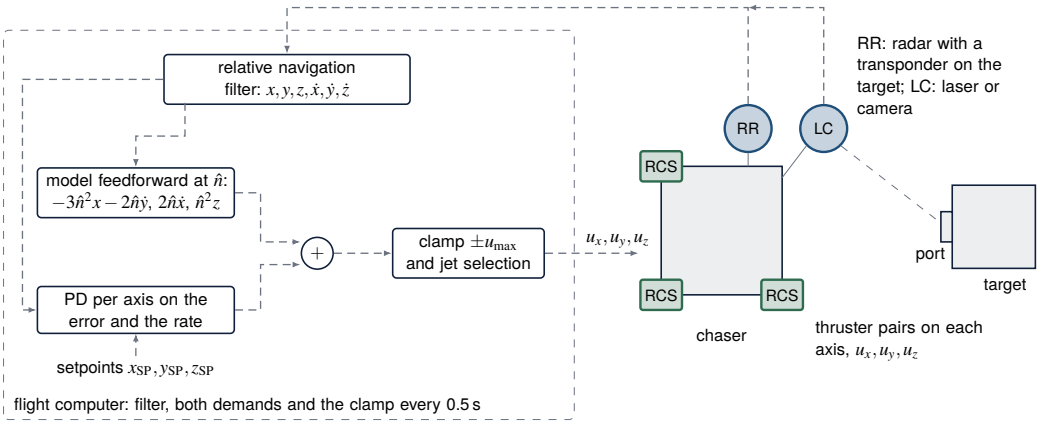


Figure 11.6 The compensated loops of Equation (11.9) on flight hardware. A radar working with a transponder on the target and a laser or camera sensor feed a relative navigation filter in the flight computer. The model feedforward and the PD demand are summed, clamped and turned into firings of thruster pairs on the chaser. The application replaces the sensors and the filter by the exact state and the jet selection by a clamp on each axis. Solid outlines are the vehicles and their devices, and dashed gray arrows are signals. Notice that the feedforward block reads the same state as the feedback block and one number, $\hat{n}$, and nothing else.

What Measures It

The controller of this chapter reads the six relative states exactly, which Section 11.3.5 names as the largest single idealization of the model. The orbiter's rendezvous radar reached the end of its tracking envelope at about 80 feet even on small payloads, and visual ranging with a sight or with camera tilt angles was imprecise, loaded the crew and gave no direct range rate [11]. Studies of docking with a space station later found the radar inadequate in minimum range and accuracy, which led to hand-held and payload-bay laser sensors [11]. In the model each reading is the true state, so none lags, drifts or jumps.

> **TECHNOLOGY** **Sensor: Rendezvous Radar With a Transponder**
>
> The Apollo lunar module measured range, range rate, line-of-sight angle and angle rate to the command module with a continuous-wave tracking radar that worked with a transponder on the other vehicle. Range came from the time delay of a three-tone phase modulation and range rate from the two-way Doppler shift. Its listed minimum range was 80 feet. In flight its range agreed with a separate radio ranging system to a mean bias of less than 0.04 percent of the median range. In flight tests the range tone tracking loop slipped cycles at low signal-to-noise ratio, and each slip moved the range reading by a multiple of 2400 feet [19], a step on a range trend that the vehicle never made.

What Moves It

The bounded acceleration u_{max} on each axis of Table 11.2 stands for a set of reaction control thrusters, which Section 11.3.5 describes as on and off devices with a minimum impulse bit. The orbiter carried primary jets of 870 pounds thrust and vernier jets of 25 pounds [11], and the plumes of the large jets shaped how it approached a target. A real vehicle commands opposed pairs of thrusters, which is what the propellant of the predictive controller counts in Section 11.4.6, while the per-axis clamp of the model delivers any command up to u_{max} exactly.

TECHNOLOGY Final Element: Reaction Control Thrusters

Thrusters give a spacecraft both translation and attitude control. A cold gas thruster has no combustion and expels a stored gas through a nozzle, which gives very small minimum impulse bits for precise maneuvering, while its low specific impulse rules out large maneuvers. The minimum impulse bit sets how precisely the vehicle translates and the thrust sets its authority. Depending on the thruster layout and the center of gravity, the thrusters also impart torques that attitude control must absorb. One recent small proximity mission lost thrust on several axes, likely after a controller failure, and others met propellant leaks [20]. On a trend a lost thruster shows as a command that no longer produces its acceleration, the gap of set-up 10.

Where the Controller Runs

On the orbiter, guidance, navigation and control software ran in two flight computers in parallel during rendezvous and proximity operations, and in three from the station flights onward. Clohessy-Wiltshire targeting was first preferred for proximity operations because it took about 5 seconds on the flight computer, against 30 to 60 seconds expected for Lambert targeting, which in practice later took less than 15 seconds [11]. The parts of the report read for this section give no execution period for the orbiter's digital autopilot. The application evaluates the law of Equation (11.9) every 0.5 s, and the four feedforward terms are added to the feedback demand before the clamp, so feedback and feedforward share one layer of the controller.

IN PRACTICE Braking Gates and the Approach Axes

Piloted approaches were flown against braking gates, a closing rate assigned to each range. A shuttle profile of the 1970s gave 45 feet per second at 9000 feet, falling in steps to 5 at 1000 feet and to zero at 100 feet, with coast periods long enough for longer burns caused by dispersions or failed jets. The pilot controlled range, range rate and line-of-sight angles directly, and final approaches ran along the velocity vector or the radius vector. An approach from below along the radius met what the source calls the natural braking effect of orbital mechanics [11], which is the term $3n^2x$ of Equation (11.1) pushing a vehicle below the target further down.

What Is Modernizing It

Three developments bear on these loops. The first is light detection and ranging, an emerging spacecraft sensor that matured in terrestrial use and already flies on larger spacecraft capable of proximity operations, with rendezvous and docking among its uses [20]. The second is camera-based navigation. A NASA demonstration launched in 2023 navigated relative to an object from angles alone, with no prior knowledge of it, and reached a relative position knowledge with 0.5 percent error relative to range [20]. Both address the navigation error that the exact state of this chapter hides. The third is autonomous guidance, navigation and control software for operations when ground communication is limited or unavailable [20]. The feedforward of this chapter is the least demanding part of such software.

11.8 Limitations

The simulator teaches a coupling structure and does not design an approach. The controller is handed a perfect state vector, so the relative navigation that dominates real terminal guidance is absent and every number in Table 11.5 is tuning-limited rather than sensor-limited. There is no attitude, so the loop that actually consumes most of a chaser's control authority and most of its engineering effort does not appear, and thrust is available along any axis at any instant. Continuous bounded acceleration replaces a set of on and off thrusters with a minimum impulse bit and a geometry that couples translation into rotation, which is a substantial softening of the terminal phase. The control structures are fixed, and the one constrained formulation among them, the predictive controller of Section 11.4.6, carries only the thrust bound and a soft closing-speed envelope, which puts the corridor and keep-out constraints of current proximity-operations work [6, 7] and the passively safe

trajectory families [5] outside anything a run can show, even though Equation (11.1) is precisely the prediction model each of them carries. The linearization itself is declared valid only inside 2 km, and it ignores the oblateness term that dominates relative motion over many orbits [14]. Disturbances are either constant or fired by the operator and never stochastic, and one fixed step against one fixed control period makes every repeat bit-identical, which no flight computer achieves. The individual failure modes of every modeling assumption are given in Section 11.3.5, and the expensive ones are the exact navigation, the absent attitude dynamics and the continuous thruster.

The predictive controller's model leaves out the drag, a degraded thruster, an error in $\hat{h}_{orb}$ and the impulsive events, which it meets only by replanning every 2 s without an estimator, and its one setting, the 26 min arrival, was chosen on one slow case, so a different value moves every delay in Table 11.8.

> ⚠ **OPEN CHALLENGE** **Proximity Operations on Small Spacecraft**
>
> A small spacecraft that is to approach another needs a suite of sensors for relative position and attitude, sometimes of an object it has no prior knowledge of, and much of its navigation must run without commands from the ground. Many recent demonstrations had trouble with their control systems and actuators, and one ran out of propellant before it could attempt its docking. The source expects the field to grow, with more demand for software that has proven flight heritage [20]. Progress would be such software flying routinely, with navigation error rather than tuning setting the docking precision, as Section 11.3.5 says it does.

11.9 Discussion

A spacecraft that cannot rendezvous is a spacecraft that must carry everything it will ever need. Every station resupply flight, every crew rotation and every proposal to refuel or repair a satellite already in orbit reduces to the same terminal problem of arriving slowly at a moving object with a propellant budget that cannot be topped up on the way. On-orbit servicing has been studied for decades and its enabling technologies are surveyed in [21], and the same approach and capture phase is the hard part of every proposal to remove a piece of large debris from a congested orbit [22]. What separates those missions from a station resupply is that the object being approached is uncooperative, tumbling and sometimes not designed to be grappled at all.

The simulator adds one statement to that picture, about where the difficulty does not lie. The terminal phase is not hard because the dynamics are unknown. Equation (11.1) is one of the best-conditioned models in engineering, its single parameter is known to many digits, and a controller that uses it collapses a coupled unstable plant to three independent second-order systems with four multiplications. The difficulty is in the navigation, the attitude, the discrete actuators and the consequences of an abort, none of which the model removes. Recognizing which part of a problem the model solves, and refusing to spend feedback authority on the part it already solved, is the transferable skill, and it is the same decomposition that appears as computed-torque control in robotics, as gravity and Coriolis compensation on gantry machines, and as static feedforward in process control [9]. In Volume 6, Chapter 2 one static feedforward from the pump command to the concentrate valve removes most of the interaction between the two loops of a reverse osmosis train, which is the same move on a process plant.

11.10 Conclusion

Relative motion near a target in orbit is written in axes that turn once per orbit, and the terms this produces make three thrusters something other than three independent actuators. The simulator of this chapter integrates the six Clohessy-Wiltshire states about a 400 km circular orbit at $n = 1.131\,\mathrm{mrad/s}$, with a bounded thruster acceleration per axis, a differential-drag disturbance, a de-ratable thruster and a geometric docking test, and closes one proportional-derivative loop per axis around them every 0.5 s.

Adding four algebraic terms evaluated at the controller's own mean motion cancels the plant's coupling identically and reduces each axis to a damped second-order system, verified against the analytic response to 0.148 m on a 100 m transfer. On the central comparison the compensated loop docks in 1785 s for 0.4191 m/s and holds its along-track axis within 0.0074 m, where the same gains without the model terms take 3516 s, spend 28.8 % more propellant and pull the along-track axis 11.475 m off a setpoint it started on.

A bandwidth sweep locates a propellant minimum at 3.14 mrad/s against a predicted $n\sqrt{3\mathrm{e}} = 3.23\,\mathrm{mrad/s}$, and shows that enough bandwidth buys away the propellant penalty but never the cross-axis excursion, which is still a factor of 340 apart at 20 mrad/s. Twelve set-ups separate the unstable radial mode, the free along-track equilibrium, a 115.22 m miss from a velocity vector pointed at the port, thruster saturation, a 15.8 % model error and a thruster delivering 30 % of its command, and each carries its own completion rule, with the graded challenge passed only by a docking that holds within 1 m on a 0.9 m/s budget. An added model predictive controller on the same exact model docks the start of set-up 6 462.5 s sooner on 14.9 % less propellant by letting the along-track axis swing 20.85 m, and against loops tuned at 10 mrad/s it saves 33 to 55 % while docking 530 to 905 s later. Throughout, a known coupling can be switched off and the cost of doing so read from the same measurements.

KEY POINTS

- The target's frame turns once per orbit, so three thrusters are a coupled plant: a radial offset drifts away, a hold behind is free.
- Four algebraic terms at the controller's own mean motion cancel the coupling and leave three damped second-order loops.
- With the same gains, the compensated loop docks in 1785 s for 0.4191 m/s, and the uncompensated one takes 3516 s and pulls the along-track axis 11.475 m off.
- More bandwidth buys away the propellant penalty but never the cross-axis excursion, still a factor of 340 apart at 20 mrad/s.
- The feedforward is itself thrust, so propellant is least near $\omega_n = n\sqrt{3\mathrm{e}}$.

Where else this applies. Volume 6, Chapter 2 removes loop interaction with one static feedforward, and Chapter 10 feeds gravity forward on another vehicle.

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] W. H. Clohessy and R. S. Wiltshire, "Terminal guidance system for satellite rendezvous," *Journal of the Aerospace Sciences*, vol. 27, no. 9, pp. 653–658, 1960.

[3] G. W. Hill, "Researches in the lunar theory," *American Journal of Mathematics*, vol. 1, no. 1, p. 5, 1878.

[4] H. B. Hablani, M. L. Tapper, and D. J. Dana-Bashian, "Guidance and relative navigation for autonomous rendezvous in a circular orbit," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 3, pp. 553–562, 2002.

[5] L. Breger and J. P. How, "Safe trajectories for autonomous rendezvous of spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 5, pp. 1478–1489, 2008.

[6] S. Di Cairano, H. Park, and I. Kolmanovsky, "Model predictive control approach for guidance of spacecraft rendezvous and proximity maneuvering," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 12, pp. 1398–1427, 2012.

[7] A. Weiss, M. Baldwin, R. S. Erwin, and I. Kolmanovsky, "Model predictive control for spacecraft rendezvous and docking: Strategies for handling constraints and case studies," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 4, pp. 1638–1647, 2015.

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

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

[10] J. L. Goodman, "History of space shuttle rendezvous and proximity operations," *Journal of Spacecraft and Rockets*, vol. 43, no. 5, pp. 944–959, 2006.

[11] ——, "History of space shuttle rendezvous," NASA Johnson Space Center, Mission Operations Directorate, Flight Dynamics Division, Houston, TX, Report JSC-63400, Revision 3, Oct. 2011, nASA Technical Reports Server 20110023479.

[12] W. Fehse, *Automated Rendezvous and Docking of Spacecraft*. Cambridge, UK: Cambridge University Press, 2003.

[13] D. C. Woffinden and D. K. Geller, "Navigating the road to autonomous orbital rendezvous," *Journal of Spacecraft and Rockets*, vol. 44, no. 4, pp. 898–909, 2007.

[14] S. A. Schweighart and R. J. Sedwick, "High-fidelity linearized J_2 model for satellite formation flight," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 6, pp. 1073–1080, 2002.

[15] 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.

[16] 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.

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

[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] P. Rozas and A. R. Cunningham, "Apollo experience report: Lunar module landing radar and rendezvous radar," NASA Manned Spacecraft Center, Houston, TX, NASA Technical Note TN D-6849, 1972, nASA Technical Reports Server 19720016521.

[20] S. V. Weston, C. D. Burkhard, J. M. Stupl, R. L. Ticknor, B. D. Yost, R. A. Austin, P. Galchenko, L. K. Newman, and L. Santos Soto, "State-of-the-art small spacecraft technology," National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA, Technical Publication NASA/TP-20250000142, Jan. 2025. [Online]. Available: https://ntrs.nasa.gov/citations/20250000142

[21] A. Flores-Abad, O. Ma, K. Pham, and S. Ulrich, "A review of space robotics technologies for on-orbit servicing," *Progress in Aerospace Sciences*, vol. 68, pp. 1–26, 2014.

[22] M. Shan, J. Guo, and E. Gill, "Review and comparison of active space debris capturing and removal methods," *Progress in Aerospace Sciences*, vol. 80, pp. 18–32, 2016.

11.11 Exercises

11.1 ○ Concept check · 1 min

In scenario 3 the chaser coasts from 250 m behind the target with 0.5 m/s of closing speed aimed exactly at the port, and it passes 73.7 m above the port. Which term of Equation (11.1) starts the rise?

(a) The Coriolis term $2n\dot{y}$ in the radial equation
(b) The gravity-gradient term $3n^2x$ in the radial equation
(c) The Coriolis term $-2n\dot{x}$ in the along-track equation
(d) The differential-drag term a_d in the along-track equation

11.2 ○ Concept check · 2 min

Which structure does the default controller of the rendezvous simulator use?

(a) A cascade with an inner thrust-acceleration loop on each axis
(b) A PD loop per axis plus a feedforward that cancels the gravity-gradient and Coriolis terms
(c) A PID loop per axis whose integral action removes the coupling as a steady load
(d) A steady-state decoupler computed from the plant's matrix of process gains

11.3 ○ Concept check · 2 min

In Table 11.6, above 6 mrad/s the propellant used with and without the compensation agrees within 1 %. At 20 mrad/s, what does the compensation still buy?

(a) A docking time about half as long
(b) A propellant saving that grows with the bandwidth
(c) An along-track excursion of 0.0074 m in place of 2.523 m
(d) Nothing that the scorecard or the exported file can show

11.4 Theory and calculation · 6 min

Use $\mu = 3.986\,004\,418 \times 10^{14}\,\mathrm{m^3/s^2}$ and $R_E = 6\,378\,137\,\mathrm{m}$.

(a) Compute the mean motion n and the period T of a 400 km circular orbit from Equation (11.2). *Checkpoint:* $n = 1.131\,367\,\mathrm{mrad/s}$, $T = 5553.6\,\mathrm{s} = 92.56\,\mathrm{min}$.

(b) A chaser is released at rest 100 m above the target. From Equation (11.3), give the slope of the secular along-track drift in m/min, and x and y at half a period. *Checkpoint:* 40.7 m/min backward. $x = 700\,\mathrm{m}$, $y = -6\pi x_0 = -1885\,\mathrm{m}$.

(c) Give the amplitude of the radial arc for a pure closing speed of 0.5 m/s, and of the cross-track swing for $\dot{z}_0 = 0.10\,\mathrm{m/s}$. *Checkpoint:* 883.9 m and 88.4 m.

11.5 Theory and calculation · 6 min

Use Equations (11.11) and (11.12) with $K_P = \omega_n^2$, $K_D = 2\zeta\omega_n$ and $\zeta = 1$.

(a) At 400 km, what fraction of the radial position gain does the gravity gradient remove from the uncompensated loop at $\omega_n = 3$ and 10 mrad/s, and at what ω_n is the radial stiffness gone? *Checkpoint:* 42.7 %, 3.84 %, 1.9596 mrad/s.

(b) Scenario 8 flies a 200 km orbit with a model for 1000 km. Compute n, $\hat{n}$, the residual $3(n^2 - \hat{n}^2)$ and the residual $2(n - \hat{n})$. *Checkpoint:* $n = 1.1834\,\mathrm{mrad/s}$, $\hat{n} = 0.9962\,\mathrm{mrad/s}$, $1.224 \times 10^{-6}\,\mathrm{/s^2}$, $3.743 \times 10^{-4}\,\mathrm{/s}$.

(c) Express the two residuals as fractions of K_P and K_D at $\omega_n = 2$ and 4 mrad/s, and say why doubling the bandwidth shrinks one fraction faster than the other. *Checkpoint:* 30.6 % and 9.4 % at 2, 7.6 % and 4.7 % at 4.

11.6 Theory and calculation · 7 min

With the cancellation exact and $\zeta = 1$, the radial error from rest at x_0 is $x(t) = x_0(1 + \omega_n t)\,\mathrm{e}^{-\omega_n t}$ and the feedback command is $a_x = \ddot{x}$. Take $x_0 = 100\,\mathrm{m}$ at 400 km.

(a) Show that $\int_0^\infty |\ddot{x}|\,\mathrm{d}t = 2\omega_n x_0/\mathrm{e}$ and $\int_0^\infty x\,\mathrm{d}t = 2x_0/\omega_n$, so that the feedforward $-3n^2x$ spends $6n^2x_0/\omega_n$. *Checkpoint:* $0.736\,\omega_n x_0$ and $2x_0/\omega_n$.

(b) Minimize the sum of the two parts over ω_n and evaluate the minimum and the sum there. *Checkpoint:* $\omega_n^\star = n\sqrt{3\mathrm{e}} = 3.23\,\mathrm{mrad/s}$, about 0.475 m/s.

(c) Table 11.6 gives 0.4189 m/s near this bandwidth. Why is the estimate higher, and why is its location still close? *Checkpoint:* The estimate is about 13 % high.

11.7 App challenge · 8 min

Use the default seed and leave every other setting as the scenario loads it.

(a) Load scenario 5, PD only, no orbital compensation, and run to 3000 s. Read the Δv used tile, the radial and along-track IAE on the scorecard, and the largest $|$`along_track_error_m`$|$ in the exported file. *Checkpoint:* 0.529 m/s, 113888 and 21 896 m s, 11.475 m, not docked.

(b) Load scenario 6, PD + orbital compensation, and run to 3000 s, pressing Run again after the pause at docking. Read the same four values and the first row of the `docked` column that reads 1. *Checkpoint:* 0.419 m/s, 66536 and 8.17 m s, 0.0074 m, docked in the 1785 s row.

(c) Explain where the along-track error of (a) comes from, and why the radial approach of (a) is slower.

Integrals within 1 %.

11.8 App challenge · 8 min

Load scenario 7, Limited thruster authority, and use the default seed. The thrust limit is 4 mm/s^2, the start 200 m above and 200 m behind. Each run lasts 2400 s. Press Run again after the pause at docking.

(a) Change nothing and run. Read the Δv used tile, the first exported row with `docked` at 1, and the time at a limit for the radial and along-track loops on the scorecard. *Checkpoint:* 2.231 m/s, 766 s, 9.71 % and 5.46 %.

(b) Compute the bound of Equation (11.13). Press Reset, set In-plane bandwidth to 4.5 mrad/s before Run and repeat. Then Reset, set 4.0 mrad/s and repeat. *Checkpoint:* 4.47 mrad/s. 1.408 m/s, 1463 s, 1.58 % and 0.04 %. 1.350 m/s, 1643 s, no time at the limit.

(c) What fraction of the propellant does the move from 10 to 4 mrad/s save, and what does it cost in time? *Checkpoint:* 39.5 %, 877 s later.

11.9 App challenge · 9 min

Load the challenge, Dock on a Δv budget, and use the default seed. Each run lasts 1500 s. Start every part from the reload button beside the scenario list, which restores the shipped settings, and make the stated settings before Run.

(a) Change nothing. Report the verdict line. *Checkpoint:* Not docked, worst in-plane error 87.86 m, Δv 0.465 m/s.

(b) Set only Control architecture to PD + orbital comp. *Checkpoint:* Docked at 998 s, worst 14.96 m, 0.376 m/s, fails.

(c) Keep PD only and set In-plane bandwidth to 10 mrad/s and Damping ratio to 1.00. *Checkpoint:* Docked at 549 s, worst 0.13 m, 0.583 m/s, meets the target.

(d) Make all three changes. *Checkpoint:* Docked at 495 s, worst 0.03 m, 0.575 m/s, meets the target.

Which changes are needed to pass, and what does the compensation add once they are made?

11.10 Application area · 3 min

Set-up 9 measures that holding 20 m above the target costs 0.2765 m/s per hour, while holding 20 m behind it costs nothing.

(a) Show that the hourly cost is $3n^2x \times 3600\,\mathrm{s}$ and give the cost per day. *Checkpoint:* 6.64 m/s per day.

(b) A 10 000 kg chaser burns propellant at an effective exhaust velocity of 2157 m/s (a specific impulse of 220 s). How much propellant does a day at the radial hold cost? *Checkpoint:* About 31 kg.

11.11 Application area · 3 min

Take a differential drag of 1 µm/s^2, the order the app's briefing for scenario 9 gives for two dissimilar vehicles at 400 km and a tenth of the 10 µm/s^2 of set-up 9b.

(a) What along-track offset does the PD loop hold at the reference tuning, 10 mrad/s, and at the slow tuning, 3 mrad/s? *Checkpoint:* 0.010 m and 0.111 m.

(b) What does holding against it cost per orbit and per hour? *Checkpoint:* 5.6 mm/s per orbit, 3.6 mm/s per hour.

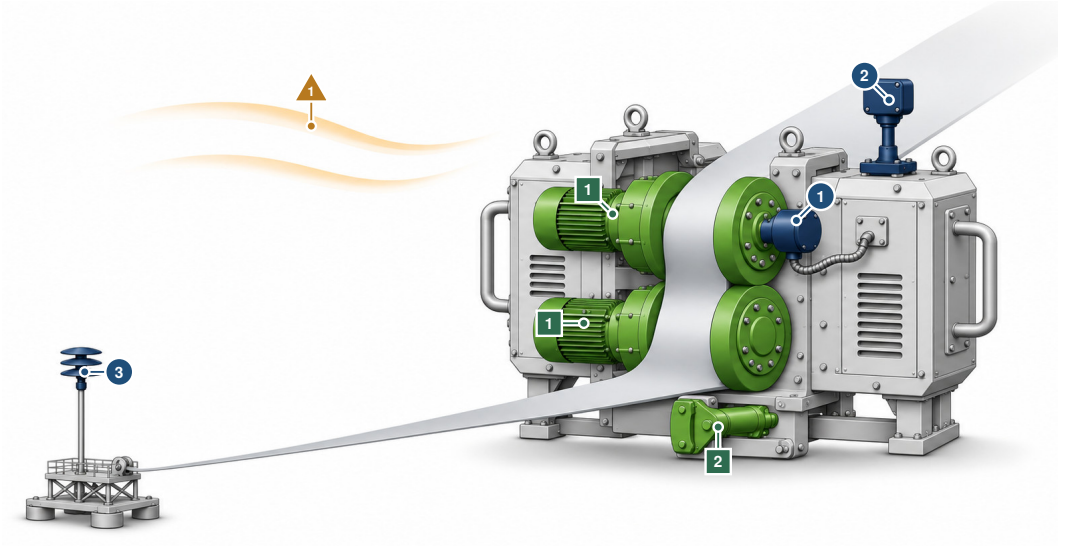
Name the three remedies for the offset and say which one an operator would choose at a 20 m hold.

11.12 Review the agent · 5 min

An engineer asks an AI assistant whether the orbital feedforward is worth keeping. It replies: "Table 11.6 shows that above 6 mrad/s the two architectures spend the same propellant within 1 %. Drop the feedforward, tune PD only at 10 mrad/s, and the chaser docks at 601 s instead of 535.5 s for the same fuel, with no exposure to an error in the mean motion." Find the flaw, and say how you would check it in the app.

CHAPTER 12

Gravity Feedforward and Input Shaping for a Speculative Space Elevator Climber



The controller moves 1 motors that drive the rollers 2 friction brake

The controller reads 1 roller speed 2 height, by ranging 3 wind report from a profiler

What disturbs it 1 crosswind and gusts

Notice that the climber hangs on a long, slowly swaying ribbon, so every change of its speed and every gust of crosswind pushes it sideways. Illustration, not to scale.

IN PLAIN TERMS Climbing a Ribbon Into Space

The main idea. A space elevator would be a very long ribbon hanging down from high above the Earth. A machine would climb it on motor-driven rollers, and its controller is told in advance how its weight shrinks with height.

The challenge. Each time the climber speeds up or slows down, the turning Earth pushes it sideways and the ribbon swings slowly. Wind, motor heat and the power supply also limit how fast it may climb.

Why it matters. No such elevator has been built, but planning a move on a long, swaying cable matters for cranes and mine hoists too.

THREE WAYS TO READ THIS CHAPTER

Explore	13 min	the first page, the fact on the ribbon's length (Section 12.1), the two history segments (Section 12.2), the two try-it segments and the fact on the descent (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 two open challenges

The control question A climber on a space elevator ribbon carries a weight that falls from 9.76 m/s^2 at the anchor to zero at the geostationary radius, and every change of its speed pushes the ribbon sideways through the Coriolis force. What do a gravity feedforward, a speed schedule, an input shaper and a feasibility supervisor each contribute when the load, the structure and the limits all change along the route?

Learning objectives

- Compute the effective gravity, the geostationary radius and the error of a constant-gravity feedforward at any height on the route.
- Estimate the period and damping of the climber's local mode from the string below it and the impedance of the ribbon above.
- Design ZV and ZVD shapers from a predicted mode and explain why they remove the ringing after a move but not the lag during it.
- Explain why a speed schedule slows the climber in a crosswind and on a descent, and read the price in trip time.
- Combine schedule, shaper and feedforward to pass the graded transfer inside every limit.

Prerequisites Sections 1.4.2 (where a disturbance enters), 1.4.3 (static feedforward), 1.6.2 (velocity form) and 1.6.4 (scenarios and the challenge). Chapter 10 for a gravity feedforward on a vehicle and Chapter 11 for motion in a rotating frame.

The app https://apmonitor.com/apps/space-elevator/

See also Volume 7, Chapter 12 for ZV and ZVD shapers on a swinging load with an assumed cable length, and Volume 20, Chapter 3 and Volume 20, Chapter 4 for load-sheet feedforward and supervised flight on vehicles whose load and authority change along the route.

12.1 Introduction

Every hoist in ordinary service carries a load whose weight does not change on the way up. An elevator car, a crane hook and a mine skip weigh the same at the top of their travel as at the bottom, and what changes with the operating point is the rope. A deep mine hoisting cable changes length as the conveyance travels, and its coupled lateral and longitudinal response is governed by equations that accommodate that non-stationary nature [1]. Command shaping grew up around flexible systems of this kind, where a commanded move excites a lightly damped mode that rings long after the move ends [2–4], and high-performance motion control pairs a trajectory with bounded derivatives with a feedforward matched to it [5].

A space elevator would reverse that arrangement. A climber rides a ribbon from an anchor on the equator through the geostationary radius to a counterweight beyond it [6–8], and in the frame that turns with the Earth its weight is set by the effective gravity $\mu/r^2 - \omega^2 r$. On the ribbon simulated here that is 9.764 m/s^2 at the anchor, 9.318 m/s^2 at the 150 km station and zero at 42 164 km from the Earth's center, so the drive force that holds an 11 t climber still falls from 107.4 kN to nothing over the route. The ribbon is a string whose slowest lateral mode has a period of 127.4 h. A climber moving outward at speed v is pushed west by the Coriolis force $2m\omega v$, only 80.2 N for 11 t at 50 m/s. Its guide rollers pass that push to the ribbon, whose primary response to an ascending climber is a rotation opposite to the Earth's rotation [9]. Every change of the climber's speed changes that push.

> ✦ DID YOU KNOW? **Three Weeks End to End**
>
> A NASA-sponsored study pictured an elevator ribbon several meters wide, thinner than a piece of paper and about 100 000 km long. Climbers splicing new ribbon onto it at 200 km per hour would each need 21 days to travel its full length [10].

The concept is speculative, and the application treats it that way. The orbital tower had to overcome buckling, strength and dynamic stability, and its strength problem was answered on paper by an

exponential taper [7]. The strength of a real carbon nanotube cable, with the defects a cable of that length cannot avoid, has been estimated at least about 70 % below the theoretical strength of the nanotubes, enough to cast doubt on the construction [11], and powering the climbers and surviving micrometeoroids, lightning and debris are problems as important as producing the cable [8]. The simulator gives its ribbon a working specific strength of 40 MJ/kg, more than eleven times the 3.5 MJ/kg its parameter panel quotes for carbon fiber, and it takes the 6 MW power supply, the guideway and the anchor platform as declared inputs. It demonstrates numerical convergence and reduced-order plausibility. It does not show that an Earth space elevator can be built, and none of its numbers are those of a real one.

The plant corrects two beliefs that a transfer function leaves intact. The first is that a vehicle's weight is a constant that a controller can be tuned around once. A feedforward of the constant $g_0 = 9.806\,65\,\mathrm{m/s^2}$ is 0.43 % high at the anchor and 5.2 % high at 150 km, while 34 km past the geostationary altitude it believes in a 107.9 kN weight that is not there and throws the climber outward to 169.1 m/s. The second is that a structure too slow to matter to the drive can be left out of the command. The climber's local mode, its own mass on the ribbon below it, has a period of 8.9 min at 150 km, and a departure that reaches cruising speed in 4 min leaves the climber 30.8 m behind its moving lean and then 21.4 m past it. The same shape, a moving load on a long and lightly damped structure, appears in hoisting cables of changing length [1], in cranes and flexible manipulators [4] and in the elevator ribbon itself [12, 13]. However, dynamics studies of the elevator have treated the climber's motion as a prescribed transit [9] or as the decision variable of an optimization of the residual libration [13], while the motion a real climber can make is also bounded by its drive, its supply, the heat of its motor, its brakes and the air, and has to be flown by feedback on measurements.

This chapter measures what a gravity feedforward, a speed schedule, an input shaper and a feasibility supervisor are each worth on a speculative space elevator climber simulator. Each of these methods, and the elevator concept itself, is established. Section 12.2 reviews the concept, the dynamics of its ribbon and the history of shaped commands. Section 12.3 sets out the model with every relation and constant: a reduced-order ribbon of six finite-element modes, a residual-flexibility local mode that carries the climber's mass, and an energy ledger that closes to 3.8×10^{-10} of the energy through it on the scheduled descent. Section 12.4 examines the effective gravity along the route and approximates the climber's local mode as its own mass on the string below it, with the ribbon above acting as a dashpot of coefficient T/c. The approximation reproduces the application's periods to within 2.7 % up to 150 km and explains why the mode is lightly damped near the anchor and well damped high up. Section 12.5 gives the five layers of the controller, and Section 12.6 reports eight focused cases and a graded transfer, in which each control layer can be removed and its cost read off the scorecard. Section 12.7 sets the loops on the hardware a climber would carry, Section 12.8 states the limitations, Section 12.9 sums up, and the exercises of Section 12.10 run the cases again.

12.2 Elevator Concepts, Ribbon Dynamics and Shaped Motion

The idea of a structure reaching from the ground past the geostationary orbit is more than a century old, and it received serious scientific attention only after carbon nanotubes were discovered [14]. A letter of 1966 examined self-supporting cables extending into space from the Earth and from other rotating or revolving bodies, and observed that masses could in principle be launched with energy drawn entirely from the rotation of the central body [6]. The orbital tower of 1975 connected a geostationary satellite to the ground, named buckling, strength and dynamic stability as the three problems it had to overcome, and solved the strength problem with a cross-section tapered exponentially with the gravitational and inertial forces, largest at the geostationary point [7]. That taper is the constant-stress ribbon of Section 12.3.

The first studies concluded that no material then available had the required properties. After carbon nanotubes appeared in 1991, a design and deployment study returned to the construction and listed the powering of the climbers and the survival of micrometeoroid impacts, lightning strikes and

debris collisions among the problems to solve [8]. A study commissioned by NASA in the late 1990s examined the feasibility of such a structure and its applications [14]. Models of a real, and therefore defective, nanotube cable some 100 Mm long then placed its strength at least about 70 % below the theoretical nanotube strength, a reduction that puts the realization of the elevator in doubt [11].

> ⧖ **HISTORY** **A NASA Workshop Weighs the Elevator, 1999**
>
> A workshop held at a NASA flight center in 1999 asked whether the space elevator could be a possible future project instead of pure science fiction. Its findings were cautious. The materials the structure needed were still in laboratory development, and an elevator from the Earth to the geostationary orbit was judged not feasible at the time, though it might become possible in the latter part of the twenty-first century. A tether hanging down from the geostationary orbit to a tall tower built up from the ground appeared the most efficient and feasible form. The workshop's vehicle rode its track by electromagnetic propulsion, because traction wheels would be much slower and would wear both the vehicle and the structure [15]. The climber of this chapter takes the other path, rollers clamped on a ribbon.

Space tether research has its own review literature [16]. The elastic oscillations of the elevator ribbon have been modeled with assumed modes for its longitudinal extension and its transverse motion about a librating reference line [12]. A planar model of climber transit showed that the ribbon's primary response to an ascending climber is a westward rotation produced by the Coriolis force on the moving climber [9], and the climber's motion along the tether has been chosen by dynamic optimization to minimize the residual in-plane libration of the ribbon and the counterweight [13]. The industrial relatives are older. Deep mine hoisting cables have been modeled as a distributed-parameter system whose lateral and longitudinal oscillations interact through quadratic and cubic terms while the cable length changes [1].

> ⧖ HISTORY **A Tether Deployed to a Planned Profile, 1993 and 1994**
>
> Two small experiments, flown as secondary payloads in 1993 and 1994, each unreeled a tether of 20 km from a rocket stage. The first deployed it open loop. At the end of the deployment its tether was still running out at about 7 m/s, which the handbook of the field gives as the reason for a huge jump in tension and the rebounds that followed. The second flight closed the loop on a profile computed before the mission. Its control logic acted on a friction brake, speeding the deployment up or slowing it down to follow the profile and bring the payload to a smooth stop along the local vertical. The final tether rate was about 2 cm/s and the final libration about 4 degrees [17]. A planned trajectory with feedback around it is also the pairing of the climber's planner and speed loop.

Shaping a command against a structural mode began with Posicast control, which excites several transient oscillations at closely spaced times, with magnitudes and phases chosen so that their phasors sum to zero, and gives a half-cycle response when a step is split into two spaced excitations [2]. Preshaped commands that stay robust under parameter uncertainty cost a move-time penalty of about one period of the first mode and cut the endpoint residual vibration of a Space Shuttle Remote Manipulator System simulator by a factor of 25 [3]. The first fifty years of command shaping for flexible systems have been reviewed [4], and Volume 7, Chapter 12 applies the same shapers to the swinging load of a gantry crane, with timings taken from an assumed cable length. In high-performance motion systems the command is itself a point-to-point trajectory with bounds on its derivatives, time-optimal in the most relevant cases, fed forward through a controller of matching order [5]. The simulator combines all three, a bounded-jerk trajectory, a shaper and a model feedforward, on a plant whose load and whose mode both change with height.

12.3 Climber, Ribbon and Loads

The simulated system is a climber of mass m on a ribbon anchored at the equator, described in the frame that turns with the Earth. Heights $z = r - R_E$ are measured outward from the anchor and lateral displacements eastward in the orbit plane. The route, the climber, its forces, its local link to the ribbon and its instruments are shown in Figure 12.1. Thirty states are integrated together, namely the height, the speed and the developed drive force, the motor and brake temperatures, eight modal coordinates and their rates of which six are used, the climber's lateral position and velocity, and seven energy ledgers. Table 12.1 lists the variables by role.

12.3.1 Effective Gravity and the Declared Ribbon

A point held on the ribbon at radius r feels gravity inward and the effect of turning with the Earth outward. Their difference is the effective gravity of Equation (12.1), the radial derivative of the effective potential Φ.

$$g_{eff}(r) = \frac{\mu}{r^2} - \omega^2 r = \frac{\mathrm{d}\Phi}{\mathrm{d}r}, \qquad \Phi(r) = -\frac{\mu}{r} - \frac{\omega^2 r^2}{2} \tag{12.1}$$

Here μ is the Earth's gravitational parameter and ω its sidereal rotation rate. The effective gravity vanishes at the geostationary radius $r_{GEO} = (\mu/\omega^2)^{1/3} = 42\,164.17\,\mathrm{km}$, which is 35 786 km above the anchor, and it points outward beyond that radius, reaching $-0.492\,\mathrm{m/s^2}$ at the 100 000 km counterweight.

The ribbon is a hypothetical constant-stress design. Every cross-section works at the same stress σ_w, so its linear density follows its tension, $\lambda = T/c^2$ with $c^2 = \sigma_w/\rho_r$ the working specific strength and ρ_r the material density, and the static balance $\mathrm{d}T/\mathrm{d}r = \lambda g_{eff}$ integrates to Equation (12.2).

$$T(r) = T_0\, \mathrm{e}^{[\Phi(r) - \Phi(R_E)]/c^2}, \qquad M_{cw} = \frac{T(r_{cw})}{\omega^2 r_{cw} - \mu/r_{cw}^2} \tag{12.2}$$

where T_0 is the anchor tension with no climber on the ribbon and M_{cw} is the counterweight that the tension at r_{cw} holds in place. With $c^2 = 40\,\mathrm{MJ/kg}$ and $T_0 = 320\,\mathrm{kN}$ the tension grows to 1073.7 kN

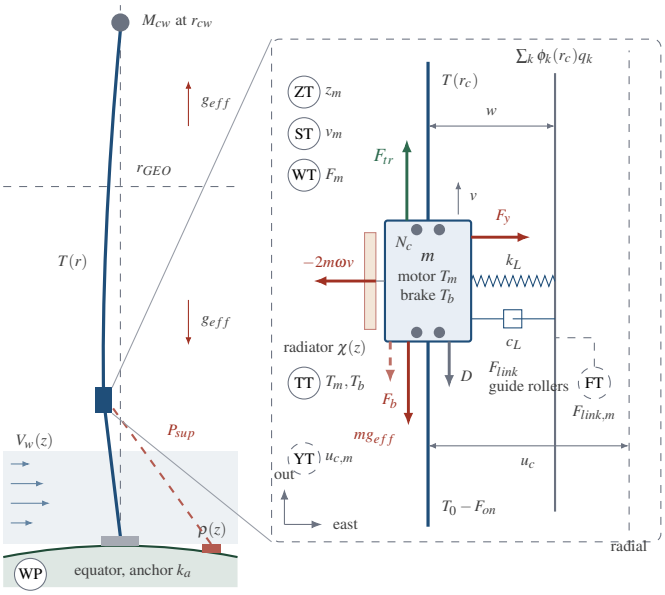


Figure 12.1 Space elevator as modeled, not to scale. Left, the route: the anchor of lateral stiffness k_a, the atmosphere $\rho(z)$ with the crosswind profile $V_w(z)$, the ribbon of tension $T(r)$ leaning west behind a climbing climber, the geostationary radius where g_{eff} changes sign, the counterweight M_{cw} at r_{cw}, the hypothetical power delivery P_{sup} and the wind profiler WP. Right, the climber of mass m on traction rollers clamped with force N_c, with its drive force F_{tr}, brake F_b, drag D, weight mg_{eff}, Coriolis force $-2m\omega v$ and side load F_y, and the local link k_L, c_L that carries F_{link} to the ribbon's modal shape. Solid instruments are read by the controller and dashed ones are only logged.

at the geostationary radius and falls to 712.7 kN at the counterweight, the density rises from 0.0080 to 0.0268 kg/m, and the ribbon weighs 2146.9 t with a 1448.9 t counterweight. The same c, 6325 m/s, is the lateral wave speed everywhere, since $\sqrt{T/\lambda} = c$. These dimensions follow from the declared inputs and are not the result of a design study.

Table 12.1 Climber Variables by Role, With Defaults and Ranges

Symbol	Role, unit	Default	Range
v	CV, speed loop, m/s	$v_{sh}+\Delta v$	$\le v_{lim}(z)$
z	CV, station stop, km	SP 150	0–40 000
F	MV, force demand, kN	MANUAL 0	−400 to 280
$F_{d,cmd}$	allocated motor command, kN	—	±280
$F_{b,cmd}$	allocated brake command, kN	—	0 to $F_{b,cap}$
Δv	MV of the height trim, m/s	0	±2
V_j	DV, crosswind at z_j, m/s	0	±30
σ_g	DV, gust intensity, m/s	0	0–8
p_s	DV, supply asked, %	100	0–100
b_v	DV, tachometer bias, m/s	0	±2
Δm	DV, cargo off the sheet, t	0	−5 to 8
F_{link}	limit, guide rollers, kN	—	≤ 6.00
T_a	limit, anchor tension, kN	320	≥ 64
T_m, T_b	limits, motor, brake, °C	20	≤ 150, ≤ 600
z, v, F_d	states, height, speed, drive	per case	—
T_m, T_b	states, temperatures, K	293.15	—
q_k, $\dot{q}_k$	states, 8 slots, 6 used, m	quasi-static	—
u_c, $\dot{u}_c$	states, climber lateral, m	quasi-static	—
$E_{in}\ldots E_{dmp}$	states, 7 ledgers, J	0	—

12.3.2 The Climber Along the Ribbon

The climber obeys the radial momentum balance of Equation (12.3), with the drive developing its force through a first-order lag.

$$\begin{aligned} m\frac{\mathrm{d}v}{\mathrm{d}t} &= F_{tr}+F_b-m\,g_{eff}(r)-D, \qquad \frac{\mathrm{d}z}{\mathrm{d}t}=v \\ \tau_F\frac{\mathrm{d}F_d}{\mathrm{d}t} &= \mathrm{clamp}(F_{d,cmd},\pm F_I)-F_d \\ D &= \tfrac{1}{2}\rho(z)C_DA_f\,|v|\,v \end{aligned} \tag{12.3}$$

Here F_d is the force the motor has developed, F_I its force at the current limit, C_DA_f the axial drag area and $\rho(z)=\rho_0\mathrm{e}^{-z/H}$ below 200 km and zero above. The traction force F_{tr} is F_d after the three limits of Equation (12.4).

$$\begin{aligned} |F_{tr}| &\le F_P=\frac{-|v|+\sqrt{v^2+4c_{cu}(P_{av}-P_{aux})}}{2c_{cu}}, \quad F_dv\ge 0 \\ |F_{tr}| &\le F_R=\frac{|v|-\sqrt{v^2-4c_{cu}(P_{aux}+P_{rg})}}{2c_{cu}}, \quad F_dv<0 \\ P_{el} &= F_{tr}\,v+c_{cu}F_{tr}^2+P_{aux} \end{aligned} \tag{12.4}$$

The first bound keeps the electrical power P_{el} within the power P_{av} the supply delivers, and the second keeps the power returned by regeneration within the P_{rg} the supply accepts, with no bound where the discriminant is not positive because the losses then absorb all of it. The third limit is the grip of the rollers. Above the static traction $\mu_sN_c=225$ kN they slide and deliver $\mu_kN_c=175$ kN. The available power P_{av} follows the power asked, $P_{sup}\,p_s/100$, at a ramp rate of 25 % of the 6 MW supply per minute.

The friction brake always opposes the motion and fades when hot, as shown in Equation (12.5).

$$\begin{aligned} F_b &= -\min\left[F_{b,cmd},\, F_{b0} f(T_b)\right] \frac{v}{\sqrt{v^2+v_\varepsilon^2}} \\ f &= \max\left[0.4,\; 1-0.6\,(T_b-T_{fade})/200\,\mathrm{K}\right] \end{aligned} \tag{12.5}$$

with $f=1$ below $T_{fade}=650\,^\circ\mathrm{C}$ and $v_\varepsilon=0.05\,\mathrm{m/s}$ smoothing the direction through zero speed. Two contacts hold the climber at rest as hybrid states. On the platform, $z\le 0$, the climber is held while $F_{tr}+F_b-mg_{eff}\le 0$ and $v\le 0$, and a landing faster than 2 m/s is recorded as an impact. The parking clamp engages on command once $|v|\le 0.05\,\mathrm{m/s}$, carries the climber's weight and slips if the net force $|F_{tr}-mg_{eff}|$ exceeds the brake capacity.

12.3.3 The Ribbon and the Climber Sideways

A small eastward displacement $u(r,t)$ of the ribbon obeys the taut-string equation with the tidal term of the rotating frame, Equation (12.6), fixed at the anchor and carrying the counterweight at its upper end.

$$\lambda\, u_{tt} = (T u_r)_r - \lambda\, \kappa(r)\, u + f, \qquad \kappa(r) = \frac{\mu}{r^3} - \omega^2 \tag{12.6}$$

The tidal stiffness κ restores below the geostationary radius and destabilizes above it. Eight hundred linear finite elements stretched toward the anchor, $r_i = R_E + L(\mathrm{e}^{\beta i/N}-1)/(\mathrm{e}^{\beta}-1)$ with $\beta=9$, lumped nodal masses $\lambda_i \ell_i$, element stiffnesses $T(r_{i+1/2})/h_i$, the tidal term $\lambda_i \kappa(r_i)\ell_i$ on the diagonal and the counterweight's mass and tidal stiffness at the top node give a symmetric tridiagonal eigenproblem whose six lowest modes are found by Sturm bisection and inverse iteration. Each mode shape ϕ_k is scaled to a unit peak, so the modal coordinate q_k is that mode's largest deflection in meters. The periods are 127.40, 7.929, 4.061, 2.722, 2.047 and 1.640 h, with modal masses from 2163 t down to 929 t. The climber drives each mode through its guide-roller force in proportion to the mode's shape at its height, as shown in Equation (12.7).

$$M_k\left(\ddot{q}_k + 2\zeta_k \Omega_k \dot{q}_k + \Omega_k^2 q_k\right) = \phi_k(r_c)\, F_{link} \tag{12.7}$$

Six string modes cannot represent the give of the ribbon between the anchor and a climber low on it, so that residual flexibility carries its own degree of freedom, the climber's lateral position u_c, in Equation (12.8).

$$\begin{aligned} m\ddot{u}_c &= -2m\omega v + F_y - F_{link} - m\,\kappa(r_c)\, u_c \\ F_{link} &= k_L w + c_L \dot{w}, \qquad w = u_c - \textstyle\sum_k \phi_k(r_c)\, q_k \end{aligned} \tag{12.8}$$

where $\dot{w}$ includes the rate $\phi_k'(r_c)\, v\, q_k$ at which the climber moves along the modal shapes. The Coriolis force $-2m\omega v$ pushes a climbing climber west and a descending one east, and it vanishes while the climber is held. It is the rotating-frame term that Chapter 11 cancels with a feedforward on its thrusters. The climber has no lateral actuator, so only the timing of its speed changes acts on this push. The flexibility and damping of the link follow Equation (12.9).

$$\begin{aligned} \frac{1}{k_L} &= G(r_c) - \sum_k \frac{\phi_k^2(r_c)}{M_k \Omega_k^2} + \frac{1}{k_a} + \frac{\Lambda^2 F_{on}}{\max(0.02z,\; z-\Lambda F_{on})} \\ c_L &= 2\zeta_s \sqrt{k_L m} + \frac{T(r_c)}{c} \end{aligned} \tag{12.9}$$

Here $\Lambda(r_c)=\int_{R_E}^{r_c} \mathrm{d}r'/T(r')$ is the series flexibility of the ribbon below the climber, and $G(r)$ is the static deflection at r under a unit lateral force there, the diagonal of the inverse stiffness, from which the share of the six modes is removed. The last term is the extra give of the ribbon below the climber while the force F_{on} the climber rides on takes tension off it, which for a uniform string is exactly the change from z/T to $z/(T-F_{on})$. The damping adds the impedance T/c of the ribbon above, which carries away the waves the climber launches upward. The sway reported by the application is the local motion beyond the quasi-static deflection under the loads of the moment, $s = w - (-2m\omega v + F_y)/k_L$.

12.3.4 Air, Wind and Side Load

The crosswind has a single-peaked profile with its maximum at $z_j = 10\,\text{km}$ and first-order gusts, and it loads the climber only, since the ribbon is taken as edge-on to the wind. The side force has a normal-force part that grows with the product of the climbing speed and the wind, the $q\alpha$ load of an ascending vehicle, and a crossflow drag, as shown in Equation (12.10).

$$\begin{aligned} V_w &= \psi(z)\,(V_j + \gamma_k), \qquad \psi(z) = \frac{z}{z_j}\,\mathrm{e}^{1-z/z_j} \\ \gamma_{k+1} &= \alpha_g\,\gamma_k + \sqrt{1-\alpha_g^2}\,\sigma_g n_k \\ F_y &= \tfrac{1}{2}\,\rho(z)\,V_w\big(C_{N\alpha}A_f|v| + C_c A_s|V_w|\big) \end{aligned} \tag{12.10}$$

Here $\alpha_g = \mathrm{e}^{-\Delta t_c/\tau_g}$ with $\tau_g = 20\,\text{s}$, n_k is a standard normal deviate drawn every control period and the profile ψ is zero above $12 z_j$. At the height of the jet the air density is $0.378\,\text{kg/m}^3$.

12.3.5 Heat

The motor and the brake are lumped thermal masses, given in Equation (12.11).

$$\begin{aligned} C_m\frac{\mathrm{d}T_m}{\mathrm{d}t} &= c_{cu}F_{tr}^2 - hA_m(T_m - T_{air}) \\ &\quad - \chi(z)\,\varepsilon\sigma A_m\big(T_m^4 - T_{sink}^4\big) \\ C_b\frac{\mathrm{d}T_b}{\mathrm{d}t} &= |F_b\,v| - hA_b(T_b - T_{air}) \\ &\quad - \varepsilon\sigma A_b\big(T_b^4 - T_{sink}^4\big) \end{aligned} \tag{12.11}$$

Here σ is the Stefan-Boltzmann constant. The forced convection of a cooler rated hA_{50} at 50 m/s at sea level scales as $hA_{50}(\rho/\rho_0)^{0.8}(\max(|v|, 5\,\text{m/s})/50\,\text{m/s})^{0.8}$, the fans supplying the 5 m/s at rest. The motor's radiator is stowed in the air and opens over 2 km around $z_{rad} = 40\,\text{km}$, $\chi = \text{clamp}[(z - z_{rad})/2\,\text{km} + 0.5,\ 0,\ 1]$, while the brake radiates at every height. The air cools at 6.5 K/km to 216.65 K at 11 km and stays there. The copper loss depends on the force and not on the speed, which is what makes a slow climb through the thin air hot.

12.3.6 Tension, Energy Ledger and Instruments

The force the climber rides on comes off the ribbon below it, so the anchor tension is $T_a = T_0 - F_{on}$, with $F_{on} = F_{tr} + F_b$ in motion, mg_{eff} on the clamp and F_{tr} on the platform. Seven ledgers integrated with the states give the energy balance of Equation (12.12).

$$E_{in} - E_{out} - E_{cu} - E_{aux} = \Delta E_{mech} + E_{bk} + E_{dr} + E_{ct} \tag{12.12}$$

The left side is the electrical energy supplied, less the energy returned, the copper loss and the auxiliary load, and the right side is the change of mechanical energy $E_{mech} = \frac{1}{2}mv^2 + m[\Phi(r) - \Phi(R_E)]$ plus the brake heat, the drag work and the kinetic energy absorbed by the clamp and the platform. A further ledger E_{dmp} accumulates what the lateral damping removes.

No loop reads the true state. The controller sees the height by ranging with 0.5 m of noise, the speed through a tachometer with 0.02 m/s of noise plus the bias b_v, the traction force with 0.2 kN of noise, the motor temperature after a 20 s first-order lag, with 0.2 K noise, and the brake temperature after a 5 s lag, with 1.0 K noise, each sampled once a second. A lateral fix of u_c (3 s lag, 1.0 m), an anemometer (2 s, 0.3 m/s), the anchor load cell (2 s, 0.5 kN) and a guide-roller load cell (30 N) are logged for the trends and the export, and no decision reads them. The speed schedule instead takes the mean crosswind V_j and the gust intensity σ_g as a wind profiler would report them.

Table 12.2 Parameters of the Climber and the Ribbon

Symbol	Meaning	Value
adjustable in the model dialog		
c^2	working specific strength	40 MJ/kg
T_0, r_{cw}	anchor tension, counterweight	320 kN, 100 000 km
N_m	lateral modes kept	6
ζ_k, ζ_s	modal, guideway damping	1 %, 1 %
k_a, $T_{a,min}$	anchor stiffness, tension floor	5000 N/m, 64 kN
m_e	climber empty mass	5.0 t
A_f, C_D	frontal area, drag coefficient	12 m^2, 0.80
$C_{N\alpha}$, C_cA_s	normal-force slope, crossflow	2.0 rad^{-1}, 36 m^2
$F_{g,max}$	guide-roller rating	6.00 kN
F_I, τ_F	current-limit force, drive lag	280 kN, 2.0 s
μ_s, μ_k, N_c	traction, sliding, clamp force	0.45, 0.35, 500 kN
c_{cu}, P_{aux}	copper loss, auxiliary load	12 W/kN^2, 40 kW
P_{sup}, $\dot{P}_{sup}$	supply, ramp rate	6.0 MW, 25 %/min
P_{rg}, v_{max}	regeneration, roller speed	2.0 MW, 100 m/s
F_{b0}, T_{fade}	brake capacity, fade onset	250 kN, 650 °C
C_m, hA_m	motor capacity, air cooling	2.0 MJ/K, 10 kW/K
A_m, z_{rad}	radiator area, opens above	260 m^2, 40 km
$T_{m,max}$, $T_{b,max}$	winding and brake limits	150, 600 °C
C_b, hA_b	brake capacity, air cooling	0.80 MJ/K, 4.0 kW/K
A_b	brake radiating area	15 m^2
ε, T_{sink}	emissivity, radiative sink	0.85, 250 K
ρ_0, H	sea-level density, scale height	1.225 kg/m^3, 8.5 km
z_j, τ_g	jet height, gust correlation	10 km, 20 s
fixed in the code		
μ	gravitational parameter	$3.986\,004\,418 \times 10^{14}$ m^3/s^2
ω	sidereal rotation rate	$7.292\,115\,9 \times 10^{-5}$ rad/s
R_E	equatorial radius	6378.137 km
g_0	constant-gravity feedforward	9.806 65 m/s^2
N, β	ribbon elements, stretching	800, 9
v_ε	brake direction smoothing	0.05 m/s
Δt, Δt_c	integration, control period	0.5 s, 1 s

12.3.7 Numerics and Starting State

Table 12.2 gives every parameter. All 42 model parameters can be changed in the application's model dialog, and the constants of the code are listed apart from them. Classical fourth-order Runge-Kutta integrates the thirty states on a 0.5 s step of the shared engine of Section 1.6.1. Its clock runs in minutes here, so the derivatives are written per second and scaled by 60 s/min. The controller and the instruments run every second, the historian logs six rows per simulated minute, and the speed selector spans 0.05 to 120 simulated minutes per wall-clock second. The contact and limit logic runs after each integration step. Every case starts on the quasi-static lateral shape for its Coriolis and mean-wind load F_{ext}, $q_k = \phi_k F_{ext}/(M_k\Omega_k^2)$ and $u_c = \sum_k \phi_k q_k + F_{ext}/k_L$, with the drive holding $mg_{eff} + D$ except on the platform or the clamp, and both temperatures start at 20 °C.

12.3.8 Assumptions

The ribbon's material, the power delivery, the guideway integrated with the ribbon, the anchor platform and the deployment are hypothetical. The model applies standard physical relations to declared inputs, and its plausibility is that of the relations, not of the inputs.

The ribbon is axially rigid, so a change of traction force reaches the anchor at once and the travel of the axial wave is left out. The lateral motion stays in the orbit plane, without out-of-plane librations and without the Moon, the Sun or the Earth's oblateness.

The coupling is one-way. The lateral motion does not feed back on the radial motion, and the radial Coriolis term it would add, $2m\omega\dot{u}_c$, stays below 8.0×10^{-5} of the weight in every case and variant below 1000 km and below 0.97 N in case 5.

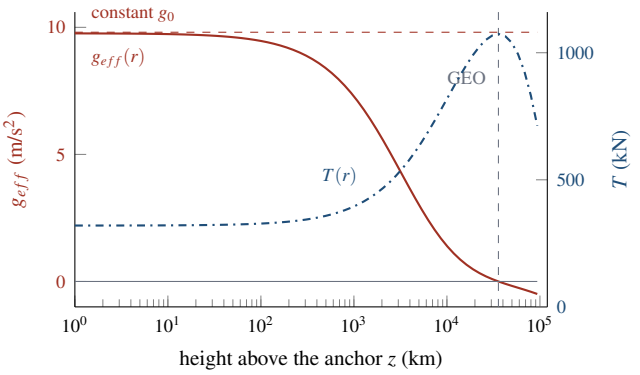


Figure 12.2 Effective gravity (thick solid line, left axis) and ribbon tension (thick dash-dotted line, right axis) along the route at the default inputs. The thin dashed horizontal line is the constant g_0 a naive feedforward would assume, and the dashed vertical line marks the geostationary altitude, 35 786 km, where T peaks at 1073.7 kN and beyond which g_{eff} points outward. The curves end at the counterweight, 93 622 km above the anchor.

Six string modes and one local degree of freedom represent the lateral motion. The local part is drawn with a static shape, and the impedance T/c stands in for the waves the climber launches, which holds while they take longer to come back than the local period.

A case that starts in motion starts on the quasi-static shape for its load. For the 127.4 h mode that is an idealization, since a real ribbon's state would depend on the history of every climber on it.

The motor and the brake are single thermal masses with illustrative coefficients, a forced-convection law, a radiator that opens above 40 km and a brake that fades linearly above 650 °C, and roller slip is a sliding coefficient with no slip dynamics.

The atmosphere is exponential with an 8.5 km scale height, the crosswind has one peak and first-order gusts and acts on the climber only, and the aerodynamic coefficients are not those of any measured vehicle.

12.4 Gravity That Fades and a Mode That Moves

12.4.1 The Weight Along the Route

The effective gravity and the ribbon tension along the whole route are plotted in Figure 12.2. Below 100 km the effective gravity stays within 3.1 % of its anchor value, which is why terrestrial hoists can treat gravity as a constant. It falls to 9.318 m/s^2 at 150 km, 7.283 m/s^2 at 1000 km and 1.399 m/s^2 at 10 000 km, a seventh of g_0, and it changes sign at the geostationary radius. The tension peaks where the effective gravity changes sign, since $\mathrm{d}T/\mathrm{d}r = \lambda g_{eff}$. A climber's weight therefore spans the whole range from the full terrestrial value to a small outward pull, and a feedforward that is right at one end of the route is wrong at the other unless it models the physics.

12.4.2 The Local Mode

Low on the ribbon the retained modes carry little of the flexibility, 0.4 % of G at 10 km and 5.9 % at 150 km, so Equation (12.9) is dominated by the string below the climber. Treating that string as uniform, at the anchor tension less the climber's weight, gives the period and damping ratio of Equation (12.13).

$$T_L \approx 2\pi\sqrt{m\Big(\frac{z}{T_0 - m g_{eff}} + \frac{1}{k_a}\Big)}, \qquad \zeta_L \approx \zeta_s + \frac{T(r_c)/c}{2\sqrt{k_L m}} \tag{12.13}$$

Table 12.3 Local Mode of the Climber From the Application and From Equation (12.13)

z (km)	m (t)	T_L app (min)	T_L Eq. (min)	ζ_L app	ζ_L Eq.	modal share
10	11	2.381	2.385	0.062	0.063	0.4 %
150	5	5.341	5.486	0.277	0.285	5.9 %
150	11	8.888	9.122	0.212	0.218	5.9 %
150	14	10.738	11.024	0.202	0.207	5.9 %
400	11	13.688	14.646	0.340	0.363	14.8 %
1000	11	19.099	22.425	0.528	0.618	32.1 %

The period grows with the square root of the height and of the mass, and a heavier climber is slower twice over because its own weight also unloads the string it rests on. The damping comes almost entirely from the ribbon above, which absorbs the local motion as a dashpot of coefficient T/c, the characteristic impedance $\sqrt{T\lambda}$ of a string. The impedance changes little with height while the stiffness falls, so the local mode is lightly damped near the anchor and well damped high up. Equation (12.13) is compared with the application's own local mode in Table 12.3. It agrees to within 2.7 % up to 150 km and overestimates the period by 7.0 % at 400 km and 17.4 % at 1000 km, where the retained modes carry 14.8 % and 32.1 % of the static flexibility that the approximation leaves in the local mode.

🕮 **Pause.** Try Exercise 12.3 before reading on.

12.4.3 The Coriolis Load as a Step

A climber leaving a station at speed v adds the lateral load $2m\omega v$, which the local mode takes up as a quasi-static lean of $2m\omega v/k_L$ to the west. The climber's own move is therefore a disturbance that enters at a known time with a known size, and it can be met only through the command. At 150 km the 11 t climber has $k_L = 1.527\,\mathrm{N/m}$, so the 86.6 N of its case 3 cruise at 54.0 m/s would lean it 56.7 m, and the lean grows as the climber rises and the link softens. The acceleration to that speed takes 3.99 min, less than half the 8.9 min period, and to the mode it is close to a step. A $0.3\,\mathrm{m/s^2}$ limit and a $5\,\mathrm{mm/s^3}$ jerk limit set that duration, which is why the planner alone cannot make the start gentle without making it long. The shapers of Section 12.5 instead split the speed change into parts timed against the mode itself.

12.5 Five Layers, One Scan per Second

The controller runs every second in five layers, outermost first, and all of them act on the measurements of Section 12.3. One scan advances the supply ramp and the gust, draws the measurements, re-reads the load sheet while the climber is parked or docked, forms the speed limit and the shaper timing, updates the supervisor, steps the planner and the shaper, computes the speed loop and allocates its demand. Table 12.4 gives the defaults, the ranges and the fixed margins.

12.5.1 The Speed Schedule

With the fixed schedule the speed limit is $v_{lim} = \min(V_{cr}, v_{max})$. With the scheduled one it is also capped at the three speeds of Equation (12.14).

$$
\begin{aligned}
v_P &: \ \left[\hat{m}(g_{eff} + 0.02) + \hat{D}\right] v + c_{cu} F^2 + P_{aux} = 0.95\, P_{av} \\
v_G &= \frac{0.8\, F_{g,max} / (\tfrac{1}{2}\rho \hat{V}) - C_c A_s \hat{V}}{C_{N\alpha} A_f} \\
\hat{V} &= \psi(z)\left(|V_j| + 2.5\sigma_g\right) \\
v_B &: \ \left(F_{hold} - F_{rg}\right) v = \dot{Q}_b \left(T_{b,max} - 80\,\mathrm{K}\right)
\end{aligned}
\tag{12.14}
$$

Table 12.4 Controller Settings, Their Defaults and Their Ranges

Setting	Default	Range
cruise speed V_{cr}	60 m/s	5–100
speed-loop gain $K_{c,v}$	0.10 s^{-1}	0.01–0.50
integral time $\tau_{I,v}$	40 s	5–300
height-trim gain K_x	0.020 s^{-1}	0–0.10
acceleration, jerk a_{max}, j_{max}	0.30 m/s^2, 5 mm/s^3	0.05–1, 1–100
shaper period T_s/T_L	100 %	50–150
cargo on the load sheet	6 t	0–20
gravity feedforward $\hat{g}$	$g_{eff}(r)$	g_0, off
speed schedule	scheduled	fixed
input shaper	off	ZV, ZVD
regeneration, thermal check	on, on	off
traction, guide, power margins	0.9, 0.8, 0.95	fixed
tension margin, gust allowance	8 kN, $2.5\sigma_g$	fixed
trim authority, reference lead	±2 m/s, 20 m	fixed
brake handover, shaper update	10 kN/s, 3 %	fixed

The power speed v_P, with F the bracketed force, applies on the way up. The guide speed v_G is the lowest over seven heights spread across the look-ahead $v^2/(2a_{max}) + v(a_{max}/j_{max} + t_n) + 200\,\text{m}$, which covers braking and the shaper's delay t_n, and it is never below 3 m/s. Here $\psi(z)$ is the wind profile of Equation (12.10). The brake speed v_B applies on the way down, with $F_{hold} = \hat{m} g_{eff} - \hat{D}$, F_{rg} the regeneration available at that speed and $\dot{Q}_b$ the heat the friction brake sheds steadily at 520 °C, taken as the lowest over five heights below the climber. The speeds v_P and v_B are found by bisection between 0.5 m/s and v_{max}.

12.5.2 The Planner and the Shaper

The planner is a time-optimal jerk-limited generator, derived from the one the air taxi of Volume 20, Chapter 3 and the lift-plus-cruise aircraft of Volume 20, Chapter 4 fly with, that keeps $|\dddot{z}_r| \le j_{max}$, $|\ddot{z}_r| \le a_{max}$ and $|\dot{z}_r| \le v_{lim}$, starts braking when the distance left equals the jerk-limited stopping distance, which from speed v at zero acceleration is $v^2/(2a_{max}) + v\,a_{max}/(2j_{max})$ when $v \ge a_{max}^2/j_{max}$ and $v\sqrt{v/j_{max}}$ otherwise, and advances in ten substeps per scan. While it accelerates it looks one substep ahead under the steepest jerk and brakes at once if waiting would carry it more than 0.5 m past its braking point, and a reference moving away from its target with more than 1 m of stopping distance is first halted and then starts a fresh move from rest. Within 0.05 m, 0.05 m/s and $0.1a_{max}$ of the target it hands over to a linear capture law, $\dddot{z}_r = 3\nu[\nu(\nu e/3 - \dot{z}_r) - \ddot{z}_r]$ with $\nu = (4j_{max})^{1/3}$ in SI units, clamped to the jerk and acceleration limits. The shaper convolves this reference with the impulse trains of Equation (12.15).

$$\begin{aligned} \text{ZV: } A &= \frac{[1,\ K]}{1+K}, \quad t = \left[0,\ \frac{\pi}{\omega_d}\right] \\ \text{ZVD: } A &= \frac{[1,\ 2K,\ K^2]}{(1+K)^2}, \quad t = \left[0,\ \frac{\pi}{\omega_d},\ \frac{2\pi}{\omega_d}\right] \\ z_{sh}(t) &= \textstyle\sum_i A_i\, z_r(t - t_i), \qquad K = \mathrm{e}^{-\zeta\pi/\sqrt{1-\zeta^2}} \end{aligned} \tag{12.15}$$

The speed and acceleration references are convolved the same way. The mode (ω_d, ζ) is predicted from the load sheet and the ribbon model at the measured height, Equation (12.9) with the load-sheet mass $\hat{m}$ and $F_{on} = \hat{m} g_{eff}(z_m)$, and its period is multiplied by the adjustable ratio T_s/T_L. For the 11 t climber at 150 km the prediction is 8.888 min with $\zeta = 0.212$, which gives ZV impulses of 0.664 and 0.336 at 0 and 4.548 min and ZVD impulses of 0.441, 0.446 and 0.113 at 0, 4.548 and 9.095 min.

The heavy radiation damping makes the second ZV impulse half the first. The impulses are rebuilt only when the predicted period moves by more than 3 % and not while the planner is braking, and each rebuild moves the stored unshaped trajectory by the amount that keeps z_{sh} continuous, within a 7200 s history. Once the unshaped reference has reached the station the freeze lifts, the tail of the shaped reference is re-timed to the period predicted near the station, and the continuity shift sends the unshaped reference on a short move back to it. While the demand is saturated in the direction of travel, a reference governor keeps the shaped height within 20 m of the climber and the reference speed within 0.2 m/s of it.

12.5.3 The Speed Loop and the Height Trim

The speed loop is the framework's shared PID in velocity form of Section 1.6.2, working in acceleration units, with its derivative term set to zero, as shown in Equation (12.16).

$$\begin{aligned} v_{sp,k} &= v_{sh,k} + \mathrm{clamp}\big(K_x(z_{sh,k} - z_{m,k}),\ \pm 2\,\mathrm{m/s}\big) \\ e_k &= v_{sp,k} - v_{m,k}, \qquad P_k = K_{c,v} e_k \\ S_k &= S_{k-1} + P_k - P_{k-1} \\ b_k &= \hat{g}(z_{m,k}) + a_{sh,k} + \hat{D}_k/\hat{m} \\ c_k &= b_k + S_k + I_{k-1} \\ u_k &= \mathrm{clamp}\big(c_k,\ F_{lo}/\hat{m},\ F_{hi}/\hat{m}\big), \quad F_k = \hat{m}\,u_k \\ I_k &= I_{k-1} + \Big[\frac{K_{c,v}}{\tau_{I,v}}\,e_k + \frac{u_k - c_k}{\tau_{I,v}}\Big]\Delta t_c \end{aligned} \tag{12.16}$$

The bias b_k is the feedforward, made of the assumed gravity $\hat{g}$ (g_{eff}, g_0 or zero), the shaped acceleration and the drag model divided by the load-sheet mass. Its gravity term is the static feedforward of a known force in the sense of Section 1.4.3: it needs only the load sheet and the gravity model, and the integral is left to correct their errors. The other symbols map onto Equations (1.10) to (1.12) as $b_k = u_b$, $c_k = v_k$ and $T_t = \tau_{I,v}$, with back-calculation on the integral only and the clamp on the emitted demand. Nothing limits the rate of the demand beyond the 2 s drive lag. A climber slower than its reference, a positive error, raises the force demand. The limits are $F_{hi} = \min(F_I,\ 0.9\mu_s N_c,\ T_0 - T_{a,min} - 8\,\mathrm{kN})$, reduced to $F_P(v_m)$ while climbing, and $F_{lo} = -[\min(F_I,\ 0.9\mu_s N_c) + F_{b0} f(T_{b,m})]$ while moving, raised to $-F_P(v_m)$ while descending. In MANUAL the operator's demand is applied, the loop tracks it with $I_k = F_{man}/\hat{m} - b_k - S_k$ as in Section 1.6.2, and the return to automatic starts the planner from the measured height and speed without a jump.

12.5.4 Braking Allocation

The force demand goes to the motor whenever it drives the motion or the climber is stopped. When it opposes the motion, regeneration takes what the supply accepts, $\min(|F|, F_R(v_m), F_I)$, the motor hands any larger share to the friction brake at no more than 10 kN/s, and the brake fills in what the motor will deliver short of the demand over the next scan, predicted from the measured drive force through the 2 s lag. In automatic the motor command stays within $\pm\min(F_I,\ 0.9\mu_s N_c)$.

12.5.5 The Supervisor

The supervisor moves the climber through the phases parked, release, go and docked, and through stopping and hold when a move becomes infeasible. Its checks run at once and then every simulated minute. A load is feasible while the load-sheet weight $\hat{m} g_{eff}(z_m)$ stays within $T_0 - T_{a,min} - 8\,\mathrm{kN}$, for the tension, and within $0.9\mu_s N_c$, for the grip, and a refusal names the first limit that fails, in the order tension, grip and heat. With the thermal check on, the motor temperature is marched along the rest of the climb in 30 s steps at the power-limited speed with the thermal model of Equation (12.11), and the climb is infeasible when the forecast peak passes the winding limit less 3 K while moving, or less 10 K before a start. A release ramps the drive to the believed weight $\hat{m}\hat{g}$ over 30 s and opens the

Table 12.5 Case Sequence Along the Route

#	Case, segment and change	Teaching point
1	5 t, anchor to 150 km	a weight that fades
2	15 t, same climb	the limit that binds
3	11 t, 150 to 400 km at 5 min	Coriolis sway on departure
4	11 t to 60 km, 15 m/s jet, gusts	slow down in the wind
5	11 t, 100 m/s across GEO	weightless, then outward
6	11 t, supply 20 % at 6 min	forecast and hold
7	stop at 150 km, sheet 3 t light	shaping a stop
8	11 t, 100 km down, 40 m/s	braking capacity
9	13 t to 150 km in the wind	graded transfer

clamp only once the measured drive force has reached 99 % of it. An infeasible climb stops from where the climber is at the jerk and acceleration limits, clamps, and holds for at least 60 s and until the forecast clears. The climber docks when the shaped reference is within 0.5 m and 0.02 m/s of the station, the planner has captured it and the climber is within 1 m and 0.05 m/s, or when it rests on the platform with the anchor as its destination. Falling toward the anchor faster than the planned stop allows, the planner brakes at up to 3 m/s^2 to protect the platform.

12.5.6 Three Models Kept Apart

Section 12.3 describes the simulation model. No identification runs inside the application, yet the loop's control model follows from Equation (12.3). With an exact feedforward the demand u reaches the speed as an integrator behind the drive lag, $v/u = 1/[s(\tau_F s + 1)]$, and the default $K_{c,v} = 0.1\,\mathrm{s}^{-1}$ puts the crossover near 0.1 rad/s, much faster than the 0.0118 rad/s of an 8.9 min local mode, so the loop tracks the reference that excites the mode rather than damping the mode. The controller's internal models are the load-sheet mass, the assumed gravity, the drag model, the ribbon model at the load-sheet mass for the shaper and the motor model for the forecast. Only the load sheet can be made wrong by a disturbance.

12.6 Cases and What They Measure

Table 12.5 lists the cases and Table 12.6 what they showed. Each case starts from the defaults of Table 12.4 apart from the changes named, and every number below comes from the engine run without its interface at that case's own seed, with times in simulated minutes from the case's start. A change is made after the case is loaded and before Run, so it acts from the first scan, unless the text says it was carried through Reset, which keeps the learner's settings as Section 1.6.4 describes. Each case covers one segment of the route. A climber lifts off when it is 0.5 m above the platform.

12.6.1 A Weight That Fades

The empty 5 t climber of case 1 rests its 48.8 kN weight on the platform until the release has proved the drive, lifts off at 1.775 min and takes the anchor tension from 320 to 271.2 kN. The tension is lowest, 260.3 kN, at 5.01 km, where the drive still accelerates the climber against 9.66 kN of drag. The drag peaks at 9.80 kN at 5.60 km and 56.8 m/s, the power at 3.56 MW at 6.91 km, and above the air the climber cruises on 46.73 kN at 140 km, the weight at an effective gravity of 9.347 m/s^2 with no drag left. It docks at 48.57 min, 1.62 m from the station mark, having drawn 2084 kWh, and its speed stays within 0.46 m/s of the reference. Without the feedforward the release has no weight to ramp to, the climber rests on the platform until the integral has built up the whole weight, lifts off at 3.56 min, falls 3.45 km behind its reference and catches up at 1.05 m/s^2.

Table 12.6 What Each Case Measured

Case	Measured outcome
1, lift-off, cruise	1.775 min on 48.8 kN, 46.73 kN at 140 km
1, drag, power	9.80 kN at 5.60 km, 3.56 MW at 6.91 km
1, dock	48.57 min, 1.62 m off the mark, 2084 kWh
1, no feedforward	lift-off 3.56 min, 3.45 km behind
2, speed limit	36.28, 37.29, 38.08 m/s at 10, 40, 100 km
2, trip, motor	69.9 min, 116.1 °C at 121.4 km
2, 12 and 22 t	forecast 145 °C, tension and traction, stays
3, departure sway	30.84 off, 26.86 jerk, 10.02 m a_{max} 0.1
3, shaped departure	ZV 20.09 m, ZVD 13.17 m
3, after the move	21.36, 2.31, 2.43 m off, ZV, ZVD
4, guide load	fixed 6.94 kN, scheduled 5.42 kN
4, lift-off to 20 km	8.1 min fixed, 16.2 min scheduled
5, $g_{eff} = 0$	14.325 min, 86.7 N to hold at the station
5, constant g_0	false 107.9 kN, overspeed in 2.5 s, 169.1 m/s
6, forecast, hold	147.6 °C at 9.08 min, held 11.98–60.93 min
6, dock, check off	108.98 min at $\leq$ 64.1 °C, off 168.3 °C
7, docked sway	28.84, 6.97, 1.19 m off, ZV, ZVD
7, dock time	24.83, 27.48, 30.73 min
8, fixed 40 m/s	brake 1037 °C, 1421 kWh back, 1438 kWh heat
8, scheduled	518 °C, 2314 kWh back, 72.5 min
8, no regeneration	570 min to the anchor

TRY IT IN THE APP **1 · Unloaded Climb**

Load the scenario and press Run. Watch the drive force on the Forces trend. The empty climber rests on the platform until its drive holds the 48.8 kN weight, lifts off at 1.775 min and, above the air, cruises on 46.73 kN at 140 km. The force it must hold has fallen with height. It docks at 48.57 min.

Tripling the mass in case 2 on the same 6 MW supply limits the speed to 36.28, 37.29 and 38.08 m/s at 10, 40 and 100 km as the weight falls, and the station is reached 69.9 min after lift-off instead of 46.8 min. The anchor tension falls to 165.5 kN and the rollers work at 68.7 % of their traction. The copper loss is nine times the empty climber's, and above 40 km only the radiator removes it, so the motor peaks at 116.1 °C at 121.4 km. On the 0.5 t steps of the load sheet, the largest cargo that leaves the platform is 11.5 t, with a forecast of 134 °C. At 12 t the forecast is 145 °C, above the 140 °C allowed for a start, and at 14 t it is 208 °C. On the same steps the supervisor names heat as the refusing limit from 12 to 15.5 t, the grip of the rollers from 16 to 20 t and the anchor tension from 20.5 t. With 22 t the load-sheet weight of 263.6 kN would take the anchor tension to 56.4 kN, below its floor, and exceed the 225 kN the rollers can grip. Heat, and not power, sets the largest load.

Near the geostationary altitude the weight is almost gone. In case 5 the effective gravity is 0.001 37 m/s^2 at the start, crosses zero 14.325 min in at 35 786 km and is −0.007 88 m/s^2 at the station, where holding the 11 t climber takes 86.7 N pulling inward. From 2 to 40 min the drive delivers between −52.9 and 44.6 N, and the climber draws 34.7 to 44.5 kW, most of it the 40 kW auxiliary load. The Coriolis force at 100 m/s is 160.4 N, the climber sits 66.0 km west of the radial line, and the slowest mode holds its 157.8 km deflection at the counterweight through the case while the 7.93 h mode moves between 2.32 and 1.56 km. The climber docks at 101.80 min. Switched to the constant g_0 at 20 min, 34 km past the geostationary altitude, the feedforward believes in a 107.9 kN weight, the rollers pass their speed rating 2.5 s after the scan that applies it and the speed reaches 169.1 m/s before the integral takes the false weight back. There, a gravity-model error of 0.1 % of g_0, 0.0098 m/s^2, is larger than the whole effective gravity at the station.

Pause. Try Exercise 12.1 before reading on.

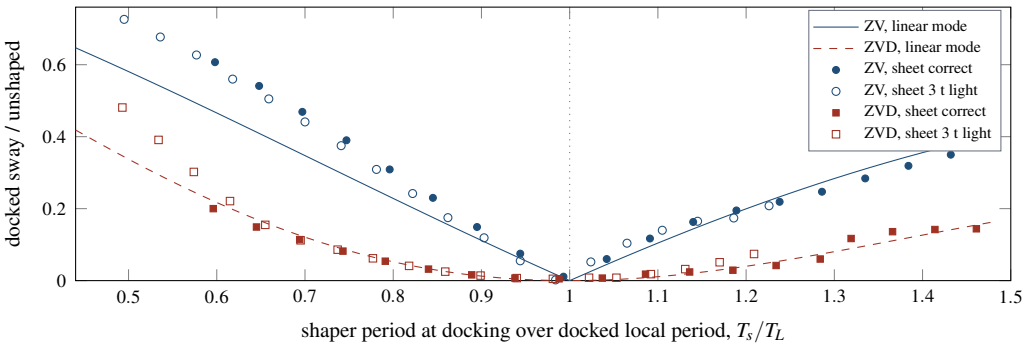


Figure 12.3 Measured sensitivity of a shaped stop to its timing, case 7 run at 19 shaper periods from 60 to 150 % of the prediction, with a correct load sheet (the cargo error set to zero and carried through Reset, since the true mass is fixed when a case starts) and the shipped 3 t light one. Each docked residual is divided by the unshaped one, 18.33 m and 28.84 m, and plotted against the shaper period in force when the climber docks, after the re-timing near the station. The thin solid line (ZV) and the thin dashed line (ZVD) are the residual of the same impulse trains on a linear mode with $\zeta = 0.21$. Circles are ZV runs and squares ZVD runs, filled with a correct sheet and open with the light one.

12.6.2 Sway at a Departure and at a Stop

The 11 t climber of case 3 holds station at 150 km on its drive and leaves for 400 km at 5 min, reaching the 54.0 m/s its supply allows in 3.99 min. The quasi-static lean its Coriolis force demands grows to 78.0 m west within 20 min as the climber rises. Unshaped, the largest sway, 30.84 m at 3.05 min after departure, is a lag rather than an overshoot, because the lean has already reached 50.3 m while the climber has moved only 19.5 m toward it. The climber then overshoots the lean by 21.36 m at 6.77 min. A jerk limit of 2 mm/s^3 barely changes the lag (26.86 m), and an acceleration limit of 0.1 m/s^2, which spreads the start over about a period, cuts it to 10.02 m at the cost of a trip 5.5 min longer. With the shapers the lag falls to 20.09 m (ZV) and 13.17 m (ZVD), 0.651 and 0.427 of the unshaped value, close to the first impulses of 0.664 and 0.441, because until the second impulse arrives the mode sees only the first share of the speed change. What the shapers remove is the ringing after the move. In the 20 min after each shaped acceleration has fallen below 0.01 m/s^2 the largest sway is 2.31 m with ZV and 2.43 m with ZVD, against the 21.36 m unshaped overshoot, and at the 400 km station the unshaped stop leaves 62.58 m ringing while docked, ZV 0.34 m and ZVD 0.07 m. The sway read during a move is this lag, which falls only by the first impulse's share, while the shapers remove nine tenths of the ringing that follows.

🕮 **Pause.** Try Exercise 12.2 before reading on.

Case 7 releases the local mode at a stop, as the Coriolis push of a 40 m/s cruise is removed, with a load sheet 3 t light. Unshaped, the climber docks at 24.83 min with 28.84 m of sway ringing. The impulses freeze when the planner starts braking, at a predicted period of 8.52 min for the 11 t on the sheet. Once the unshaped reference reaches the station they are re-timed to the period predicted there, 8.82 min for ZV at 20.92 min and 147.7 km and 8.78 min for ZVD at 20.22 min and 146.1 km, and the continuity shift sends the reference 89 m and 159 m back to the station. Carried through Reset, ZVD is re-timed at 20.15 min and 146.0 km instead, because a shaper switched on during the cruise first moves the unshaped reference by 6.5 km to keep the shaped one continuous. Both periods remain short of the 10.74 min mode that the 14 t on board gives the docked climber. ZV docks at 27.48 min and leaves 6.97 m, and ZVD docks at 30.73 min and leaves 1.19 m. Each shaper costs part of a period.

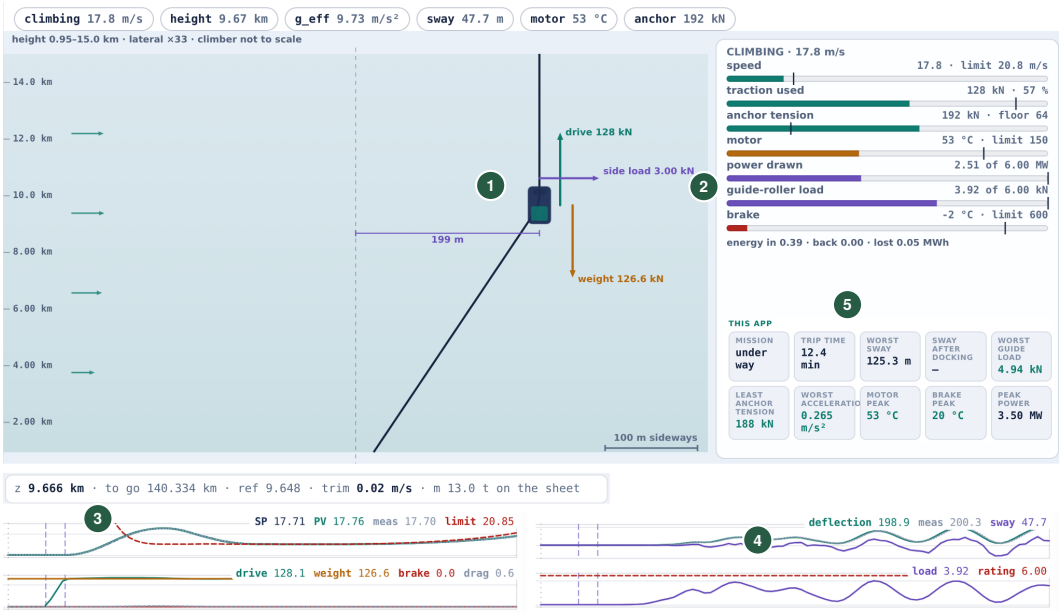


Figure 12.4 The application in the graded transfer, *Challenge, Cargo to the 150 km station*, run with the suggested configuration (scheduled speed limit, ZV shaper and $g_{eff}(r)$ feedforward) and paused at $t = 14.00$ min. The picture is a composite of parts of one screenshot: the Climber view with its plate, the scorecard tiles set into the plate's empty lower part, the station-stop status line and four of the trend strips. The 13 t climber is in the jet at 9.666 km, climbing at 17.76 m/s under a 20.85 m/s limit. In the run at that moment its drive develops 128.0 kN against a 126.6 kN weight and 0.6 kN of drag. The scene draws the height from 0.95 to 15.0 km and the lateral direction at 33 times that scale, and the 3.00 kN side-load arrow and the 199 m eastward deflection at the climber (1) are each read on their own labeled scale. The plate reads 3.92 of 6.00 kN on the guide rollers (2), 192 kN of anchor tension over its 64 kN floor, 57 % of the traction, 2.51 of 6.00 MW drawn, 53 °C at the motor and −2 °C at the brake in the cold air. On the strips the dashed vertical marks are the supervisor's release and departure, at 1.02 and 1.57 min of the run, the speed limit dips ahead of the jet (3), to 10.1 m/s in the run, and rises again, the deflection reads 198.9 m with 47.7 m of local sway (4), and the guide-roller load swings below its rating. The scorecard's worst sway of 125.3 m (5) is mostly the climber lagging a lean that the gusts move faster than its 2.49 min local mode at 8.4 km can follow, while its worst guide load is 4.94 kN.

A sweep of the shaper period is carried out in Figure 12.3 over 76 runs, all of which dock. Plotted against the ratio of the shaper period in force at docking to the docked local period, the two load sheets fall on one curve for each shaper, and the residual of the same impulses on a linear mode describes them with a root-mean-square difference of 0.053 of the unshaped residual. The notches of both shapers sit at 0.98 to 0.99. With a correct sheet ZV leaves 0.21 m at 0.993, where the unmodified prediction re-timed near the station, 8.82 min, is within 0.7 % of the docked 8.888 min, and with the light sheet, retimed to 120 % of the prediction, it leaves 0.06 m at 0.984. The period that matters is therefore the one in force for the last impulses, and the 8.52 min frozen at the start of braking would place the notches near 0.95. The residual stays below a tenth of the unshaped value between the samples at 0.74 and 1.21 with ZVD, and only between 0.94 and 1.02 with ZV. The shipped case sits at 0.822 for ZV and 0.818 for ZVD, where they leave 0.242 and 0.041 of the unshaped residual.

Table 12.7 Transfer Configurations Including Twenty Docked Minutes. Result Names the Failed Parts, Guide Is the Worst Guide-Roller Load and Docked the Worst Sway While Docked

Configuration	Result	Guide (kN)	Docked (m)	Dock (min)
shipped[a]	guide, sway	8.15	27.47	60.97
scheduled only	sway	5.10	27.51	69.72
ZV only	guide	8.12	0.22	63.98
g_{eff} only	guide, sway	8.15	27.42	60.97
scheduled, g_{eff}	sway	5.10	27.47	69.93
ZV, g_{eff}	guide	8.12	0.24	64.20
scheduled, ZV, g_0	passes	4.94	0.36	75.10
suggested[b]	passes	4.94	0.36	75.07
suggested with ZVD	passes	4.87	0.10	79.62
suggested, no feedforward	acceleration	4.73	0.27	75.43

[a] fixed 60 m/s, unshaped, constant g_0 [b] scheduled, ZV, $g_{eff}(r)$

12.6.3 Wind, Power and Brakes

In the crosswind of case 4 the fixed schedule meets the jet at 50.2 m/s, 5.17 km up in an 11.7 m/s wind, where the speed-dependent part of the side load is 4.70 kN against 1.64 kN of crossflow drag, and the guide rollers see 6.94 kN against their 6.00 kN rating, an overload the scorecard records. Scheduled, the limit plans to 80 % of the rating with 2.5 gust intensities in hand, the climber slows to 10.1 m/s at 6.0 km and the load peaks at 5.42 kN. The price is time, 16.2 min from lift-off to 20 km against 8.1 min. On seven other gust seeds the scheduled peak ranges from 4.26 to 6.53 kN, past the rating on two of them, and the fixed one from 7.28 to 9.10 kN, so the gust allowance makes an overload rare rather than impossible. Hurrying through the windy layer is the wrong instinct, because 74 % of the side load at the fixed schedule's peak grows with the climbing speed.

> **TRY IT IN THE APP** **4 · Wind in the Atmospheric Segment**
>
> Run the scenario with the scheduled speed and watch the climber slow as it nears the jet, to 10.1 m/s at 6.0 km, while the guide-roller load peaks at 5.42 kN. Press Reset, set the speed schedule to Fixed and run again. The load reaches 6.94 kN, past its 6.00 kN rating, and the climber is 20 km up 8.1 min after lift-off instead of 16.2 min.

Six minutes into case 6 the supply ramps to 20 % over 3.2 min, which leaves the 11 t climber 8.97 m/s. At that speed the motor would spend about 49 min between the air that cools it and the 40 km where its radiator opens. Before the ramp has finished, at the check at 9.08 min, 13.4 km up, the forecast reaches 147.6 °C, past the 147 °C at which a moving climber is stopped. The supervisor stops, holds on the clamp from 11.98 to 60.93 min at 13.9 km, where the motor carries no current and cools while the forecast at the reduced supply reads 168 to 173 °C, and resumes once the supply has returned at 60 min, docking at 108.98 min with the motor never above 64.1 °C. With the check off the climber carries on and the motor reaches 168.3 °C at 57.34 min, 39.6 km up, a recorded over-temperature within 0.3 K of the 168.0 °C forecast at 10.10 min. On six noise seeds the stop falls at the same check and the peak with the check off spans 167.9 to 168.3 °C.

Going down in case 8, the climber's weight must be held back all the way, and at the fixed 40 m/s that is about 4 MW of braking. The motor returns 1421 kWh to the supply, and the friction brake takes 1438 kWh, passes its 600 °C limit 7.79 min in and peaks at 1037 °C. Scheduled, the descent is held to the speed that regeneration and the brake's cooling at 520 °C can absorb, 23.1 m/s on average between 60 and 90 km, and the brake peaks at 518 °C while 2314 kWh return to the supply and 448 kWh become brake heat, in a trip of 72.5 min instead of 45.5 min. Without regeneration the schedule allows 2.7 m/s and the climber reaches the anchor after 570 min.

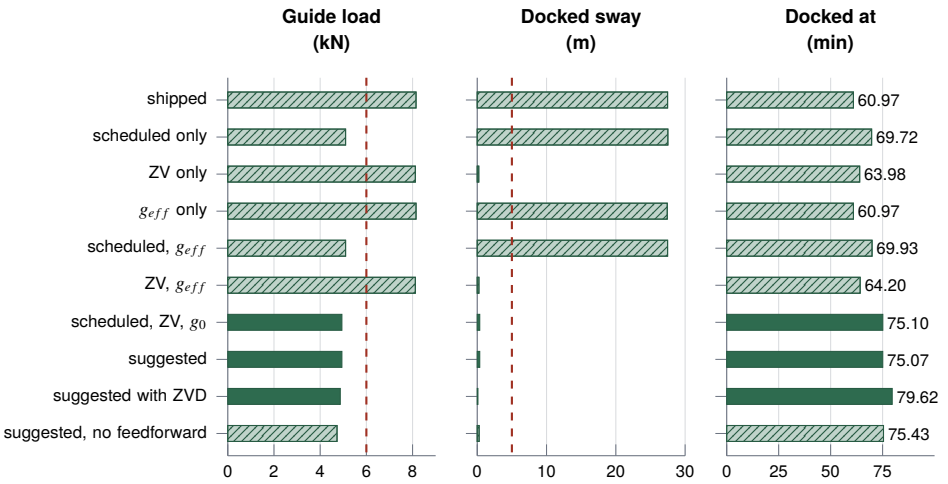


Figure 12.5 Table 12.7 redrawn: the graded transfer in ten configurations. Solid bars pass, hatched bars fail, and the dashed lines are the 6 kN guide-load and 5 m docked-sway limits. Notice that each failing configuration crosses one dashed line or the other, the last row failing on acceleration instead, and that every passing one docks later than any that fails on guide load or sway.

> ✦ **DID YOU KNOW?** **Slower Down, More Energy Back**
>
> From the simulator: descending 100 km at a fixed 40 m/s, the friction brake peaks at 1037 °C and 1421 kWh return to the supply. Held to the scheduled speed, the brake peaks at 518 °C and 2314 kWh return, in a trip of 72.5 min (case 8 of Table 12.6).

12.6.4 The Graded Transfer

The transfer carries 8 t of cargo from the anchor to the 150 km station through the 15 m/s jet with 2 m/s gusts. It passes only if the climber docks within 2 m of the mark and stays 20 min, the guide load stays within 6 kN, the docked sway within 5 m, the anchor tension above 64 kN, the acceleration below 0.5 m/s^2, the motor below 150 °C and the brake below 600 °C, with no slip, no overspeed, no excess of the supply and a closed ledger, and every part is scored on its worst moment. A changed model parameter or disturbance voids the attempt, and the crediting otherwise follows Section 1.6.4. Table 12.7 lists the configurations measured. Figure 12.5 draws the same table as bars against the two limits that most configurations miss.

As shipped, at a fixed 60 m/s, unshaped and with the constant g_0, the climber docks at 60.97 min, but the guide rollers see 8.15 kN at 6.13 km, 2.15 kN over their rating, and 27.47 m of sway rings while it is docked, 22.47 m over its limit. The schedule alone removes the overload and ZV alone removes the sway, and each leaves the other failure in place. Together they pass. The suggested configuration, scheduled with ZV and the $g_{eff}(r)$ feedforward, keeps the guide load to 4.94 kN, 1.06 kN inside the rating, the anchor tension above 188.4 kN, the acceleration below 0.265 m/s^2 and the motor below 93.5 °C, and docks 0.742 m from the mark at 75.07 min with 0.36 m of sway, 14.1 min later than the shipped run. Applied before Run or carried through Reset, the same settings give the same trajectory and result. The constant g_0 in its place still passes, docking 0.353 m from the mark. These errors lie inside the 1 m window on the measured height, read with 0.5 m of noise, within which the supervisor docks, so they reflect that window and not the feedforward. Below 150 km the constant g_0 is at most 6.35 kN wrong for the 13 t climber, and the integral carries it. Removing the feedforward altogether fails the acceleration limit at 0.542 m/s^2. ZVD in place of ZV passes with 0.10 m of docked sway and arrives 4.55 min later. The gusts follow the case's own seed, so every

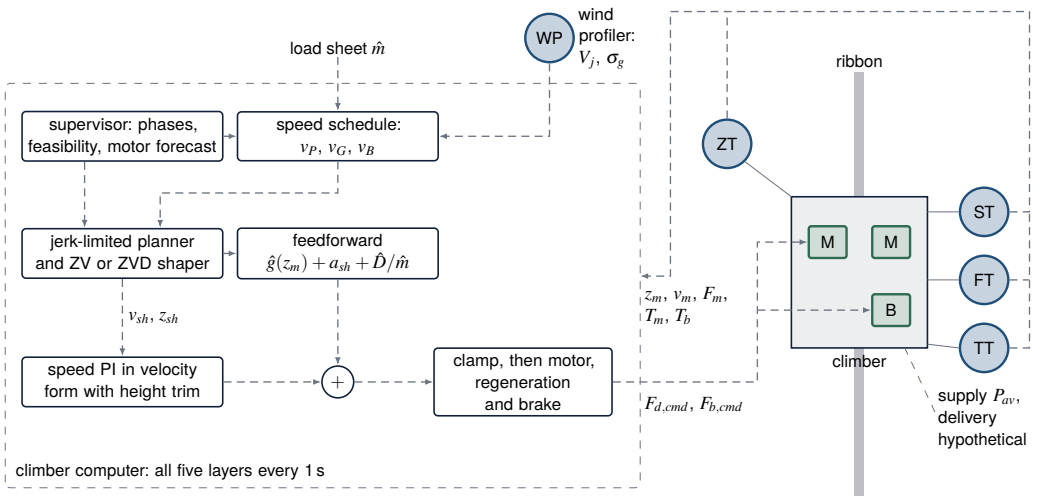


Figure 12.6 The five layers of Section 12.5 on the hardware a climber would carry, which no one has built. Sensors of the kinds flown on tethers and used on drives report height (ZT), roller speed (ST), traction force (FT) and the motor and brake temperatures (TT), and a wind profiler (WP) reports the jet to the schedule. One computer runs the supervisor, the schedule, the planner with its shaper, the feedforward and the speed loop, and splits the clamped demand between the roller motors (M) and the friction brake (B). Dashed gray arrows are signals. Notice that the feedforward block needs only the load sheet, its gravity and drag models at the measured height and the planner's acceleration, while the speed loop corrects whatever they get wrong.

attempt meets the same wind. On seven other seeds the suggested configuration passes six times and on the seventh the guide load reaches 7.02 kN, while the shipped one fails the guide load and the sway on all seven. The feedforward that the transfer barely needs is the one that case 5 cannot do without.

12.6.5 Convergence

The application's 38 checks pass, and they are the evidence for the numerical side of the claim. One of them runs the planner over 360 moves from 10 m to 20 km, from rest and at 9 m/s toward or away from the target, at three sets of limits, and requires every move to capture within 4 h. Halving the 0.5 s step leaves the transfer docking on the same 1 s scan and leaves its worst guide load, docked sway and motor peak unchanged to the digits reported here. Doubling the ribbon grid to 1600 elements moves the six periods by 0.0028 to 0.0346 %, and keeping eight modes instead of six moves the worst sway of case 7 by 0.47 %. The ledger of Equation (12.12) closes to 3.8×10^{-10} of the electrical energy through it on the scheduled descent, 4.5×10^{-10} on the fixed one and 4.8×10^{-11} or better on the unloaded climb, the power restriction and the transfer. Without regeneration, when little electrical energy passes, the residual is 1.2×10^{-5} of it. A Coriolis step at 10 000 km drives modes 2 to 4 to twice their static share $\phi_k \Delta F / (M_k \Omega_k^2)$ within 2.4 %, and a tachometer reading 0.5 m/s fast leaves the case 5 cruise 24.97 m behind its height reference, the $0.5/K_x = 25$ m at which the trim cancels the bias. The transfer as it runs is shown in Figure 12.4.

12.7 Technology in Practice

No space elevator climber has been built, so its hardware can only be borrowed from tethers that have flown and from drives that work on the ground. Figure 12.6 places the five layers of Section 12.5 on such devices.

What Measures It

Flown tether systems measured what this chapter's climber needs. A deployer flown in 1993 and 1994 counted the turns of its tether with light-emitting diodes and carried a tensiometer and a temperature sensor on its brake, and the payload at the far end carried a three-axis tensiometer and a three-axis accelerometer [17]. These are the length, speed, force and temperature readings of Section 12.3, where the height is ranged with 0.5 m of noise and the speed is read by a tachometer, each once a second.

> **TECHNOLOGY Sensor: Counting What Passes**
>
> A flown tether deployer measured its tether by counting. Two light-emitting diodes counted the turns of tether leaving a fixed spool, and a third detected when the spool was almost empty. To give the deployed length and its rate, the turns had to be mapped and converted into length [17]. An optical encoder on a shaft counts the same way, with a slotted wheel that chops a light beam into a train of pulses [18]. A count of this kind gives position and speed directly, but only of the surface it touches and only as well as the map from counts to meters. In the model the speed reading carries noise and the adjustable bias b_v of Table 12.1, which the height trim cancels at a standing offset.

What Moves It

The same deployer controlled its tether with a friction brake: a stepper motor wrapped the tether around a brake post or unwrapped it to change the tension and with it the deployment speed, the friction growing with the contact area [17]. A climber must also push, and the 1999 workshop judged that traction wheels would be slow and would wear the vehicle and the structure [15]. The model gives its rollers a current limit of 280 kN, a grip of 225 kN and a drive lag of 2 s, and its brake a fade above 650 °C.

> **TECHNOLOGY Final Element: A Motor That Brakes as a Generator**
>
> A load lowered by an electric hoist does work on the hoist. A friction brake turns that energy into heat, while regenerative braking switches the motor into a generator and sends the energy back to a battery or the grid [18]. In an alternating-current drive this takes extra power transistors that invert the drive's internal direct voltage back to the line, which makes the drive more expensive and potentially more prone to faults, a cost repaid where the recovered energy is significant. Every electronic braking method depends on the drive working, and where the motor itself can brake, a separate mechanical brake may not be needed [18]. The climber keeps both: regeneration up to the supply's acceptance P_{rg} and a friction brake that fades when hot.

Where the Controller Runs

A terrestrial speed drive closes its own loop on a tachometer and so becomes the secondary controller of a cascade, taking its speed setpoint from the process controller [18]. The flown deployer ran its control logic in its own electronics box and followed a profile computed before the mission [17]. The parts of the handbook read for this section give no execution period for that logic. The application runs all five layers on one computer every 1 s: the supervisor, the schedule, the planner with its shaper, and the speed loop whose bias term carries the gravity feedforward.

> **IN PRACTICE Pay Out When the Tension Is High**
>
> Tethers flown in orbit left rules of practice. The field's handbook states that tension laws can control both the swing of a tethered system and the oscillations of its tether. A swing in the orbit plane is damped by paying out tether when the tension is greater than usual and retrieving it at other times, which absorbs energy from the swing. Because that swing causes large variations in tension through the Coriolis effect, such maneuvers damp it quickly [17]. The climber has no reel and no lateral actuator, so in this chapter only the timing of its speed changes acts on its swing.

What Is Modernizing It

Three developments bear on the climber. The first is power delivered by light. A 2005 assessment took laser power beamed from a ground station to the climber's electric motors as its baseline, demonstrated on the ground but not for a laser that tracks a climber to 100 000 km [10]. A NASA test of a high-intensity laser receiver drew 13.6 W/cm^2 of electrical power from a single photovoltaic cell at a conversion efficiency of 24 percent [19]. In the model this supply is the declared P_{sup} of 6 MW. The second is competition: a NASA prize grant funded tether strength and climber contests meant to advance robotic climbers and beamed power [10]. The third is simulation. A NASA program written for tethered satellite missions was extended to elevator-sized tethers to study deployment, climber dynamics, atmospheric effects and waves induced to avoid objects [10].

12.8 Limitations

The largest limitation is the concept. The ribbon's strength, the power delivery, the guideway, the anchor platform and the deployment are hypothetical, and the dimensions of Section 12.3 follow from declared inputs rather than from a design. The application shows that its equations converge and that its reduced-order behavior is plausible. It does not show that an Earth space elevator can be built, and its numbers are not the numbers of one.

> ⚠ **OPEN CHALLENGE** **A Ribbon No One Can Make Yet**
>
> The most critical technology for an Earth space elevator is a carbon nanotube reinforced composite with a strength in the 100 GPa range. Stresses in the ribbon have been calculated at 60 to 80 GPa. Single nanotubes have been reported above 150 GPa, but making a fiber of that strength from them had been attempted without success, which in a 2005 assessment put the whole Earth-based system at the second of nine technology readiness levels [10]. Progress would be a fiber, and then a ribbon, that keeps that strength after fabrication, splicing and the coatings it needs along its length.

The ribbon is axially rigid and moves laterally in the orbit plane only, the lateral motion does not feed back on the radial, the local mode is drawn with a static shape, and its radiation damping T/c holds only while the waves it launches return more slowly than its period. Cases that start in motion start on the quasi-static shape, an idealization for the 127.4 h mode. The thermal, aerodynamic and structural coefficients are illustrative, and the limits and margins are teaching values. No decision reads the lateral fix, the anemometer or the guide-load measurement, so the lateral dynamics are handled by the schedule and the shaper alone, with no lateral feedback.

The scorecard's worst sway is taken against the quasi-static lean of the moment. In gusts it is dominated by the lean moving faster than the local mode, since the first 14 min of the suggested transfer reach 125.3 m with the 2 m/s gusts and 13.9 m with the gust intensity set to zero, which also removes the schedule's gust allowance, while the worst guide load moves only from 4.85 to 4.94 kN, and on a departure it records the lag rather than the ringing that shaping removes.

> ⚠ **OPEN CHALLENGE** **Ribbon Control Is Still on Paper**
>
> A built elevator would need control systems that hold the ribbon in its orbital slot and govern its tension, induced waves, climber dynamics and atmospheric effects. A climbing climber changes the ribbon's angular momentum, and if the ribbon cannot absorb it, climbers may need thrusters of their own. In 2005 ribbon control was only at an analytical stage, studied with a simulation program built for tethered satellite missions [10]. Progress would be such control demonstrated on a long tether, with the climbers' own speed commands among its inputs.

12.9 Conclusion

The space elevator climber of this chapter is a plant whose weight fades to nothing and reverses along the route, and whose structure answers every change of speed with a slow sideways swing.

The effective gravity falls from 9.764 m/s^2 at the anchor to zero 35 786 km up, where a constant-g_0 feedforward believes in 107.9 kN that is not there and drives an 11 t climber to 169.1 m/s. The climber's local mode, its own mass on the string below it with the ribbon above as the dashpot T/c, is approximated to within 2.7 % up to 150 km, and a departure excites it through 86.6 N of Coriolis force, leaving 21.36 m of ringing unshaped and 2.31 m with ZV. Swept over a stop, the shaped residual follows the linear curve with a root-mean-square difference of 0.053. The schedule cuts a 6.94 kN side load to 5.42 kN and a 1037 °C brake to 518 °C, and the supervisor holds for 48.95 min rather than let the motor reach 168.3 °C. The ribbon's strength and the power delivery remain hypothetical, and the model is planar, axially rigid and coupled one way. A fuller model would add out-of-plane motion, axial waves, several climbers and a lateral estimator fed by the measurements the application already logs. The simulator turns a structure that may never be built into a working example of what feedforward, scheduling and shaping each carry when the load, the structure and the limits all move with the route.

KEY POINTS

- The weight a climber's drive must hold falls with height and reverses past the geostationary radius, so a feedforward right at one end of the route is wrong at the other unless it models the physics.
- Every change of speed changes the Coriolis push, and a departure shorter than half the local period is close to a step for that mode.
- Shapers remove the ringing after a move, 21.36 m unshaped against 2.31 m with ZV, but not the lag during it.
- The schedule alone removes the transfer's guide overload and ZV alone its docked sway. Together they pass.

Where else this applies. Volume 7, Chapter 12 times the same ZV and ZVD impulses against a swinging crane load, and Volume 20, Chapter 3 feeds a load sheet forward on another vehicle whose load changes along the route.

References

[1] S. Kaczmarczyk and W. Ostachowicz, "Transient vibration phenomena in deep mine hoisting cables. Part 1: Mathematical model," *Journal of Sound and Vibration*, vol. 262, no. 2, pp. 219–244, 2003.

[2] O. J. M. Smith, "Posicast control of damped oscillatory systems," *Proceedings of the IRE*, vol. 45, no. 9, pp. 1249–1255, 1957.

[3] N. C. Singer and W. P. Seering, "Preshaping command inputs to reduce system vibration," *Journal of Dynamic Systems, Measurement, and Control*, vol. 112, no. 1, pp. 76–82, 1990.

[4] W. Singhose, "Command shaping for flexible systems: A review of the first 50 years," *International Journal of Precision Engineering and Manufacturing*, vol. 10, no. 4, pp. 153–168, 2009.

[5] P. Lambrechts, M. Boerlage, and M. Steinbuch, "Trajectory planning and feedforward design for electromechanical motion systems," *Control Engineering Practice*, vol. 13, no. 2, pp. 145–157, 2005.

[6] J. D. Isaacs, A. C. Vine, H. Bradner, and G. E. Bachus, "Satellite elongation into a true "sky-hook"," *Science*, vol. 151, no. 3711, pp. 682–683, 1966.

[7] J. Pearson, "The orbital tower: A spacecraft launcher using the Earth's rotational energy," *Acta Astronautica*, vol. 2, no. 9–10, pp. 785–799, 1975.

[8] B. C. Edwards, "Design and deployment of a space elevator," *Acta Astronautica*, vol. 47, no. 10, pp. 735–744, 2000.

[9] S. S. Cohen and A. K. Misra, "The effect of climber transit on the space elevator dynamics," *Acta Astronautica*, vol. 64, no. 5–6, pp. 538–553, 2009.

[10] D. V. Smitherman, Jr., "Critical technologies for the development of future space elevator systems," in *56th International Astronautical Congress*, 2005, paper IAC-05-D4.2.04; NASA Marshall Space Flight Center, NTRS 20060000015.

[11] N. M. Pugno, "On the strength of the carbon nanotube-based space elevator cable: from nanomechanics to megamechanics," *Journal of Physics: Condensed Matter*, vol. 18, no. 33, pp. S1971–S1990, 2006.

[12] S. S. Cohen and A. K. Misra, "Elastic oscillations of the space elevator ribbon," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 6, pp. 1711–1717, 2007.

[13] P. Williams and W. Ockels, "Climber motion optimization for the tethered space elevator," *Acta Astronautica*, vol. 66, no. 9–10, pp. 1458–1467, 2010.

[14] P. K. Aravind, "The physics of the space elevator," *American Journal of Physics*, vol. 75, no. 2, pp. 125–130, 2007.

[15] D. V. Smitherman, Jr., "Space elevators: Building a permanent bridge for space exploration and economic development," in *AIAA Space 2000 Conference and Exposition*, 2000, aIAA Paper 2000-5294; NASA Marshall Space Flight Center preprint, NTRS 20000105202.

[16] M. P. Cartmell and D. J. McKenzie, "A review of space tether research," *Progress in Aerospace Sciences*, vol. 44, no. 1, pp. 1–21, 2008.

[17] "Tethers in space handbook," Smithsonian Astrophysical Observatory for NASA Marshall Space Flight Center, NASA Contractor Report NASA/CR-97-206807, 1997, nTRS 19980018321.

[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] D. E. Raible, D. Dinca, and T. H. Nayfeh, "Optical frequency optimization of a high intensity laser power beaming system utilizing VMJ photovoltaic cells," NASA Glenn Research Center, Cleveland, OH, NASA Technical Memorandum NASA/TM-2012-217256, 2012, nTRS 20120007096.

12.10 Exercises

12.1 ○ Concept check · 2 min

In case 5 the feedforward is switched from $g_{eff}(r)$ to the constant g_0 at 20 min, when the 11 t climber is 34 km past the geostationary altitude. What happens next?

(a) Little: the effective gravity there is tiny, so the integral absorbs a tiny error unseen

(b) The climber falls back inward, because the bias no longer holds up any weight

(c) The demand saturates on its lower limit and the supervisor holds the climber on the clamp

(d) The bias asks for a 107.9 kN weight that is not there and throws the climber outward

12.2 ○ Concept check · 2 min

In case 3 the ZV shaper lowers the worst sway of the departure only from 30.84 m to 20.09 m, yet it lowers the ringing after the acceleration from 21.36 m to 2.31 m. Why does it do so little during the move?

(a) Until the second impulse arrives, the mode sees only the first impulse's share of the speed change

(b) The shaper is timed for the wrong period, because the load sheet of case 3 understates the cargo

(c) The planner's jerk limit overrides the shaper during the acceleration and restores the unshaped start

(d) The scorecard measures sway from the radial line, so the growing lean itself is counted as sway

12.3 ○ Concept check · 1 min

Table 12.3 gives a damping ratio of 0.062 for the local mode at 10 km and 0.528 at 1000 km. Why is the mode so much better damped high up?

(a) The guideway damping ζ_s rises with height as the motor's radiator opens

(b) The impedance T/c of the ribbon above changes little while the stiffness k_L falls

(c) The six retained string modes lend the local mode their 1 % damping

(d) The Coriolis force grows with height and acts as a damper on the sideways motion

12.4 ▭ Theory and calculation · 6 min

Use Equation (12.1) with the constants of Table 12.2.

(a) Compute the effective gravity at the anchor and at the 150 km station, and by how many percent the constant g_0 overstates each. *Checkpoint:* 9.764 m/s^2 and 9.318 m/s^2, high by 0.43 % and 5.24 %.

(b) Compute the geostationary radius and its height above the anchor. *Checkpoint:* 42 164 km, 35 786 km up.

(c) The graded transfer carries a 13 t climber to 150 km. By how much does a constant-g_0 feedforward overstate its weight there, and which part of the speed loop carries the error? *Checkpoint:* 6.35 kN, carried by the integral.

12.5 Theory and calculation · 6 min

The 11 t climber holds station at 150 km with $T_0 = 320\,\mathrm{kN}$, $k_a = 5000\,\mathrm{N/m}$, $\zeta_s = 1\,\%$ and $c = 6325\,\mathrm{m/s}$. The ribbon tension at the station is 331.7 kN.

(a) Compute the link stiffness k_L of the uniform-string approximation and the period T_L of Equation (12.13). *Checkpoint:* $k_L = 1.450\,\mathrm{N/m}$, $T_L = 9.12\,\mathrm{min}$.

(b) Compute the damping ratio ζ_L of Equation (12.13) with that k_L, and compare both results with the application's 8.888 min and 0.212. *Checkpoint:* $\zeta_L = 0.218$, and the period is 2.6 % long.

(c) The climber leaves at 54.0 m/s. Compute the Coriolis push and the quasi-static lean it demands with the application's $k_L = 1.527\,\mathrm{N/m}$. *Checkpoint:* 86.6 N and 56.7 m.

12.6 Theory and calculation · 6 min

Design the shapers of Equation (12.15) for the local mode of the previous exercise, with an undamped period of 8.888 min and $\zeta = 0.212$.

(a) Compute K, the two ZV amplitudes and the time of the second impulse. *Checkpoint:* $K = 0.506$, then 0.664 and 0.336, the second at 4.548 min.

(b) Compute the three ZVD amplitudes and their times. *Checkpoint:* 0.441, 0.446 and 0.113 at 0, 4.548 and 9.095 min.

(c) The unshaped departure of case 3 lags its lean by 30.84 m. Predict the lag with each shaper from the share of the move that the first impulse carries, and compare with the measured 20.09 and 13.17 m. *Checkpoint:* About 20.5 and 13.6 m.

12.7 App challenge · 8 min

Load scenario 1, Unloaded climb, at its default seed, with every setting as loaded.

(a) Run until the run stops by itself, 10 min after docking. From the exported CSV, read `drive_force_kN` on the first row with `alt_km` at or above 140, and read the Worst acceleration tile. Compare the drive force with mg_{eff} at that height. *Checkpoint:* 46.7 kN, equal to 5000×9.347, and 0.311 m/s^2.

(b) Press Reset, set Gravity feedforward to off before the departure at 1 min, and run until the climber is 10 km up. Read the time of lift-off from the height trend and the Worst acceleration tile, and compare both with part (a). *Checkpoint:* Lift-off near 3.6 min against 1.8 min, and 1.048 m/s^2.

Accept 0.1 kN on the force, 0.2 min on a lift-off time and 0.005 m/s^2 on an acceleration.

12.8 App challenge · 8 min

Load scenario 3, Acceleration-induced sway, at its default seed. Each run goes to 15 min, 10 min after the climber sets off. Read the Worst sway tile.

(a) Run as loaded. *Checkpoint:* 30.8 m.

(b) Press Reset, set Jerk limit to 2 mm/s^3 before the departure at 5 min, and run again. *Checkpoint:* 26.9 m.

(c) Reload the scenario, set Input shaper to ZV before pressing Run, and run. Then reload again and repeat with ZVD. Divide each result by part (a). *Checkpoint:* 20.1 m and 13.2 m, ratios 0.65 and 0.43.

Accept 0.2 m on each tile.

12.9 App challenge · 9 min

Load the challenge, Cargo to the 150 km station, at its default seed. Each run stops by itself 20 min after docking. Read the Worst guide load, Sway after docking and Mission tiles, and the verdict.

(a) Run as loaded (fixed schedule, shaper off, constant g_0). *Checkpoint:* 8.15 kN, 27.5 m, docked at 61.0 min, failed.

(b) Press Reset, set Speed schedule to Scheduled and run. Then press Reset, set Speed schedule back to Fixed and Input shaper to ZV, and run again. *Checkpoint:* Scheduled: 5.10 kN, 27.5 m. ZV: 8.12 kN, 0.2 m. Both fail.

(c) Press Reset, set Scheduled, ZV and Gravity feedforward $g_{eff}(r)$, and run. Read also the docking error, column `metric_stop_error_m` of the exported CSV. *Checkpoint:* Passes: 4.94 kN, 0.4 m, docked at 75.1 min, 0.74 m from the mark.

Accept 0.02 kN on a guide load, 0.1 m on a sway and 0.1 min on a time.

12.10 Application area · 4 min

Before every start the supervisor checks that the load-sheet weight $\hat{m}g_{eff}$ stays within $T_0 - T_{a,min} - 8\,\mathrm{kN}$ for the anchor tension and within $0.9\mu_s N_c$ for the grip of the rollers. The empty climber is 5 t, and $g_{eff} = 9.764\,\mathrm{m/s^2}$ at the anchor.

(a) Compute the largest weight each rule allows and the cargo it corresponds to. *Checkpoint:* Tension 248 kN, 20.4 t of cargo, grip 202.5 kN, 15.7 t of cargo.

(b) Section 12.6.1 reports refusals for heat from 12 to 15.5 t, grip from 16 to 20 t and tension from 20.5 t. Which limit sets the largest cargo, and why is it not the power? *Checkpoint:* Heat, from 12 t, so 11.5 t is the largest that leaves.

12.11 Application area · 4 min

Case 8 brings the 11 t climber down 100 km to the anchor.

(a) Estimate the braking power at the fixed 40 m/s near 100 km, where $g_{eff} = 9.464\,\mathrm{m/s^2}$, and compare it with the 2 MW the supply accepts. *Checkpoint:* About 4.2 MW, roughly twice what regeneration may return.

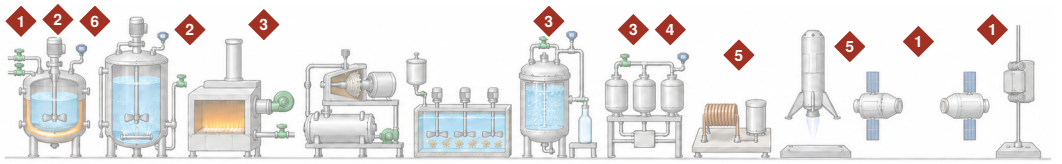
(b) The potential energy released, $m[\Phi(R_E + 100\,\mathrm{km}) - \Phi(R_E)]$, is about 2937 kWh. What share returns to the supply on the fixed and on the scheduled descent, and what does the schedule cost? *Checkpoint:* 48 % and 79 %, and 27 min more.

12.12 Review the agent · 5 min

An AI assistant reviews the departure of case 3 and writes: “The 30.8 m sway is a disturbance that the speed loop rejects too slowly. Raise the proportional gain $K_{c,v}$ from 0.10 to 0.30 s^{-1} so the loop holds the climber against the Coriolis push. The worst sway should fall by about a third.” Find the flaw, and say how you would check the claim in the application.

CHAPTER 13

Summary and Outlook



Open gaps 1 feedforward model range 2 inner-loop speed rules 3 trims and instruments 4 streams in a blend 5 undershoot, authority 6 structure from data

Notice that the tank with two upsets carries three gaps and that the trim gap runs from the heater to the formulation skid, while the vehicles met the gaps of model range and actuator authority. Illustration, not to scale.

IN PLAIN TERMS **What Eleven Plants Taught About Acting Early**

The main idea. Every chapter of this book tried to correct an upset before it reached the thing being controlled, from a water tank to a spacecraft. Some plants measured the upset and acted at once. Others closed a small, fast loop around the part the upset hit first.

The challenge. An early action is only as good as the meter and the model behind it. A meter that drifts, or a model used far from where it was made, can push the plant the wrong way.

Why it matters. Asking what each early action reads, and where its model holds, separated the designs that worked from those that failed.

Chapter 1 reduced the volume to one rule: a structure helps against a disturbance only if it measures the disturbance, or a variable on its path, earlier than the controlled variable does. Eleven chapters then put that rule on eleven plants, from a tank with two upsets to a climber whose weight fades to nothing on the way to orbit. This chapter sets the results side by side, draws a short guide for choosing a structure, 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.

13.1 What the Chapters Showed

Tables 13.1 and 13.2 return to the path map of Table 1.3 with the measured result of each pairing. Each row names the disturbance path that made the plant hard, a structure that met it, one that did not, and the measured result with the table or section of the chapter that reports it.

The columns of the two tables make five points.

The entry point decided the result before any tuning did. On the tank of Chapter 2 each structure removed the upset on its own path and left the other untouched. The cascade divided the pressure error by 4.7 and did not improve the demand error, and the outlet feedforward did the reverse. Several chapters kept a disturbance that no structure reached: the probe lag of Chapter 3, the condenser of Chapter 5, the deceleration authority of a rocket with too small an engine. Those limits were stated, and no tuning moved them. Figure 13.1 sets one like pair of numbers from each row side by side.

A feedforward is only as good as its model, at the bandwidth paid for. The orbital feedforward was exact, and a mean motion 15.8 % low cost 22.6 % more propellant on a slow loop and 2.7 % on one twice as fast. The tank's trim was exact only at the mean supply pressure, which left 0.0444 m of

Table 13.1 Which Structure Matched Which Disturbance Path, Chapters 2 to 7

Chapter	Difficulty	What worked	What failed	Measured result
2, tank level	Supply pressure enters at the valve. Outlet demand enters at the vessel.	A flow cascade for the pressure, feedforward from the outlet flow for the demand, and both together.	Feedforward from the inlet flow, which the valve itself moves. An inner gain past its stability limit.	Worst error 0.2989 single loop, 0.0136 both structures, 0.3590 m inlet feedforward (Table 2.4). Cascade divides the pressure error by 4.7, feedforward the demand error by about 250 (Section 2.6).
3, fermenter oxygen	Agitator errors at the actuator. A feed step on the culture. Drive 15.4 s, holdup 12.4, probe 20 s.	A speed cascade settling in 2.9 s, with a carbon-balance feedforward.	The same cascade with its inner loop detuned to 12.3 s. Any structure against the probe lag.	IAE 10.6 single, 3.5 cascade, 2.1 with feedforward (Table 3.4). Detuned 72.3 against 23.8 (Section 3.6). Break-even at 7.4 s or 13.4 s by inner reset time (Figure 3.4).
4, fired heater	A 2 s fuel valve and a 12 s fan. Unmeasured heating value. A drifting oxygen analyzer.	Cross-limiting selectors. Oxygen trim with a carbon monoxide override.	Air slaved to the measured fuel on a load increase. A trim at ±25 % following its analyzer.	λ_b 0.990 and 3.7 s rich against at least 1.136, for a dip of 5.8 against 4.6 °C (Table 4.5). Over the permit 118 against 0 s (Table 4.6).
5, waste-heat cycle	Lost source heat reaches the superheat through the evaporator. The condenser sets a limit no loop moves.	The flow cascade with feedforward from the source meters.	The superheat loop alone, independent or in cascade. Any setting on 55 °C server water.	IAE 276.7, 437.2 and 29.0 K s, expander isolated 54.9, 83.4 and 0 s (Table 5.5). Best net output of 192 settings −0.59 kW (Section 5.5).
6, water plant	The dose acts 193.6 min to 63 % on a 30 min storm. Organic carbon only from the laboratory.	A dose computed from raw turbidity, with a slow multiplying trim.	Feedback alone, which integrated past the coagulation optimum to its 140 mg/L stop.	Above 1.00 NTU 47.15 against 0 min, alum 415 against 306 kg (Table 6.3). Trim at 600 min: 2.04 against 1.23 NTU target (Section 6.5.3).
7, carbonation line	Filler rate measured. Syrup strength unmeasured. A water header dip at the water valve.	Throughput feedforward, a ratio on the measured water flow and a Brix trim.	The ratio on the water setpoint when the valve cannot follow. The trim switched off.	Trim off: 4.60 m^3 of 6.67 off specification (Table 7.5). Header dip: 10.80 °Bx against 11.33 °Bx on the setpoint basis (Section 7.6).

the combined upset. Near the geostationary altitude a gravity error of 0.1 % of g_0 exceeds the whole effective gravity, so the model error is the disturbance. The water plant's model used the morning's laboratory organic carbon, and the settled water came to rest at 2.13 against 1.21 NTU.

The speed of an inner loop is a measured number, not a ratio from a rule. The fermenter's cascade stopped paying at an inner time constant of 7.4 s with one inner reset time and at 13.4 s with another. Below about 15 %/A the converter's outer loop inherited the inner lag. On the rocket a separation of 3.0 failed with a fast outer loop, while the same ratio reached by slowing the inner loop still landed. On the tank the inner loop's stability limit moved with the supply pressure inside each cycle. The textbook ratio of three to five (Section 1.4.5) is a starting point, and each of these chapters found the true boundary only by sweeping.

Table 13.2 Which Structure Matched Which Disturbance Path, Chapters 8 to 12

Chapter	Difficulty	What worked	What failed	Measured result
8, nanoparticle skid	Four feeds in ratio. Weak lots no flow meter sees. Size reported a minute late.	A supervisor choosing loop and flow, a size trim, a conductivity probe and a window guard.	Coordinated ratio control alone. The supervisor on the metered blend or without the guard.	Rate 441.7 against 600.0 mL/min, passing 6 of 6 seeds. Metered blend 776.7 mL off specification, no guard 1928.5 mL (Table 8.5).
9, boost converter	A right-half-plane zero that moves 17-fold with load. An overload.	An inner current loop with its clamp, and load-current feedforward.	The single voltage loop pushed harder. An inner gain below about 15 %/A.	Overload peak 9.47 against 11.32 A. Excursion 58.6 against 161.9 % (Section 9.6.4). Feedforward: recovery 43 against 804 ms, dip 7.82 against 5.69 V (Section 9.6).
10, rocket landing	The outer loop commands an angle. Zero at 2.80 rad/s. Weight carried at a tilt.	Bandwidth separation 8.7, gravity feedforward with $1/\cos\theta$, no position integral.	Separation 3.0 (a scripted run). The feedforward switched off before a Reset.	Position IAE 123.0 to 82.4 m s while the attitude IAE rose from 0.264 to 4.607 rad s, a hard arrival (Table 10.7). With a 10° tilt, descent IAE 6.83 against 12.94 m (Table 10.5).
11, orbital rendezvous	Gravity gradient and Coriolis terms couple the axes. The coupling is known.	Four model terms that cancel the coupling at the controller's mean motion.	The same gains without them. A mean motion 15.8 % low on a slow loop.	Docked at 1785 against 3516 s, along-track excursion 0.0074 against 11.475 m (Table 11.5). Wrong model: propellant penalty 22.6 at 2 mrad/s, 2.7 % at 4 (Section 11.6).
12, elevator climber	The weight fades to zero at 35 786 km and reverses. The climber's own moves excite a ribbon mode.	Feedforward of the effective gravity. ZV and ZVD shaping.	No feedforward at lift-off. A constant g_0 past the geostationary altitude.	Lift-off at 1.775 against 3.56 min. A false 107.9 kN drives the climber to 169.1 m/s (Table 12.6). Ringing after a move 21.36 against 2.31 m (Section 12.6.2).

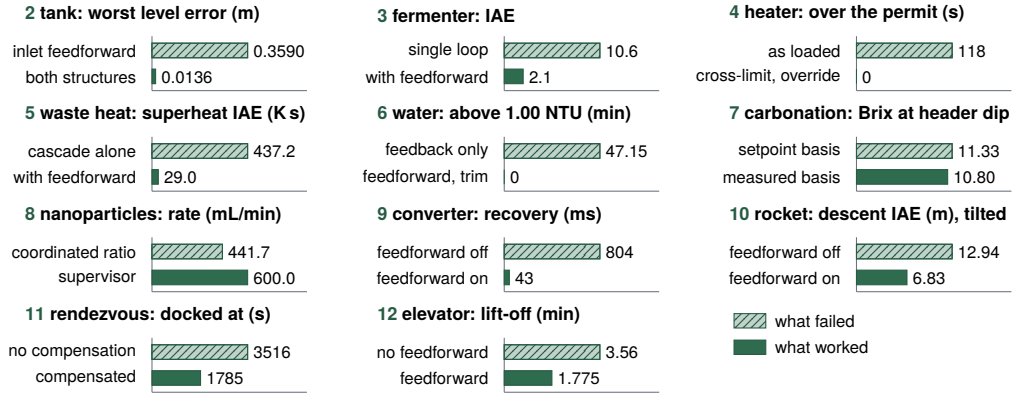


Figure 13.1 One like pair of numbers from each row, as the "Measured result" column of Tables 13.1 and 13.2 prints them. Each panel has its own scale from zero. The hatched bar is the structure or setting that failed and the solid bar the one that worked; a value of zero has a label and no bar. Notice that the carbonation panel is the one where the two bars nearly match: the setpoint basis moved the filler only from 10.80 to 11.33 degrees Brix, and that was enough to put product off specification.

An inner loop or a selector often earns its place by what it limits. The converter's single loop recovered faster than the cascade, 135 against 196 ms, with a smaller first dip. The cascade's value was a current held to 9.47 A in an overload and an outer loop that could be pushed. Cross-limiting on the heater bought no temperature at all. It deepened the outlet dip from 4.6 to 5.8 °C and removed every second of fuel-rich firing. Such structures are judged on the constraint they protect, not on the IAE.

A slow trim corrects what was not measured, slowly, and follows its instrument. The water plant's trim, with a 300 min integral time on a 190 min plant, had not converged six hours later. The heater's oxygen trim followed a drifting analyzer until the burners went rich, and only the carbon monoxide override held the permit. The mRNA skid's size trim made the rate, and the conductivity probe made the quality. In each plant the trim needed its own protection: a clamp, an override, or a second measurement that did not share the first one's fault.

13.2 Choosing a Structure

The chapters suggest an order of questions for a new disturbance. Each names the chapter where its answer decided the result, and Figure 13.2 draws the same order.

1. *Where does the disturbance enter, and which measurement sees it first?* Place it on the map of Section 1.4.2. A lag in the measurement itself is met by no structure (Chapter 3).
2. *Is it measured, and can its effect be computed?* Then feed it forward from a balance or from physics (Chapter 6, Chapter 11). Estimate the model error and ask whether the feedback loop is fast enough to absorb it.
3. *Does the measured signal depend on the manipulated variable?* Then it is not a disturbance measurement. The inlet feedforward of Chapter 2 put a pole at $z = -1$ and reversed the valve on 70.0 % of its moves.
4. *Does the upset enter between the controller output and a faster measurable variable?* Close an inner loop there, and sweep its speed against the outer lags before trusting it (Chapter 3, Chapter 10).
5. *Is one direction of a ratio error hazardous?* Cross-limit, and accept the slower response in the safe direction (Chapter 4).
6. *Is part of the disturbance unmeasured or the model uncertain?* Add a slow multiplying trim, with a clamp and an independent override (Chapter 4, Chapter 7).
7. *Is the plant nonminimum phase or short of actuator authority?* Then a feedforward trades undershoot for speed as a higher gain does (Chapter 9), and a capability shortfall needs a change of target or equipment, not of structure (Chapter 10).

A predictive controller earned its place where it carried a forecast or a hard constraint. Told of a turndown, it held the waste-heat cycle's superheat at 5.80 K where the cascade fell to 2.43 K (Table 5.7). The converter's explicit MPC reached the band 1.9 to 6.8 times sooner than the cascade (Section 9.6.6), and the orbital MPC docked 462.5 s sooner on 14.9 % less propellant (Table 11.8). It lost where its model was wrong. Under fouling the waste-heat MPC never settled, an IAE of 5583 against 335 K s for the cascade, and the landing MPC tipped over with a slow gimbal where the cascade landed (Sections 5.5 and 10.7). In the steady crosswind a bias feedforward added to the cascade cut the offset from 5.771 to 0.0308 m, as small as the MPC's, without a horizon. Volume 9 treats that choice in full.

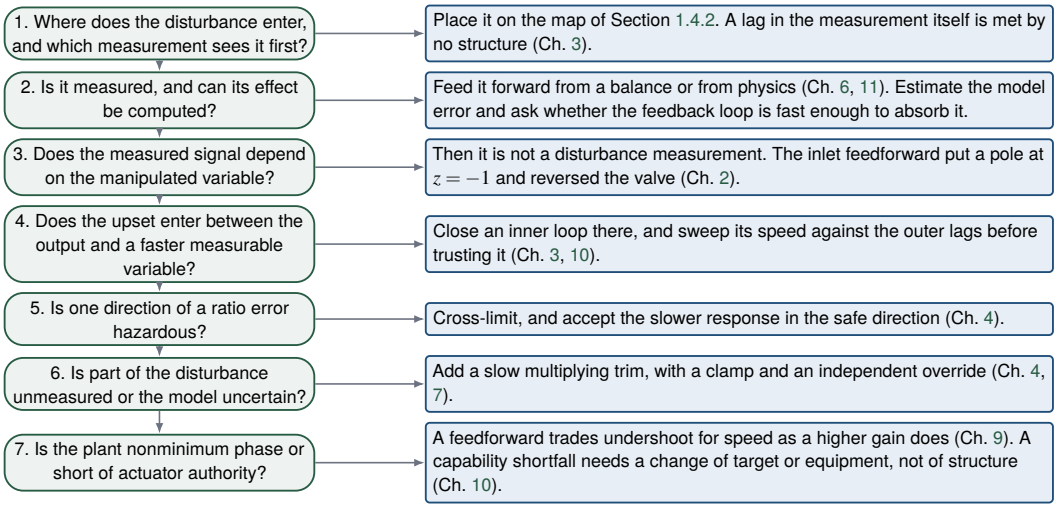


Figure 13.2 The seven questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that the entry point and the measurement come first and the trim comes late: a slow trim corrects only what the structure above it has already placed correctly.

13.3 Where Practice Is Heading

Feedforward is tuned by rule and from data. Simple tuning rules for feedforward compensators have been published for process control [1], along with the optimal performance of low-order compensators and the practical considerations in using them [2]. A 2024 monograph is devoted to feedforward control and closes with a summary of its tuning rules [3]. In motion control, feedforward is tuned iteratively from measured data [4], and because accurate data are inexpensive to collect on such machines, models and controllers are identified and learned from experiments [5]. The orbital and elevator chapters show what that work protects against: a feedforward whose model is slightly wrong at the operating point that matters.

Unmeasured disturbances are estimated and fed forward. Disturbance-observer-based control lumps disturbance and model uncertainty into one total disturbance, estimates it and compensates it, and has been applied across industrial sectors for four decades [6]. Active disturbance rejection control is one member of that family, built around an observer of the total disturbance [7]. In the editor's view this is the natural successor to the slow trim of this volume: the crosswind of Chapter 10, the heating value of Chapter 4 and the organic carbon of Chapter 6 are all disturbances that an estimator could feed forward, provided its estimate is checked as carefully as an analyzer's.

Trims run on inferred quality. Data-driven soft sensors infer a quality from routine process measurements [8]. The mRNA skid of Chapter 8 trims on such an estimate between laboratory results, and it missed one of twelve seeds when neither the estimate nor the last reported size saw the true size leave its window. In the editor's view the trims of the next decade will run more often on inferred values, and the work of the engineer moves toward deciding which second measurement can catch an inference that is wrong.

Inner loops meet optimization. Predictive control has been applied to power converters and drives because it handles constraints and nonlinearities directly, at a high computational cost [9]. Powered-descent guidance has been posed as a convex program, solved by methods with deterministic

convergence, with a view to running it onboard [10]. In the editor's view cascades will remain the inner layer of such systems, as in Chapter 9 and Chapter 10, because they bound a current or an angle with a few lines of code that can be read and verified.

> ⧖ HISTORY **1956 to 2020: From a Cascade to a Hierarchy of Loops**
>
> The timeline of Chapter 1 places the first quantitative analysis of cascade control in 1956. A thesis on hierarchical control traces its field to those first results and calls cascade control a precursor of hierarchical control. The theory of multilayer control that followed in the 1960s rested on aggregation and on multiple time scales [11]. A 2020 review of predictive control in power electronics shows the same division built into hardware. In its example the fast controller runs in a gate array and takes its reference from an outer loop that runs at a lower frequency on a processor [12]. In the editor's reading, the tank of Chapter 2 and the converter of Chapter 9 stand at the two ends of that arc, and the sweeps of Section 13.1 show where the separation of time scales stops holding.

Feedforward is computed from an inverted plant model. On a waste-heat recovery cycle for a heavy-duty diesel engine, a dynamic feedforward obtained by inverting a nonlinear model was added to the commonly used gain-scheduled PID that holds the superheat. On an engine test bed the largest superheat error fell to about 1.9 K, against about 10 K without the feedforward, and the feedforward computation was shown robust enough for an embedded implementation [13]. The cycle of Chapter 5 feeds forward from its source meters at the same point. In the editor's view the step from a static balance to an inverted dynamic model is where feedforward in these plants is heading, and gap 1 of Section 13.5 is the price of that step.

Cascades are closed over networks. Cascade control systems, with at least two nested loops, are widely used in industrial processes, power plants and chemical processes. When the loops are closed through a real-time network, delays and packet loss in both loops can reduce performance or cause instability. Event-triggered execution has been proposed to lighten the load on the network, and cyber-attacks change the closed-loop dynamics [14]. In the editor's view this carries the inner-loop question of Section 13.1 onto the network. The inner gain of Chapter 2 was already capped by the delay of its own sampling, and a network adds delays that vary from one scan to the next.

Vehicles find their own relative state. A survey of small spacecraft technology reports a vision-based payload on a mission launched in 2023 that gave angles-only relative navigation of an object without prior knowledge of it, with a relative position error of 0.5 % of range. The same survey notes that much of this navigation must be done without commands from the ground, that guidance for proximity work is still primarily research based, and that many of these missions had difficulties with their control systems and actuators [15]. The feedforward of Chapter 11 cancels a coupling computed from the relative state and the mean motion, so in the editor's view the quality of that on-board estimate sets how far the feedforward can be trusted. Figure 13.3 sets developments of feedforward and cascade practice in three threads.

13.4 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 [20]. 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.

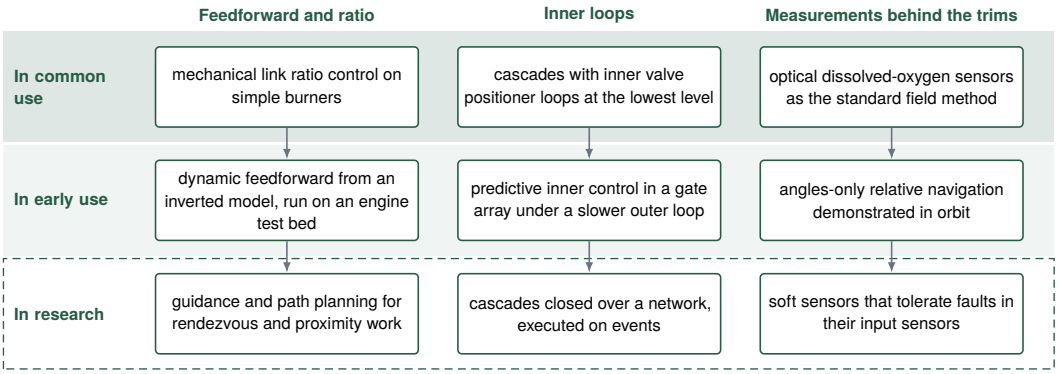


Figure 13.3 Developments in three threads. The top lane holds what the cited source calls standard, customary or common practice, the middle lane what it reports in service at some sites, on a test bed or in orbit, and the bottom lane what it reports from research, as a proposal or as a need, in the editor's reading. Sources by thread: feedforward and ratio [13, 15, 16], inner loops [12, 14, 17] and measurements behind the trims [15, 18, 19]. Notice that each thread moves from a fixed device toward a computation that must be checked while the plant runs.

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 [21]. 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.

13.4.1 How This Volume Was Made

Each of the eleven apps was built and tested by agents. An agent then wrote the paper on it, keeping verification notes that trace each number to a run, and ten of the eleven papers received a referee report, at least four of them a self-review by the agent that wrote the paper. Defects went into a log for each app. For this book one agent per chapter reran every number an exercise uses. The editor set the scope of each round and what each agent could change, and statements that runs contradicted were left for the editor. The record holds the following errors and the checks that caught them.

The defect logs. Ten of the eleven logs give each entry a severity, 102 entries in all. Of these, 32 are wrong behavior of an app (a wrong output, a wrong lesson, or a controller reading what it should not), 46 are documentation, and 19 are cosmetic or display. The water plant's log has no severity field. It records that eight quantitative claims in the app's teaching text were wrong when the model was driven, and that the model was right in every case.

A feedforward that knew too much. The fermenter's carbon-balance feedforward read a hidden flag that told it when a quarter of the sparger had plugged. No instrument reports that, and the challenge credited the feedforward with knowledge no plant has. On the converter, the cascade, its preload and its feedforward read the inductor resistance of the run in place of the declared value. With a changed resistance its duty differed from that of a controller fed the same measurements. Both were found by reading the code for what it reads, both now use what a plant would have, and the converter's fix moved no shipped run.

A teaching number that made a slow inner loop look safe. The fermenter's teaching text placed the cascade's break-even at an inner time constant of 10 s, 0.8 times the oxygen holdup, where the sweep gives 7.5, 0.6 times the holdup. The error ran in the direction that makes a slow inner loop look acceptable. The same text gave the drive lag as eight seconds where the code computes 15.4 s. The water plant's text said its trim closes the organic-carbon offset "completely" in about two hours. In the run it had moved the dose from 31.32 to 33.42 mg/L by 600 min, short of target. Each was caught by driving the shipped model and comparing.

A stability rule that was not a stability condition. The tank app's help text placed the inner flow loop's limit where a product of gains passes one, and called a gain of 1.5 the boundary. Writing the sampled recursion from the code gave the Jury limit of 1.925 at 12 bar, and runs found the loop well behaved at 1.5 and broken at 2.0. The same derivation showed the inlet feedforward sitting exactly on the unit circle at the shipped gain. The chapter now carries both results.

A metric that flattered the tuning. On the converter the overshoot tile read 49.3 % where the logged output peaked at 58.2 % of the step. The tile latched the starting value one sample after the step, when an inverse-response plant had already moved the wrong way. The paper agent found it by checking the tile against the exported trajectory. The rocket's feedforward-off comparison gave two answers, a descent IAE of 6.668 or 12.94 m, depending on whether Reset came before or after the switch. Its chapter reports both paths.

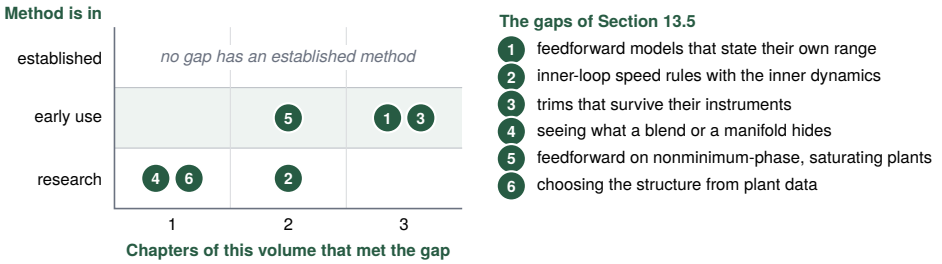


Figure 13.4 The six gaps of this section. Across is the number of chapters that each item names as evidence: three for gap 1 (Chapters 2, 11 and 12), two for gap 2 (3 and 2), three for gap 3 (4, 7 and 8), one for gap 4 (8), two for gap 5 (9 and 10) and one for gap 6 (2). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Notice that the two gaps met by the most chapters, feedforward models that state their range and trims that survive their instruments, have methods in early use, while rules for inner-loop speed, measurement per stream and choosing a structure from data are still research.

An explanation that was wrong. A draft of the heater paper said the trim clamp did nothing in the challenge because the trim had "not yet" reached 8 %. Measured, the drift drove the trim down, to only –0.3 % when carbon monoxide crossed the permit. The self-review caught it by measuring the trim instead of reasoning about it.

The book round. The eleven chapter records list 184 rows of numeric verification, each a stated value beside its reproduction. Nine chapters showed no drift. Two showed one number each: the water plant's storm IAE of 1553.5 now runs to 1555.6, and the elevator's constant-g_0 transfer docked 0.353 m from the mark where the chapter said 0.375. Both were listed for the editor, in a list of six entries for this volume that also held app texts with stale numbers and an orbital sentence implying a compensation the challenge does not need. A later repair round corrected both numbers in the papers and the chapters, and the orbital sentence now reads correctly in the chapter. The agent writing Chapter 1 narrowed two of its own first drafts, "every chapter enters the feedforward as a PID bias" among them, after checking them against the sources and its script.

Every error above was caught by a check that could fail: a run beside a quoted number, a recursion written from the code, a hidden value read where a measurement belonged, a tile set against the trajectory. None was caught by rereading the text that contained it. For feedforward and cascade work the record adds one rule to those of the shared section. Ask of every feedforward what it reads, and confirm that a plant could measure it.

13.5 Gaps in Feedforward, Cascade and Ratio Practice

Six gaps belong to the topic of this volume, and the gaps common to agent-era practice follow in Section 13.5.1. 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 13.4 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. *Feedforward models that state their own range.* The tank's trim was exact at one supply pressure and the orbital terms at one mean motion, and a constant-g_0 weight on the elevator holds only near the ground. A textbook of feedback systems notes that feedforward, because it matches two signals, requires good process models, without which its corrections may have the wrong size or be badly timed [17]. A standard for models and simulations requires a record of each validated model's domain of validation and allows use outside it only within the model's limits [22].

The gap is hard because the operating point that leaves the model's range is often the one the feedforward exists for, as at the geostationary altitude. A feedforward that reports when the operating point has left the region where its model holds, and backs off, is not standard practice.

2. *Inner-loop speed rules that include the inner loop's own dynamics.* The fermenter's break-even moved from 7.4 to 13.4 s with the inner reset time alone, and the tank's inner limit moved within each pressure cycle. An open textbook states as non-negotiable that the inner loop respond faster than the outer, and that cascade control cannot function if the relation is reversed, but gives no ratio [16]. The gap is hard because the inner integral, the sample period and a moving operating point each change the closed inner loop that the outer design assumes. A design rule that accounts for all three would replace a sweep.

3. *Trims that survive their instruments.* The heater's trim followed a drifting analyzer, the carbonation analyzer read dissolved air as carbon dioxide, and the mRNA soft sensor and the last reported size both stayed in the window while the true size left it. Clamps and overrides were set by hand in each chapter. The Open challenge box after this list takes the gap further.

4. *Seeing what a blend or a manifold hides.* In Chapter 8 13 500 mL of the 15 409 mL a passing run accepted was made while a channel was outside its own window, and one flow meter read its setpoint throughout. An industrial review of lipid nanoparticle manufacturing notes that scaling out by parallel mixing lines risks mixer-to-mixer variability, with inconsistent flow rates, pressure imbalance or channel clogging, often leading to non-uniform product [23]. The gap is hard because a meter on the shared line sees only the sum, and a meter on every stream multiplies the instruments as the streams multiply. Ratio control of many parallel streams needs measurements or estimates per stream.

5. *Feedforward on nonminimum-phase and saturating plants.* The converter's feedforward deepened the dip from 5.69 to 7.82 V, and the rocket's engine could not supply the deceleration the profile asked for. A review of predictive control in power electronics names solution in real time, within a few microseconds, as the main challenge, and notes that converter models are often linearized at a loss of accuracy [12]. The gap is hard because the undershoot belongs to the plant, so a faster correction buys recovery only with a deeper first dip. Design methods that respect undershoot and actuator authority from the start are thinner than those for a minimum-phase loop.

6. *Choosing the structure from plant data.* This volume chose each structure by reading the disturbance path from a model. The tank of Chapter 2 offered two flow meters, and the one its own valve moves made a feedforward that reversed the valve on 70.0 % of its moves. A survey of industrial predictive control listed which variables should be used for control, and how to design a control system for an entire plant, among the needs industry perceived [24]. The gap is hard because routine records hold the plant under its existing structure, so the path a disturbance would take under another structure is seldom in the data. A method that proposes the feedforward signal and the inner variable from plant records, and says which entry point the records cannot resolve, would turn the path map of Table 1.3 into a tool.

⚠ OPEN CHALLENGE A Trim That Knows When Its Instrument Is Wrong

Three chapters met gap 3: an oxygen analyzer that drifted, a carbon dioxide analyzer that counted dissolved air and a size estimate that missed its window. A critical review of soft sensors for bioprocesses names erroneous inputs from sensor faults among the greatest challenges, finds that tolerance to sensor faults remains the greatest, and notes that a departure of a prediction from the hardware reading indicates a sensor fault [19]. In the editor's view the gap is hard because a drifting instrument still reports a plausible value, so nothing marks the reading the trim should stop following. Progress would be a trim that carries a second, independent measurement and a stated bound on the harm a failed instrument can cause.

13.5.1 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 [25]. 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 [21]. 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 [26]. 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 [27]. 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 [28]. 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

- A structure helped only against a disturbance that it saw earlier than the controlled variable did.
- A feedforward was only as good as its model, and the speed an inner loop needed was found by sweeping.
- Inner loops and selectors earned their place by the limits they kept, and slow trims needed guarded instruments.

References

[1] J. L. Guzmán and T. Hägglund, "Simple tuning rules for feedforward compensators," *Journal of Process Control*, vol. 21, no. 1, pp. 92–102, 2011.

[2] M. Hast and T. Hägglund, "Low-order feedforward controllers: Optimal performance and practical considerations," *Journal of Process Control*, vol. 24, no. 9, pp. 1462–1471, 2014.

[3] J. L. Guzmán and T. Hägglund, *Feedforward Control*. Berlin: De Gruyter, 2024.

[4] F. Boeren, T. Oomen, and M. Steinbuch, "Iterative motion feedforward tuning: A data-driven approach based on instrumental variable identification," *Control Engineering Practice*, vol. 37, pp. 11–19, 2015.

[5] T. Oomen, "Advanced motion control for precision mechatronics: Control, identification, and learning of complex systems," *IEEJ Journal of Industry Applications*, vol. 7, no. 2, pp. 127–140, 2018.

[6] W.-H. Chen, J. Yang, L. Guo, and S. Li, "Disturbance-observer-based control and related methods: An overview," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1083–1095, 2016.

[7] J. Han, "From PID to active disturbance rejection control," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 3, pp. 900–906, 2009.

[8] 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.

[9] S. Vazquez, J. I. Leon, L. G. Franquelo, J. Rodriguez, H. A. Young, A. Marquez, and P. Zanchetta, "Model predictive control: A review of its applications in power electronics," *IEEE Industrial Electronics Magazine*, vol. 8, no. 1, pp. 16–31, 2014.

[10] B. Açıkmeşe and S. R. Ploen, "Convex programming approach to powered descent guidance for Mars landing," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 5, pp. 1353–1366, 2007.

[11] D. Gromov, "Analysis of hierarchical structures for hybrid control systems," Ph.D. dissertation, Technische Universität Berlin, Berlin, 2010.

[12] P. Karamanakos, E. Liegmann, T. Geyer, and R. Kennel, "Model predictive control of power electronic systems: Methods, results, and challenges," *IEEE Open Journal of Industry Applications*, vol. 1, pp. 95–114, 2020.

[13] J. Peralez, M. Nadri, P. Dufour, P. Tona, and A. Sciarretta, "Organic Rankine cycle for vehicles: Control design and experimental results," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 952–965, 2017.

[14] M. Fallahnejad, "A survey of recent breakthroughs in networked cascade control systems," *International Journal of Robust and Nonlinear Control*, vol. 36, no. 5, pp. 2531–2541, 2026.

[15] S. V. Weston, C. D. Burkhard, J. M. Stupl, R. L. Ticknor, B. D. Yost, R. A. Austin, P. Galchenko, L. K. Newman, and L. Santos Soto, "State-of-the-art small spacecraft technology," National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA, Technical Publication NASA/TP-20250000142, Jan. 2025. [Online]. Available: https://ntrs.nasa.gov/citations/20250000142

[16] 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.

[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] U.S. Geological Survey, "Dissolved oxygen," U.S. Geological Survey, Techniques and Methods, book 9, chap. A6.2, 2020.

[19] V. Brunner, M. Siegl, D. Geier, and T. Becker, "Challenges in the development of soft sensors for bioprocesses: A critical review," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 722202, 2021.

[20] 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.

[21] 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.

[22] National Aeronautics and Space Administration, "Standard for models and simulations," Office of the NASA Chief Engineer, Washington, DC, NASA Technical Standard NASA-STD-7009B, 2024, approved 2024-03-05, standards.nasa.gov.

[23] J. Hong Hoang, M. Ott, E. Samaridou, M. Beck-Broichsitter, and J. Simon, “Industrial perspective on the manufacturing of lipid nanoparticles for nucleic acid delivery,” *Pharmaceutics*, vol. 18, no. 4, p. 489, 2026.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] L. Bainbridge, “Ironies of automation,” *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

13.6 Exercises

13.1 Discussion · 8 min

An agent proposes a feedforward for a level loop on a tank fed through a control valve. Its signal is the inlet flow transmitter, because “the inlet flow is the largest disturbance on the balance”. Using Chapter 2, explain what will happen, list the checks you would ask for before the feedforward is enabled, and say which structure the agent should have proposed for each of the tank’s two upsets.

13.2 Discussion · 8 min

A design note says “the inner loop of a cascade must be three to five times faster than the outer loop”. Using Chapter 3, Chapter 9 and Chapter 10, argue why the note is a starting point and not a design rule, and describe the measurement you would make on a new cascade instead.

13.3 Discussion · 8 min

The heater of Chapter 4 trims its air ratio on stack oxygen, the water plant of Chapter 6 trims its dose on settled turbidity, and the skid of Chapter 8 trims its buffer on a soft sensor of particle size. For each, say what the trim corrects that the feedforward or ratio cannot, what can make the trim do harm, and what protects against that harm in the chapter.

13.4 Discussion · 7 min

An engineer must choose between spending a month improving the model behind a feedforward and spending the same month making the feedback loop faster. Using Chapter 11 and Chapter 12, argue which choice pays in which circumstances.

13.5 Discussion · 10 min

Section 13.4.1 describes errors that the agents building this volume made. For the feedforward that read a hidden fault flag, the teaching number that made a slow inner loop look safe, the stability rule that was not a stability condition, and the overshoot tile, name the check that caught each. Then write a procedure, in five to eight steps, for an agent that will add a feedforward to a running loop on a plant, stating what the agent may read, what it may write and what a person must approve.

13.6 Discussion · 8 min

A vendor proposes to replace the cascades of a waste-heat power cycle and a rocket lander with model predictive controllers. Using Chapter 5 and Chapter 10, write the questions you would ask before agreeing, and say what evidence would change your mind in each direction.

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 (b). The header pressure enters at the valve, upstream of the steam flow meter, so an inner flow loop removes it within seconds.

1.3 6 PV units times minutes, with the sign of the uncompensated disturbance. The feedforward is late.

1.5 (a) 12, then 10 % per kg/s. (b) The added bias is −10 %. At 8 kg/s it fires 86 % against the 80 % needed, 3.75 °C too hot.

1.7 (a). The outlet demand enters outside the inner loop, so the inner flow measurement does not see it.

1.9 (d). The inlet flow responds to the valve, so the term closes an algebraic loop from the valve to itself.

Chapter 2. Matching Loop Structure to Disturbance Entry

2.1 (c). The demand enters at the vessel, outside the inner loop, so only the slow outer loop can answer it.

2.3 (a). The pole is at $-K_{\text{ff}}K_v = -1.000$ and the static gain of the recursion is $1/(1+K_{\text{ff}}K_v) = 0.5$.

2.5 (a) 0.762 and −0.455. (b) $K < 2/(3K_v) = 1.925$, and $K = 1.5$ is unstable above 19.75 bar. (c) About 13.7 s per cycle.

2.7 (a) 0.0214 m and 0.0047 m, a ratio of about 4.6 against the linear 3.7. (b) Feedback holds the lift within 21.6 to 22.1 % and lets the inflow swing from 3.06 to 10.33 kg/s. The cascade moves the lift from 15.1 to 42.0 % and holds the inflow within 5.26 to 8.86 kg/s.

2.9 (a) 0.861 m, within 0.002 m of the setpoint, and 1.280 m. (b) $0.501 \times 0.2775 = 0.139$ m low and $0.999 \times 0.2775 = 0.277$ m high.

2.11 (a) The leak slope is $c/(2\sqrt{L})$, 2.5 kg/s per m at 1 m, so $\tau = 2000$ s and $K_p = 0.139$ m per percent. At 2 m the slope is 1.77 and $\tau = 2828$ s. (b) No. It stays $1/K_v = 2.887$ at every level.

Chapter 3. Fermenter Oxygen Cascade and Inner Loop Speed

3.1 (c). DIC-301 writes a speed demand, which is the setpoint of the speed loop SIC-301.

3.3 (a). The feedforward uses what the plant measures. The biomass is a state, not a measurement.

3.5 (a) 15.4 s. (b) $K_{p,N} = 0.736$ rev/min per percent, $\tau_{inner} = 2.85$ s and 12.3 s. (c) About 1.47. The equation counts the proportional gain and not the inner integral time.

3.7 (a) 23.8, 11.83 %, 25 rev/min, and 12.0, 5.18 %, 95 rev/min. (b) 18.2, 9.34 %, 18 rev/min, and 3.5, 4.82 %, 23 rev/min. (c) At the shipped gain the outer loop is slow and the drive lag is a small part of its response. At the high gain the single loop is limited by the drive lag, which the cascade has removed.

3.9 (a) 5.9, −2.90 % and +1.75 %, 18 rev/min. (b) 15.1, +22.86 % at 4.25 min, 38 rev/min. (c) 18.2, the plain cascade. The lag, not the gain, decides whether the feedforward helps.

3.11 (a) 14 %. The feed pump sets the growth rate, and oxygen limits growth only near 3.5 % true dissolved oxygen. (b) A second probe, an off-gas balance that would fail to close, or a calibration check. A tighter controller holds the wrong value more exactly.

Chapter 4. Fired Heater Cross-Limiting and Oxygen Trim

4.1 (b). The low selector of Equation (4.17) picks $(1+\delta)F_{a,m}/R$ while the fan is still below the new demand.

4.3 (d). The trim starts the drift near +5.3 % and never comes near −8 %, so the clamp never acts.

4.5 (a) $\lambda_b = 1.15/1.06 = 1.085$. (b) $\rho = +6.0$ %, within the default ±8 %. (c) About 196 ppm at the burners, 190 ppm at the stack and 2.0 % oxygen.

4.7 (a) 3.7 s rich, 7268 ppm and 0.990 with fuel leading. None, 50 ppm and 1.136 cross-limited. (b) 4.6, 5.8 and 6.3 °C. The more the fuel waits for the fan, the later the heat arrives and the deeper the dip.

4.9 (a) 3.7 s, 118 s, 7566 ppm, 0.989. (b) 0.0 s, 101 s, 295 ppm, 1.073. (c) 0.0 s, 0 s, 173 ppm, 1.088 and 80.70 %, a pass. Cross-limiting answers the load step, and the override answers the drifting analyzer.

4.11 (a) 0.45, 0.28 and 0.20 kg/s. (b) −25 Pa. A positive firebox pushes flame and hot gas out of every opening, so the margin is set by the worst gust and not by the average, and the spread over seeds shows the counts themselves are not reliable.

Chapter 5. Organic Rankine Superheat Control

5.1 (a). The override shares the thermowell-lagged measurement with the superheat controller, so it learns of the loss as late as the loop does.

5.3 (c). The loop is out of authority, and fan power falls with the cube of the flow, so a higher setpoint nets more.

5.5 (a) 245.6 kJ/kg. (b) 716.4 kW and 2.917 kg/s. (c) 62.5 and 46.7 %, 1.28 % per K.

5.7 (a) 1.10 K at 141 s, about 83 s isolated, tile $0 \cdot 1$. (b) 9.41 to 11.07 K, tile $0 \cdot 0$. (c) 6.77 against 8.40 kWh. The feedback run loses the generator while the expander is out.

5.9 (a) 5.80 K at 134 s. (b) 2.45 K at 149 s, about 21 s below 5 K. (c) 5.11 K. The forecast is worth about 0.7 K and the structure about 2.7 K.

5.11 (a) 12.19 % and 10.06 %, 2.56 %. (b) No setting nets positive power at 20 °C cooling water, the best being −0.59 kW. Stop the cycle, or use the heat directly.

Chapter 6. Coagulant Feedforward in Water Treatment

6.1 (d). The dose ratio is $140/36.36 = 3.85$, on the far side of the optimum of Equation (6.1), where 1.1 % of the particles are destabilized.

6.3 (b). A sink speeds the tank up, so the controller would see a much shorter lag than the three-hour transport response.

6.5 (a) 184.5 min. $\theta \approx 32.25$ min and $\tau \approx 161.35$ min. (b) $+0.184$ and -0.262 NTU per mg/L. (c) -1.05 below the optimum and $+0.74$ above it, where the loop is positive feedback.

6.7 (a) 141.9 min, 20.92 NTU, and 32.2 mg/L against 52.73 mg/L. (b) 0.0 min, 3.60 NTU, and 52.6 mg/L against 52.73 mg/L. (c) 414.5 kg and 306.0 kg.

6.9 (a) 2.9 min, 8 washes, 215 m^3 and 12 193 m^3. (b) 0.0 min, 8 washes, 518 m^3 and 12 172 m^3, or 4.3 % against 1.8 %. (c) 0.59 NTU and 0.17 NTU. The log keeps one row every 0.2 min and misses the top of the spike.

6.11 (a) $415 \times 0.32 = \$132.8$ and $306 \times 0.32 = \$97.9$, a saving of about \$35 per storm. (b) About 86 storms. The case rests on the 141.9 min off specification and 47.15 min of reportable violation that the feedforward removes.

Chapter 7. Soft Drink Carbonation and the Foam Limit

7.1 (d). The analyzer measures a content, and Henry's law gives the same content a higher CO_2 pressure at 10.5 °C than at 3 °C, so the margin falls.

7.3 (b). The 0.32 m^3 made while the blend ramps down raises the level and compresses the gas above it, and the pressure passes 3.3 bar(g).

7.5 (a) 0.2195 and 10.44 °Bx. (b) $R = 0.2288$, $\Delta R = +0.0093$, 4.2 % more syrup. (c) 149.2 kg/h and 147.1 kg/h.

7.7 (a) 2.29 m^3 in and 4.37 m^3 off specification, foam index 70, the limit passed at 413 s, the tank at 10.5 °C with 0.48 bar of margin. (b) All 6.67 m^3 in specification, foam index 40, the tank again at 10.5 °C but with 0.95 bar. (c) The shipped tank floats at 3.02 bar(g), inside its make-up and vent gap, so the setpoints differ by 0.48 bar at the tank.

7.9 (a) 10.98 m^3 in and 0.82 m^3 off, 10.6 kg vented, foam index 58, level 70.7 %. The in-specification, off-specification, vent and foam bounds fail. (b) 11.80 m^3 in, none off, 0.8 kg, 39, 50.6 %, 14.1 kWh per m^3, target met. (c) 3.16 m^3 would foam. The feedforward and stop handling keep the level and the CO_2, and the top pressure keeps the margin.

7.11 (a) $6.67 - 4.60 = 2.07\,m^3$, so the one-time 2.4 kg is about 1.2 kg per m^3 saved. (b) No. The foam index would be 55, over the limit. The cut has to take the rest, or the top pressure has to rise again if the tank and filler are rated for it.

Chapter 8. Timing the Quench in Nanoparticle Formulation

8.1 (c). At a constant flow the age is V_{loop}/TFR, so the total flow is the clock as well as the throughput.

8.3 (a). The two extra feed pumps exist so the buffer strength is free to move on its own.

8.5 (a) 97.49 nm and 1.0474 nm/ms^{p}. (b) 117.0 and 440.0 ms, giving 140.2 nm and 217.2 nm against the 139.9 nm and 215.8 nm the plant makes. (c) Only the growth term grows: it rises from 42.7 to 119.7 nm, a factor of 2.80, while the assembly size does not move.

8.7 (a) 139.9 nm at 117 ms and 215.8 nm at 440 ms, so 3.76 times the age gives 1.54 times the size. (b) 215.5 nm against 216.0 nm: three and a half volumes of quench instead of two move the size by about half a nanometer. (c) The discharge pressure rises from 2.49 to 4.10 bar, and the extra flow is presented to the buffer exchange.

8.9 (a) 126.6, 163.6, 132.9 and 150.8 nm, blending to 142.8 nm at a spread of 0.364. (b) Yes: the blend is 4.2 nm from the target and the spread is under the limit, so a light-scattering measurement of the tank would pass the material. (c) The pressure rises from 2.49 to 4.29 bar, and the channels separate further, to 125.3 and 190.1 nm.

8.11 (a) Every window, limit and delay is a teaching value rather than a specification, and the size law is an interpolation inside one box with teaching factors carrying it outside. (b) The blend is what leaves the valve and reaches the tank, and a measurement of that tank would report the blend's mean and width. The per-channel figure is reported beside it as teaching truth.

Chapter 9. Boost Converter Cascade and Inverse Response

9.1 (b). In $C\,\mathrm{d}v/\mathrm{d}t = q\,i_L - v/R$ the factor q falls with the duty while the state i_L has not moved, so the capacitor loses $i_L\,\Delta d$ of charging current.

9.3 (d). The cascade does not shrink the first dip. It regulates and limits i_L, tolerates $K_{c,v} = 0.45\,\mathrm{A/V}$ with a 58.6 % excursion, and gives the load-current feedforward a loop to enter.

9.5 (a) 134.8 and 139.8 rad/s. (b) $\omega_n = 77.68\,\mathrm{rad/s}$, $\zeta = 0.300$, $z/\omega_n = 1.74$. (c) $z = 34.4\,\mathrm{rad/s}$ at $4\,\Omega$. The bound falls from 67 to 17.2 rad/s.

9.7 (a) 23.28 V at 0.208 s, back through 24 V at 0.216 s. (b) 29.41 V against 31.35 V requested, and 748 to 1413 mJ. (c) The capacitor loses 33 mJ while the inductor gains 66 mJ, so the withheld charge is stored as inductor current.

9.9 (a) 34 ms, 0.43 V, 7.58 A, two clauses fail. (b) 24 ms, 0.21 V, 7.44 A, met. At 0.35 A/V, 16 ms and 0.28 V but 8.22 A. (c) 16 ms, 0.00 V and 8.09 A, so the MPC fails on current.

9.11 (a) About 2 W more loss, so $48/51.72 \approx 92.8$ %. (b) Whenever stored energy is leaving the converter, for example after a load release. It is arithmetic, not a fault.

Chapter 10. Rocket Landing by Commanding an Attitude

10.1 (a). The vehicle has no lateral actuator, so the position loop can only ask for a tilt and wait for the attitude loop to produce it.

10.3 (c). Two proportional-derivative loops in series can hold a constant attitude only by holding a constant error.

10.5 (a) $\sqrt{2/a} = 0.505\,\text{s}$, $\sqrt{6/a} = 0.874\,\text{s}$ and $\sqrt{12/a} = 1.237\,\text{s}$. (b) $1.5g\delta/a = 0.164\,\text{m}$. (c) 0.540 s. No: the times belong to the open-loop plant and depend on a alone.

10.7 (a) 32.45 s and 0.40 m/s, corner at $(10 - 0.5)/0.35 = 27.1\,\text{m}$. (b) 39.32 s and 0.47 m/s, landed. 25.38 s and 3.29 m/s, a hard landing although the throttle peaks at 96 %. (c) A hard landing at 10.00 m/s after $250/10 = 25.0\,\text{s}$.

10.9 (a) 5.77 m downwind. (b) −1.00°, −3.53°, −4.37° and 5.82 m against predicted −1.01°, −3.54°, −4.38° and 6.09 m. The rate term $K_{Dx}v_x/K_{Px}$ accounts for 0.24 m of the 0.27 m gap. (c) 2.96 m, a ratio of 1.95.

10.11 (a) $a = 8.58\,\text{s}^{-2}$, $\omega_{n\theta} = 3.21\,\text{rad/s}$ and $\sqrt{a} = 2.93\,\text{rad/s}$, so the attitude loop and the zero move up as the tanks empty, where the model has them move down. (b) The planar rigid body. Slosh in a part-full tank is a lightly damped mode close in frequency to the attitude loop.

Chapter 11. Orbital Rendezvous With Coriolis Feedforward

11.1 (a). At the start $x = 0$ and $\dot{x} = 0$, so the only nonzero term in the radial equation is $2n\dot{y}_0 = 1.13\,\text{mm/s}^2$.

11.3 (c). The coupling sits inside the loop, so gain attenuates its effect but never removes it. The along-track axis is still pulled 2.523 m off a setpoint it started on, 340 times the compensated value.

11.5 (a) 42.7 % and 3.84 %, gone at $\sqrt{3}n = 1.9596\,\text{mrad/s}$. (b) $n = 1.1834\,\text{mrad/s}$, $\hat{n} = 0.9962\,\text{mrad/s}$, $1.224 \times 10^{-6}/\text{s}^2$, $3.743 \times 10^{-4}/\text{s}$. (c) 30.6 % of K_P and 9.4 % of K_D at 2, 7.6 % and 4.7 % at 4. K_P grows as ω_n^2 and K_D as ω_n.

11.7 (a) 0.529 m/s, 113 888 m s, 21 896 m s, 11.475 m. (b) 0.419 m/s, 66 536 m s, 8.17 m s, 0.0074 m, docked at 1785 s. (c) The term $-2n\dot{x}$ turns the radial descent into an along-track push, and $3n^2x$ removes 42.7 % of the radial stiffness.

11.9 (a) Fails: not docked, 87.86 m, 0.465 m/s. (b) Fails: docked at 998 s but 14.96 m off in the judged window. (c) Meets it: 549 s, 0.13 m, 0.583 m/s. (d) Meets it: 495 s, 0.03 m, 0.575 m/s. Bandwidth and damping are needed. The compensation is not, but it docks sooner, holds four times closer and spends a little less.

11.11 (a) a_d/K_P: 0.010 m and 0.111 m. (b) $a_dT = 5.6\,\text{mm/s}$ per orbit and 3.6 mm/s per hour. Accept the offset, add an integral term, or feed the estimated drag forward. At a 20 m hold, accept it.

Chapter 12. Space Elevator Climber: Gravity Feedforward

12.1 (d). Where g_{eff} is almost zero, the constant g_0 is a false weight of mg_0, which the drive applies outward at once.

12.3 (b). In Equation (12.13) the damping term is $(T/c)/(2\sqrt{k_L m})$, which grows as the string below softens.

12.5 (a) $k_L = 1.450\,\text{N/m}$ and $T_L = 9.12\,\text{min}$. (b) $\zeta_L = 0.218$, against 0.212, and a period 2.6 % longer than 8.888 min. (c) $2m\omega v = 86.6\,\text{N}$ and a lean of 56.7 m to the west.

12.7 (a) 46.7 kN, the weight at $9.347\,\text{m/s}^2$ with no drag left, and $0.311\,\text{m/s}^2$. (b) Lift-off near 3.6 min instead of 1.8 min, and a worst acceleration of $1.048\,\text{m/s}^2$, more than three times the value with the feedforward.

12.9 (a) Guide 8.15 kN, docked sway 27.5 m, docked at 61.0 min: fails on both. (b) The schedule alone gives 5.10 kN but 27.5 m, and ZV alone gives 0.2 m but 8.12 kN. (c) Passes with 4.94 kN, 0.4 m and 0.74 m from the mark, 14.1 min later than the shipped run.

12.11 (a) $11000 \times 9.464 \times 40 = 4.2\,\text{MW}$, about twice the 2 MW of regeneration, so the friction brake takes the rest. (b) $1421/2937 = 48\,\%$ and $2314/2937 = 79\,\%$, and the schedule costs 27 min (72.5 min against 45.5 min).

Chapter 13. Summary and Outlook

13.1 A good answer notes that the inlet flow is moved by the valve the controller drives, so it is not a disturbance measurement. On the tank this feedforward made an algebraic loop with its pole at $z = -1$ at the shipped gain, reversed the valve on 70 % of its moves, used about 158 times the valve travel of feedback, and gave a worse worst error (0.3590 against 0.2989 m). The checks: whether the signal depends on the manipulated variable, a derivation of the loop the feedforward closes, and a test reporting

travel and reversals as well as error. The right structures are a flow cascade for the supply pressure and a feedforward from the outlet flow for the demand, which together reached 0.0136 m.

13.3 A good answer says that each trim corrects an unmeasured or modeled quantity: the heating value of the fuel gas, the organic carbon the dose model takes from the morning's laboratory result, the strength of a concentrate lot. Harm comes from the instrument: a drifting oxygen analyzer made the heater fuel-rich within half an hour at ± 25 % authority. A trim slower than the plant leaves an offset for hours. A soft sensor that, like the last reported size, missed the true size leaving its window let 295.2 mL be accepted at one seed. Protection: a clamp and a carbon monoxide override on the heater, an honest statement of the trim's speed on the water plant, and the conductivity probe and window guard on the skid.

13.5 A good answer pairs the hidden flag with reading the code for what the feedforward reads, the break-even number with driving the model and comparing with the sweep, the stability rule with writing the sampled recursion from the code and testing gains on each side, and the tile with comparing it against the exported trajectory. The procedure gives read-only access to the historian and the configuration, and asks for the candidate signal's source with proof that it does not depend on the manipulated variable. It asks for a model of the disturbance path from data or a balance with its valid range, and an offline test on recorded upsets reporting error, travel and reversals. A person approves through management of change. The feedforward is enabled at a reduced gain with a stated rollback, and the result is compared with the baseline and logged.

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.

actuator saturation The state in which an actuator sits at a limit, so a larger command moves the process no further. *Chapter 1*.

air ratio The air supplied to a burner divided by the air its fuel needs to burn completely. *Chapter 1*.

alkalinity The capacity of water to neutralize acid, which coagulation consumes and which sets how far the pH falls. *Chapter 6*.

analyzer An instrument that measures a composition or a quality, such as oxygen, turbidity or particle size, often slowly or by sample. *Chapter 1*.

anti-windup Logic that stops a controller's integral term from growing while the actuator is held at a limit. *Chapter 1*.

back-calculation An anti-windup method that feeds the difference between the requested and the delivered output back into the integral term. *Chapter 1*.

bandwidth separation Keeping an inner loop several times faster than the loop around it, so the outer loop can treat it as nearly instant. *Chapter 10*.

boost converter A switching circuit that raises a direct-current voltage by storing energy in an inductor and releasing it to the output. *Chapter 9*.

bumpless transfer A switch of mode or structure that does not jolt the actuator. *Chapter 1*.

burner management system A separate safety system that watches a burner and shuts off the fuel on a dangerous condition such as loss of flame. *Chapter 4*.

captive flow The flow in a ratio that is manipulated so that it follows the measured wild flow. *Chapter 1*.

carbonation Dissolving carbon dioxide into a drink under pressure, in an amount that rises with pressure and falls with temperature. *Chapter 7*.

cascade control Two loops in series in which an outer controller sets the setpoint of an inner, faster loop. *Chapter 1*.

Clohessy-Wiltshire equations Linear equations for the motion of one spacecraft relative to another in a circular orbit. *Chapter 11*.

coagulation Adding a chemical to raw water so that fine particles clump together and can settle or be filtered out. *Chapter 6*.

control layers The levels of a control system, from field devices and inner loops up to the outer loops, trims and supervisors. *Chapter 1*.

control structure The arrangement of measurements, controllers and actuators around a plant, chosen before any tuning. *Chapter 1*.

control valve A valve moved by an actuator so that a controller can set the flow through it. *Chapter 1*.

Coriolis meter A flow meter that vibrates its tubes and senses the inertial force of the fluid, so it reads mass flow directly. *Chapter 1*.

cross-limiting A pair of selectors on fuel and air setpoints that keeps the fuel from leading the air up or lagging it down. *Chapter 1*.

current limit A cap on the current a controller may request, enforced by the inner current loop of a converter. *Chapter 9*.

dead time An interval after a change during which the output shows nothing of it, as when fluid travels down a pipe. *Chapter 1*.

deaerator A vessel that strips dissolved air from water under vacuum before the water is carbonated. *Chapter 7*.

dissolved oxygen The oxygen held in a liquid, which cells in a fermenter must have to grow. *Chapter 3*.

distributed control system A plant control system of many controllers and stations on a network, abbreviated DCS. *Chapter 1*.

disturbance An outside change that moves the controlled variable and that the controller does not set. *Chapter 1*.

disturbance entry point The place in a plant where a disturbance comes in, which decides which measurement can see it first. *Chapter 1*.

duty cycle The fraction of each switching period during which a converter's power switch is on. *Chapter 9*.

effective gravity The net pull on a climber along a space elevator, gravity less the outward effect of rotation, which reverses above geostationary height. *Chapter 12*.

excess air Air supplied to a burner beyond what its fuel needs, which wastes heat up the stack but guards against unburned fuel. *Chapter 1*.

external reset Feeding a controller's integral term from the signal actually used downstream, so a

controller not selected does not wind up. *Chapter 4*.

fed-batch fermentation A culture that is fed during the run but not drained, so its volume grows until harvest. *Chapter 3*.

feedback trim A slow feedback controller that corrects the factor of a ratio or feedforward from a measured result. *Chapter 1*.

feedforward control Control action computed from a measured disturbance or a known force before it has moved the controlled variable. *Chapter 1*.

final control element The device that a controller moves to change the process, such as a valve, a pump, a drive or a thruster. *Chapter 1*.

flow transmitter An instrument that measures a flow and sends it as a signal to a controller. *Chapter 1*.

foam index A score in the carbonation app for how close the product comes to foaming at the filler. *Chapter 7*.

gain scheduling Changing a controller's settings with the operating point, so the loop behaves alike where the plant's gain differs. *Chapter 10*.

gravity feedforward A bias that gives the actuator the force needed to hold a vehicle's weight before any error appears. *Chapter 10*.

Henry's law The rule that the amount of gas a liquid dissolves is proportional to the gas's pressure above it. *Chapter 7*.

ideal compensator The feedforward that would cancel a measured disturbance exactly, the disturbance path divided by the process path with a minus sign. *Chapter 1*.

inner loop The fast secondary loop of a cascade, which takes its setpoint from the outer loop. *Chapter 1*.

inner-loop speed rule The rule that an inner loop must be substantially faster than its outer loop for a cascade to pay. *Chapter 1*.

input shaping Splitting a command into timed steps so that the vibration excited by one step cancels that of another. *Chapter 12*.

integrated absolute error The area between the controlled variable and its setpoint over a run, abbreviated IAE, a common score of a loop. *Chapter 1*.

integrator windup Growth of a controller's integral term while the actuator sits at a limit, which later causes a large overshoot. *Chapter 1*.

inverse response A response that first moves the wrong way before it settles in the right direction. *Chapter 9*.

Kalman filter An estimator that blends a model's prediction with noisy measurements, weighting each by its expected error. *Chapter 5*.

lead–lag A dynamic element with one zero and one pole, used to make a feedforward act faster or slower than a pure gain. *Chapter 1*.

lipid nanoparticles Tiny fat-based particles that carry a fragile cargo such as messenger RNA, formed by fast mixing of two streams. *Chapter 8*.

measurement lag The delay with which a sensor follows the true value, as when a probe or thermowell must warm or soak. *Chapter 3*.

model predictive control Control that uses a model to predict the plant over a horizon and plans moves that respect limits, abbreviated MPC. *Chapter 1*.

multiplying trim A trim that multiplies a ratio rather than adding to it, so its correction scales with the wild flow. *Chapter 1*.

nephelometric turbidity unit The unit of turbidity read from the light that a water sample scatters at right angles, abbreviated NTU. *Chapter 6*.

nonminimum phase Describes a plant whose response starts the wrong way or lags more than its gain suggests, which limits how fast feedback can act. *Chapter 9*.

organic Rankine cycle A heat engine that boils an organic fluid instead of water, so it can turn low-temperature waste heat into power. *Chapter 5*.

outer loop The primary loop of a cascade, which holds the main controlled variable by setting the inner loop's setpoint. *Chapter 1*.

override control A selector that lets a protective controller take over an actuator when a limit is near. *Chapter 4*.

oxygen trim A slow loop that corrects a burner's air-to-fuel ratio from the oxygen measured in the flue gas. *Chapter 4*.

PI controller A controller with proportional and integral action, the form most loops of this volume use. *Chapter 1*.

positioner A small controller on a valve that moves the stem until its position matches the signal it receives. *Chapter 1*.

programmable logic controller A rugged industrial computer that runs logic and control loops for a machine or skid, abbreviated PLC. *Chapter 1*.

pulse-width modulation Switching a device fully on and off at a fixed rate and setting the average by the fraction of time on. *Chapter 9*.

quench A stream added right after mixing to dilute the solvent, which stops particles from growing further. *Chapter 8*.

ratio control Holding one flow in a set proportion to another, measured flow. *Chapter 1*.

ratio station A device or function block that multiplies the wild flow signal by a factor to give

the captive flow's setpoint. *Chapter 1*.

reaction control thruster A small rocket jet that fires in short pulses to push or turn a spacecraft. *Chapter 11*.

refractometer An instrument that reads the sugar content of a liquid from how it bends light. *Chapter 7*.

residence-time distribution The spread of times that parcels of material spend in a process, which sets how long an upset takes to pass. *Chapter 8*.

right-half-plane zero A feature of a model that makes its output start in the wrong direction, which caps how fast feedback can act. *Chapter 9*.

sampled analyzer An analyzer that returns a result only at intervals, after a sample has been taken and measured. *Chapter 8*.

selector A block that passes the lower or the higher of two signals, used for cross-limiting and overrides. *Chapter 1*.

setpoint The value at which a controller is asked to hold its controlled variable. *Chapter 1*.

SIMC rule A tuning rule that sets a PI controller from a first-order model with dead time and one chosen closed-loop time. *Chapter 1*.

soft sensor A model that computes an unmeasured quantity from measurements that are available. *Chapter 8*.

static feedforward A feedforward that is only a gain, holding the steady state but not the timing of the response. *Chapter 1*.

superheat How far a vapor is above the boiling temperature at its pressure, which must stay positive before an expander. *Chapter 5*.

supervisory control A higher layer that writes setpoints to regulatory controllers rather than moving actuators itself. *Chapter 1*.

tachometer A sensor on a rotating shaft whose output is proportional to speed. *Chapter 3*.

thermowell A closed tube that protects a temperature sensor from the process and adds a lag to its reading. *Chapter 5*.

three-element control Boiler drum level control that adds a steam flow feedforward and a feedwater flow loop to the level loop. *Chapter 2*.

thrust vector control Steering a rocket by tilting its engine so the thrust pushes partly sideways. *Chapter 10*.

turbidity The cloudiness of water caused by fine suspended particles, the main quality measure of a treatment plant. *Chapter 6*.

two degrees of freedom A controller with separate paths from the setpoint or disturbance and from the measurement, combining feedforward and feedback. *Chapter 1*.

variable-frequency drive An electronic drive that sets an electric motor's speed by the frequency of the power it supplies. *Chapter 3*.

velocity form A digital PID that computes only the change of its output at each execution. *Chapter 1*.

wild flow The flow in a ratio that is measured but not manipulated, which the other flow follows. *Chapter 1*.

zirconia oxygen probe A heated ceramic cell whose voltage follows the logarithm of the oxygen ratio between flue gas and reference air. *Chapter 4*.

Index of Control Topics

active constraint, 76, 155
actuator saturation, 213, 273
  bandwidth limit, 266
actuator travel, 61, 109
AI agents
  engineering practice, 318
  open problems, 323
  reproduction, 321
  truth audit, 320
  verification record, 320
  while studying, 16
analyzer drift, 85
analyzer validation, 157, 164, 322

back-calculation, 31, 40, 58, 105, 131, 156, 205, 299
bandwidth limit, 203
bandwidth separation, 225
bumpless transfer, 15, 58, 156, 206, 238, 266, 299

capacity limit, 65, 110, 157
cascade control, 8, 22, 24, 57, 94, 225
  current limit, 316
  history, 318
  inner current loop, 196, 217
  inner flow loop, 11, 105
  inner-loop speed, 8, 11, 31, 46, 212, 235, 314
  inner-loop stability, 33, 40
combustion control, 75
computed-torque control, 280
conditional integration, 157
conservation check, 178
constraint forecast, 299
control allocation, 274
control structure, 1
  selection, 10, 22, 316
cross-limiting, 9, 75, 316
current limit, 212
current-mode control, 199, 219

data export, 16
data reconciliation, 157
dead time
  apparent, 126

decoupling, 265
derivative action
  measurement noise, 66
dissolved oxygen control, 46
disturbance
  entry point, 5, 23, 313
disturbance observer, 317
disturbance rejection, 5, 109
draft control, 82

economic optimum
  on a constraint, 84
energy balance, 201
  ledger check, 294
external reset, 105
external reset feedback, 82

fault detection, 274
fault diagnosis, 85
feedforward control, 3, 5, 22
  from the wrong measurement, 34
  ideal compensator, 29
  lead–lag, 5
  measured heat, 95
  model error, 265, 301, 313
  model-based, 97, 255, 265, 318
  nonminimum-phase plant, 211
  open problems, 321
  static, 5, 58, 81, 125, 157, 205, 236, 299
  throughput, 157
  timing, 64, 118
  tuning, 317
finite-element model, 293
finite-volume model, 97, 150
  one pressure per side, 99

gain scheduling, 248
gravity feedforward, 236, 286

input shaping, 286
  zero vibration (ZV), 298
  zero vibration and derivative (ZVD), 298
integrated absolute error, 108
integrating process, 157
integrator windup, 135, 237
inverse response, 197, 235

jerk-limited trajectory, 298
Jury stability test, 33

Kalman filter, 15, 107
    extended, 208

linear quadratic regulator, 206
linearization, 29, 53

material diversion, 182, 190
measurement lag, 53, 67, 102, 155
    tuning for, 111
minimum phase, 204
modal reduction, 293
model mismatch, 111
model predictive control, 41, 106, 140, 238
    against cascade, 316
    economic objective, 266
    explicit, 206, 220
    in the apps, 15
    power electronics, 207, 317
    preview, 107
    relinearized, 238
    soft constraint, 106
    terminal cost, 267
model validation
    withheld data, 180
moving horizon estimation, 15
multivariable coupling, 255

nonminimum phase, 197, 226

offset
    proportional control, 236, 274
offset-free model predictive control, 238
offset-free MPC, 208
open-loop instability, 260
override control, 75, 105
    high pressure, 106
    high select, 82
overshoot
    inverse response, 320
    under inverse response, 217

pairing, 256
PD control, 236, 264
PID control
    anti-windup, 14
    velocity form, 14, 31, 81, 105, 131, 156, 181, 204, 237, 299
PID tuning
    internal model control, 32
    SIMC rule, 32
Posicast control, 290
process gain
    sign change, 128
protection layer, 103

radiation damping, 297
rate limit, 228
ratio control, 2, 7, 137, 145, 171
    air leads, 83
    asymmetric, 76
    cross-limiting, 9, 76
    feedback trim, 7, 82, 88, 125, 156, 181
    fuel leads, 83
    several ratios at once, 174, 189
    trim, 316
reference governor, 299
residence-time distribution, 171, 191
    tanks in series, 176
residual flexibility, 293
right-half-plane zero, 196, 226

sampled analyzer, 178
scenarios, 16, 158
    completion rules, 16, 37, 82, 132
scorecard, 16
second-order system
    natural frequency and damping, 264
selector, 76
selector, minimum, 105
self-regulating process, 26, 59
sensitivity function, 30, 203
sensor bias, 65, 68, 162, 307
sequential control, 132
simulation engine, 13
simulator framework, 12
soft sensor, 171, 317
    bias update, 182
speed scheduling, 286
    constraint-based, 297
state feedback, 264
state-space averaging, 198
step test, 34, 157
successive loop closure, 228
superheat control, 94
supervisory control, 2, 11, 157, 182, 299

throughput manipulation, 145
trajectory feedforward, 290
transfer function
    control-to-output, 203

undershoot, 210

voltage-mode control, 205

Index of Applications

aging loop, 174
agitator, 50
  drive, 53
alkalinity, 128
antifoam, 53
arch draft, 80
automated rendezvous, 257, 280

back pressure, 52
backwash sequencing, 132
boost converter, 196
breakout margin, 152
buffer tank, 152
burner management system, 76

capture radius, 261
carbon dioxide
  solubility, 149, 165
carbon dioxide analyzer, 153, 164
carbon monoxide analyzer, 77
carbon monoxide knee, 79
carbon nanotube, 288, 308
carbonation, 145
carbonator, 151, 164
Carnot bound, 113
center of pressure, 231
charge ratio, 174
chiller
  refrigeration limit, 151
clearwell, 132
climber (space elevator), 292
Clohessy-Wiltshire equations, 258
coagulation, 127
  demand, 129
  restabilization, 127
condenser, 97
continuous conduction, 200
continuous pharmaceutical manufacturing, 172
control valve
  square-root law, 26, 39
cooling water
  fan power, 110
Coriolis acceleration, 259
Coriolis force, 287
counterweight, 290
crosswind, 231, 294

data center, 97
  server heat, 112
dc-dc converter, 196
deaerator, 151
delta-v, 261
descent-speed profile, 232
diafiltration, 178
differential drag, 261
discontinuous conduction, 201
disinfection
  by-products, 130
  CT credit, 130
dissolved air, 150
dissolved oxygen, 46
dissolved oxygen probe, 49
docking, 261
duty cycle, 199

effective gravity, 290
encapsulation, nominal, 177
engine exhaust heat, 96, 119
engine gimbal, 228, 248
evaporator, 97
excess air, 77
expander
  positive-displacement, 97, 101
  scroll, 97

fed-batch fermentation, 49
feed pump, 102
  suction head, 100
fermenter, 46, 69
filler, 152
filter ripening, 130, 138
filter-to-waste, 130
filtration, 127, 130
fired heater, 75
flocculation, 130
flow transmitter, 11, 27, 39
foam, 53
foam index, 152, 165
forced-draft fan, 76, 88
fouling, 111
free drift, 260

friction brake
  fade, 293
fuel ethanol, 49
fuel gas
  header, 84
  heating value, 78, 89

gas-fired water heater, 4
gassed power ratio, 52
generator rating, 113
geostationary orbit, 290
  crossing, 301
gravity gradient, 259
guide rollers, 294

headspace, 152
Henry's law, 52, 147
hold point, 261, 279

impinging-jet mixer, 172, 173
inductor, 199

jar test, 127

lander, planar rigid-body model, 228
level transmitter, 27
lipid nanoparticles, 171
local-vertical local-horizontal frame, 258

manifold maldistribution, 178
mean motion, 259
mine hoist, 287
Monod kinetics, 52
motor heating, 294
mRNA formulation, 171

natural draft, 77

orbital rendezvous, 255, 257, 278
organic Rankine cycle, 94, 96
oxygen analyzer
  zirconia, 88
oxygen transfer rate, 51
oxygen trim, 82
oxygen uptake rate, 51, 69

polydispersity index, 177, 189
positioning control, 77
power stage, 199, 218
powered descent, 227, 228, 249
process heater, 77
propellant budget, 261
propellant depletion, 231
property table, 102

pulse-width modulation, 200

quench, 174

R1233zd(E), 97
reaction control thruster, 263, 279
receiver, 97
refractometer, 148
regenerative braking, 292, 307
relative motion, 258
ribbon (space elevator), 290, 308
rocket landing, 225, 247
Rushton turbine, 50

scale-out, 174, 190
sedimentation, 130
soft drink, 145
space elevator, 286, 289
spacecraft, 255
sparger, 50
specific impulse, 231
stack damper
  characteristic, 80
stack loss, 79
start-up, 184
station keeping, 273
sulfite oxidation method, 48
supply header, 25
switching converter, 197, 220
syrup proportioner, 146

tachometer, 53
tank level control, 22
target and chaser, 258
tether dynamics, 289, 290, 307
thermowell, 102, 118
thrust vector control, 225
touchdown classification, 232
tramp air, 77
turbidimeter, 126, 137, 139
two-burn transfer, 267

variable-frequency drive, 57, 67, 118
volumetric mass transfer coefficient, 51

waste heat recovery, 94
wastewater aeration, 48
water treatment, 125
  coagulant dosing, 125
wind gust, 231
working fluid
  dry and wet, 97

yeast, 49

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.