INTERACTIVE CONTROL ENGINEERING

Nonlinear Dynamics, Oscillations, and Stabilization

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 7

Nonlinear Dynamics, Oscillations, and Stabilization

Plants that will not stay put without control

John D. Hedengren
Editor

Advanced Process Solutions, LLC

Interactive Control Engineering: Principles and Applications
Volume 7: Nonlinear Dynamics, Oscillations, and Stabilization
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/v07-nonlinear-stabilization/

How to cite this volume. John D. Hedengren, Ed., *Nonlinear Dynamics, Oscillations, and Stabilization*, vol. 7 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 plants that do not stay put without control. Some have more than one steady state, some are unstable at the operating point, some ring, and some change the sign of their gain as they move. Chapters 2 to 6 introduce the behavior on a reactor at its ignition point, a pendulum driven by a propeller, a braking wheel at its friction peak, a ball on a beam and a levitated ball. Chapters 7 to 11 balance and steer mechanical systems with state feedback. Chapters 12 to 14 deal with oscillation itself: a swinging load, a chattering cutter and a string of vehicles that amplifies a disturbance.

Volume 2 tunes loops that are stable on their own. Volume 20 treats motion tasks once stability is in hand.

John D. Hedengren
2026

Contents

Preface v

Notation x

1 Introduction: Stabilizing Unstable and Oscillating Plants 1

1.1 The Question This Volume Answers . . . 1
1.2 A Short History . . . 2
1.3 Notation and Conventions . . . 3
1.4 Stability Theory on One Pendulum . . . 3
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 Exothermic Reactor Simulator for Temperature Control Across an Ignition Boundary 22

2.1 Introduction . . . 23
2.2 Background . . . 24
2.3 Reactor Model . . . 25
2.4 Steady-State Multiplicity and Stability . . . 29
2.5 Dynamic Analysis . . . 31
2.6 Controller Design . . . 33
2.7 Results . . . 37
2.8 Predictive Control at the Unstable Operating Point . . . 41
2.9 Estimation After a Runaway . . . 43
2.10 Technology in Practice . . . 44
2.11 Limitations . . . 46
2.12 Discussion . . . 47
2.13 Conclusion . . . 47
2.14 Exercises . . . 49

3 Aeropendulum Simulator in Which the Setpoint Chooses the Plant 53

3.1 Introduction . . . 54
3.2 Propeller Arms on the Control Bench . . . 55
3.3 The Kit as Modeled . . . 56
3.4 One Arm Linearized at Every Angle . . . 60
3.5 The Kit's Controller as Coded . . . 62
3.6 Nine Cases and a Setpoint Program . . . 65
3.7 Technology in Practice . . . 69
3.8 Limitations . . . 71
3.9 Conclusion . . . 72
3.10 Exercises . . . 73

4 Anti-Lock Braking at the Friction Peak Where the Process Gain Changes Sign 77

4.1 Introduction . . . 78
4.2 Background . . . 79
4.3 Vehicle Model . . . 80
4.4 Stability Analysis of the Slip Dynamics . . . 85

4.5 Controller Design . . . 87
4.6 Results . . . 90
4.7 Technology in Practice . . . 98
4.8 Limitations . . . 99
4.9 Conclusion . . . 101
4.10 Exercises . . . 103

5 A Desktop Ball and Beam Whose Only Damping Comes From the Controller 107
5.1 Introduction . . . 108
5.2 The Ball and Beam in the Laboratory and the Literature . . . 109
5.3 The Kit as Simulated . . . 110
5.4 Two Integrators and a Lag . . . 113
5.5 The Kit's Controller as Coded . . . 115
5.6 Eleven Cases, Measured . . . 118
5.7 Technology in Practice . . . 123
5.8 Limitations . . . 125
5.9 Conclusion . . . 126
5.10 Exercises . . . 127

6 Magnetic Suspension Virtual Laboratory to Teach Control of Unstable Systems 131
6.1 Introduction . . . 132
6.2 Background . . . 133
6.3 Suspension Model . . . 134
6.4 Linearization and the Unstable Pole . . . 138
6.5 Controller Design . . . 140
6.6 Stability Conditions for the Closed Loop . . . 142
6.7 Upper Bounds on the Proportional Gain . . . 144
6.8 Results . . . 145
6.9 Numerical Implementation . . . 150
6.10 Technology in Practice . . . 150
6.11 Limitations . . . 152
6.12 Discussion . . . 152
6.13 Conclusion . . . 153
6.14 Exercises . . . 155

7 Inverted Pendulum Simulator for Balancing a Pole Without Losing the Cart 158
7.1 Introduction . . . 159
7.2 Balancing a Pole on a Moving Base . . . 160
7.3 Cart and Pole Model . . . 161
7.4 Linear Analysis of the Balancing Problem . . . 164
7.5 Implemented Control Structures . . . 167
7.6 Results Across the Scenario Sequence . . . 172
7.7 Repeatability of a Quantized Loop . . . 180
7.8 Technology in Practice . . . 180
7.9 Limitations . . . 182
7.10 Conclusion . . . 183
7.11 Exercises . . . 185

8 Double Inverted Pendulum Simulator for Model-Based Balance and Computed Swing-Up 188
8.1 Introduction . . . 189
8.2 A Benchmark Held, Then Swung Up . . . 190

8.3 The Cart and Its Two Links . . . 191
8.4 What the Linearization Allows . . . 195
8.5 The Digital Controller . . . 196
8.6 The Cases, One Requirement at a Time . . . 201
8.7 Technology in Practice . . . 209
8.8 Limitations . . . 211
8.9 Conclusion . . . 212
8.10 Exercises . . . 213

9 A Riderless Benchmark Bicycle That Rides Itself Only Between Two Speeds 217
9.1 Introduction . . . 218
9.2 Self-Steering Bicycles, From the Safety Bicycle to the Benchmark . . . 219
9.3 The Bicycle as Modeled . . . 220
9.4 A Window in Speed . . . 226
9.5 The Controller as Coded . . . 228
9.6 Nine Cases and a Slalom, Measured . . . 230
9.7 Technology in Practice . . . 233
9.8 Limitations . . . 236
9.9 Conclusion . . . 236
9.10 Exercises . . . 238

10 Holding a Drifting Car at an Unstable Equilibrium With Steer and Throttle 241
10.1 Introduction . . . 242
10.2 From the Handling Limit to Automated Drifting . . . 243
10.3 The Car and the Pad . . . 244
10.4 The Drift as an Equilibrium . . . 248
10.5 Implemented Control System . . . 250
10.6 Ten Cases, Measured . . . 251
10.7 Technology in Practice . . . 255
10.8 Limitations . . . 257
10.9 Conclusion . . . 258
10.10 Exercises . . . 259

11 Active Suspension Simulator for Comfort and Road Holding Through One Strut Force 263
11.1 Introduction . . . 264
11.2 Suspension Control From Fixed Dampers to Road Preview . . . 265
11.3 Quarter Car on a Road Defined in Distance . . . 266
11.4 One Force on Two Masses . . . 270
11.5 The Force Request and Its Three Struts . . . 272
11.6 Scenarios and Measured Results . . . 277
11.7 Technology in Practice . . . 281
11.8 Limitations . . . 283
11.9 Conclusion . . . 283
11.10 Exercises . . . 285

12 Input Shaping on an Overhead Crane Where One Actuator Serves Two Outputs 288
12.1 Introduction . . . 289
12.2 Background . . . 290
12.3 Crane Model . . . 291
12.4 Dynamic Analysis . . . 295
12.5 Controller Design . . . 298
12.6 Results . . . 301

12.7 Technology in Practice . . . 308
12.8 Limitations . . . 309
12.9 Conclusion . . . 310
12.10 Exercises . . . 312

13 Stability Lobes and Adaptive Feed Control in a CNC Milling Simulator 315
13.1 Introduction . . . 316
13.2 Chatter Theory, Speed Selection and Adaptive Machining . . . 317
13.3 Tool, Chip and Machine Model . . . 318
13.4 Two Controllers on One Machine . . . 324
13.5 Results on the Pocket . . . 326
13.6 Technology in Practice . . . 331
13.7 Limitations . . . 332
13.8 Conclusion . . . 333
13.9 Exercises . . . 334

14 String Stability of Well-Tuned Followers in a Vehicle Platoon Simulator 338
14.1 Introduction . . . 339
14.2 From Vehicle Strings to Cooperative Platoons . . . 340
14.3 The Line, Its Sensors and Its Radio . . . 341
14.4 The Criterion From the Cars' Own Law . . . 346
14.5 Cruise Control, Following Laws and the Supervisor . . . 349
14.6 Eleven Runs Along the Line . . . 352
14.7 Technology in Practice . . . 358
14.8 Limitations . . . 360
14.9 Conclusion . . . 361
14.10 Exercises . . . 363

15 Summary and Outlook 366
15.1 What the Chapters Showed . . . 366
15.2 Choosing a Method . . . 369
15.3 Where Practice Is Heading . . . 370
15.4 Engineering With AI Agents . . . 372
15.5 Gaps in Nonlinear Stabilization . . . 375
15.6 Exercises . . . 379

Answers to Selected Exercises 381

Glossary 386

Index of Control Topics 389

Index of Applications 392

About the Editor 394

Volumes in This Series 395

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: Stabilizing Unstable and Oscillating Plants



The chapters 2 exothermic reactor 3 aeropendulum 4 braked wheel 5 ball and beam 6 levitated ball 7 pole on a cart 8 two links on a cart 9 riderless bicycle 10 drifting car 11 quarter car 12 gantry crane 13 end mill 14 line of cars

Notice that the pole, the two links, the levitated ball and the drifting car would fall or spin out without their controllers, while the crane, the strut and the mill settle by themselves but ring. Illustration, not to scale.

IN PLAIN TERMS **Holding What Would Fall or Shake**

The main idea. Some machines will not stay put: a pole on a cart falls, a ball under a magnet drops or jumps to it, and a hot reactor can run away. Others, like a load on a crane, settle but keep swinging.

The challenge. A controller holds such a machine only near the point it was designed for, and only while its push is strong and quick enough. A push that comes too late or too weak cannot catch the machine again.

Why it matters. The same ideas keep braked wheels from locking, crane loads from swinging and lines of cars from passing a slowdown back larger.

THREE WAYS TO READ THIS CHAPTER

Explore	7 min	This page, Section 1.2 with its timeline and two History boxes, the Did you know? boxes in Sections 1.4.2, 1.4.7 and 1.4.11, Figure 1.11, and Key points in Section 1.8. No equations.
Study	40 min	The whole chapter in order, with each number of the pendulum example worked, then Exercises 1 to 10.
Practice	5 min	This page, Section 1.4.11 with Table 1.2, Section 1.5 with Figures 1.7 and 1.9, and Section 1.8 with Figure 1.11.

1.1 The Question This Volume Answers

Most loops in the earlier volumes act on a plant that would settle by itself. The plants of this volume do not. A reactor runs between two folds of its steady-state curve, a ball hangs under a magnet that pulls harder the closer it comes, a pole stands on a cart, a car drifts sideways at the edge of its tires, and a line of cars passes each brake tap back larger than it came. Others settle but ring: a crane load swings, a car body bounces, a cutting tool chatters. This volume answers one question: what holds such a plant, and how small is the margin by which it is held?

Thirteen chapters answer it on thirteen plants, each with an app that the reader runs in a browser. Chapters 2 to 4 are plants whose steady-state gain changes size and sign with the operating point. Chapters 5 to 10 are unstable mechanical plants held by feedback, from one unstable pole to two, and from one input to two. Chapters 11 to 13 are oscillating plants, where a resonance or a delay sets the limit. Chapter 14 closes the volume on a line of cars, where each loop is stable and the line may not be.

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

- Find the steady states of a nonlinear model, the gain at each, and the folds where steady states meet and vanish.

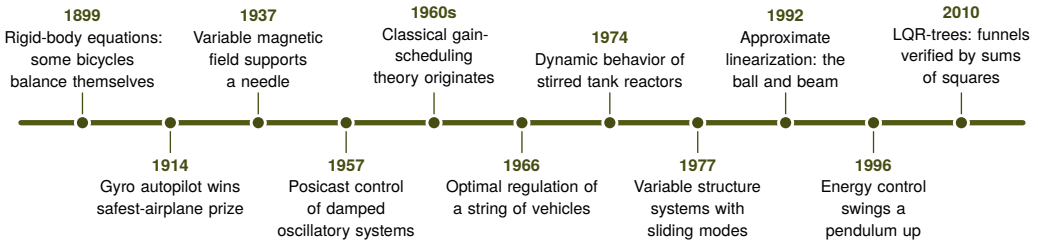


Figure 1.1 Milestones of stabilization in the systems of this volume. The dates are from [4] (1899), [6] (1914), [5] (1937), [7] (1957), [8] (1960s), [9] (1966), [10] (1974), [11] (1977), [12] (1992), [13] (1996) and [14] (2010). Notice that the bicycle, the airplane and the floating needle come before the linear theory of Section 1.4, and that the methods for large motions come last. Events are evenly spaced, not to a time scale.

- Linearize the model at an operating point, write it in state space and read stability, speed and mode shape from its eigenvalues.
- Design a state-feedback gain by pole placement or as a linear quadratic regulator (LQR), and an observer for the states not measured.
- Say what a right-half-plane pole or zero, an actuator limit, a lag or a delay costs, before any controller is tuned.
- Remove a resonance from a command by input shaping, predict a limit cycle, and test a chain of loops for string stability.

This chapter develops each of these on one small example and names the chapter that uses it. Standard texts give fuller treatments [1–3]. PID forms and tuning are taken as known from Volume 2.

1.2 A Short History

The milestones of Figure 1.1 follow the methods of this volume through very different machines, some of them held before a theory explained them. Rigid-body equations of 1899 showed that some safety bicycles balance themselves in the right range of speed [4], the window that Chapter 9 measures. A paper of 1937 described a vertical ferromagnetic needle supported by magnetic forces alone with a variable field [5], the single axis of Chapter 6.

> ⧖ HISTORY **1903 to 1914: Feedback Holds an Unstable Airplane**
>
> The Wright brothers settled the trade between stability and maneuverability with an airplane that was unstable but maneuverable, flown at Kitty Hawk in 1903, and its pilot had to keep adjusting the rudder. Sperry then used a gyro-stabilized pendulum to indicate the vertical and a feedback mechanism that pulled the stick to raise the nose when the plane pointed down, and the reverse. It was the first use of feedback in aeronautical engineering, and it won a prize in a competition for the safest airplane, in Paris in 1914. Feedback had stabilized an unstable machine [6], as Section 1.4.5 does for the upright arm.

Oscillation was first taken out of the command. Posicast control, published in 1957, split a step into two excitations spaced so that their ringing cancels [7], the root of the shapers of Chapter 12. The state-space tools of Section 1.4 followed: a new approach to filtering and prediction in 1960 [15], then the observer [16] and the question of when a linear control system is optimal [17], both in 1964.

Nonlinearity entered next, first as a family of linear models, a story the second box below tells. A paper of 1974 on the dynamic behavior of continuous stirred tank reactors [10] gave the S-shaped curve of Chapter 2, and one of 1977 set out variable structure systems with sliding modes [11], the switching law of Chapter 4. Feedback linearization was a new approach in flight control in the

mid-1980s, under the name dynamic inversion [18], as Chapter 3 tells, and a paper of 1992 took the ball and beam as its example of approximate input-output linearization [12].

The last milestones concern large motions and chains of loops. An energy method for swinging a pendulum up, presented in 1996, left the upright point to another controller to catch [13], the switch of Chapter 7. In 2010 regions of attraction became funnels around trajectories stabilized by LQR and verified by sums of squares [14], the handover test of Chapter 8. For chains, a paper of 1966 regulated a string of moving vehicles by optimal control [9], and one of 1996 set out the string stability of interconnected systems [19] that Section 1.4.10 tests.

> ⧖ HISTORY **From the 1960s: Divide and Conquer by Gain Scheduling**
>
> Wide variations of operating conditions in aerospace stimulated the development of gain scheduling and adaptive control [6]. Much of the classical theory of gain scheduling originates from the 1960s. It decomposes a nonlinear design into linear subproblems, so that well established linear methods apply, and safety certification requirements are often based on linear methods. A review of 2000 called it perhaps one of the most popular nonlinear design approaches, applied from aerospace to process control. Relaxing its condition that the plant pass between operating regions slowly was still very much an open question [8].

1.3 Notation and Conventions

A model is written $\dot{x} = f(x,u)$ with the state x and the input u. An equilibrium, or steady state, is a pair (x_0,u_0) with $f(x_0,u_0) = 0$. Deviations from it are written δx and δu, or x and u again where no confusion arises. The linearized model is $\delta\dot{x} = A\,\delta x + B\,\delta u$ with a measurement $y = Cx$. A state feedback is $u = -Kx$ with a row of gains K, an observer has the gain L, and the weights of an LQR are Q and R, as in the predictive controller of Section 1.6.3. An unstable real pole is p and a right-half-plane (RHP) zero is z, both in 1/s or rad/s. The control period is Δt, an actuator lag is τ and a pure delay is τ_d.

The running example uses seconds, radians and rad/s^2 for its input. Each chapter keeps the units of its plant: kelvin and minutes for the reactor, milliseconds for the levitated ball and the double pendulum, and meters per second for the vehicles. Several chapters write the unstable root differently. Chapter 7 writes it λ_0, Chapter 6 and Chapter 4 write p, and Chapter 4 also uses p for a stable slip pole. Chapter 14 writes the time gap h, which this chapter writes t_g, because h is the integration step of Section 1.6.1. The framework's $a = \pm 1$ is the controller action, so the cart's acceleration here is $\ddot{x}$.

1.4 Stability Theory on One Pendulum

1.4.1 The Running Example

The running example is one simple pendulum, a point mass on a light rod of length $\ell = g/4 = 2.45\,\text{m}$, so that $g/\ell = 4/\text{s}^2$ and $\sqrt{g/\ell} = 2\,\text{rad/s}$. The pivot is frictionless. The input u is an angular acceleration in rad/s^2, and it enters in two ways, as shown in Figure 1.2. In the arm, a motor at a fixed pivot applies a torque, and u is that torque divided by $m\ell^2$. On the cart, the pivot rides a cart whose drive sets its acceleration $\ddot{x}$, and $u = \ddot{x}/\ell$. With θ measured from hanging, the two give Equation (1.1).

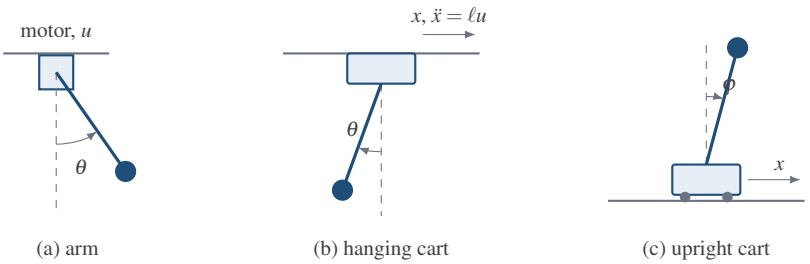
$$\text{arm:}\quad \ddot{\theta} = -4\sin\theta + u, \qquad \text{cart:}\quad \ddot{\theta} = -4\sin\theta - u\cos\theta, \quad \ddot{x} = \ell u \tag{1.1}$$

Hanging, the pendulum swings with a period of $\pi = 3.14\,\text{s}$. Upright, a small tilt grows by a factor of e every 0.5 s. The upright arm, the upright pole on a cart and the crane load of the hanging cart are the three plants of this section. The arm also stands for the aeropendulum of Chapter 3, the hanging cart for the crane of Chapter 12, and the upright cart for the pendulums of Chapter 7 and Chapter 8.

1.4.2 Steady-State Gain and Multiplicity

A constant input holds the arm at rest where gravity balances it, as shown in Equation (1.2).

$$u_0 = 4\sin\theta_0, \qquad \frac{\mathrm{d}\theta_0}{\mathrm{d}u_0} = \frac{1}{4\cos\theta_0} \tag{1.2}$$



(a) arm (b) hanging cart (c) upright cart

Figure 1.2 The Running Example. One Pendulum With $g/\ell = 4/\mathrm{s}^2$, Driven (a) by a Motor at a Fixed Pivot, (b) by the Acceleration of a Cart Below Which It Hangs and (c) by the Same Cart With the Pendulum Upright. The Tilt From Upright Is $\varphi = \theta - \pi$.

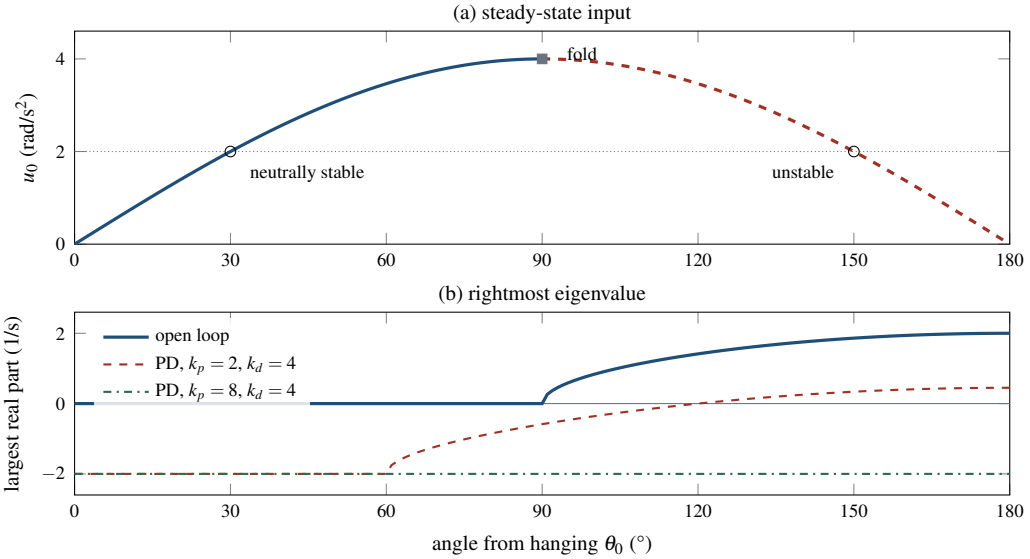


Figure 1.3 Steady States of the Arm. (a) The Input That Holds Each Angle, Neutrally Stable Below Horizontal (Solid) and Unstable Above (Dashed), With the Two Steady States of $u_0 = 2$ (Circles) and the Fold (Square). (b) The Largest Real Part of the Eigenvalues: Open Loop (Thick Solid) and With the PD and Gravity Feedforward of Equation (1.3) at $k_p = 2$ (Dashed) and $k_p = 8$ (Dash-Dotted).

The second expression is the steady-state gain, the slope of the inverse of the first. Figure 1.3(a) draws the steady-state curve. An input of $2\,\mathrm{rad/s}^2$ has two steady states, at 30° and 150°. Their gains are equal and of opposite sign, ± 0.289 rad per $\mathrm{rad/s}^2$. The largest input that holds the arm is $4\,\mathrm{rad/s}^2$, at horizontal. There the two branches meet, the gain is infinite, and above it no steady state exists. Such a turning point is a fold, or saddle-node bifurcation [1].

The two branches differ in stability. Linearizing Equation (1.1) at θ_0 gives $\delta\ddot{\theta} = -4\cos\theta_0\,\delta\theta + \delta u$, a stiffness $4\cos\theta_0$ that gravity supplies below horizontal and takes away above it. The eigenvalues are $\pm\sqrt{-4\cos\theta_0}$. At 30° they are $\pm 1.86\mathrm{j}$, an undamped swing, and at 150° they are ± 1.86, one of them unstable, as Figure 1.3(b) shows. The equilibrium loses stability exactly where the holding input is largest, because there the stiffness passes through zero.

The same picture organizes three chapters. The reactor of Chapter 2 has an S-shaped curve with two folds, three steady states inside a window of jacket temperature, and an open-loop unstable interval that contains the default setpoint [10]. The wheel of Chapter 4 has the arm's shape exactly:

the brake torque that holds a slip has a maximum, the equilibrium loses stability at that maximum, and a larger torque has no equilibrium at all. The aeropendulum of Chapter 3 is the arm with a propeller, and its stiffness changes sign at horizontal. Chapter 10 solves for its equilibrium before it designs anything, and Chapter 6 has one at every gap and none stable.

> ✦ **DID YOU KNOW?** **Climbing to the Peak Since 1951**
>
> An extremum-seeking controller varies its reference slowly and sinusoidally and reads from the phase of the response which side of the optimum it is on. Extremal control was first applied to aircraft engines in 1951 [6], and Chapter 4 names it as one way to stay on a tire's friction peak.

1.4.3 Linearization at an Operating Point

A linearization describes the plant near one point. The arm shows how much it changes from point to point. Consider the proportional-derivative (PD) law of Equation (1.3) with a gravity feedforward.

$$u = 4\sin\theta_{\mathrm{SP}} - k_p(\theta - \theta_{\mathrm{SP}}) - k_d\dot{\theta} \tag{1.3}$$

At a setpoint θ_0 the closed loop is $s^2 + k_d s + (k_p + 4\cos\theta_0) = 0$. The controller adds stiffness to what gravity supplies, and the loop is stable while the sum is positive. With $k_p = 8$ and $k_d = 4$ the damping ratio is 0.577 hanging, 0.707 at horizontal and 1.0 upright, so one tuning serves every angle. With $k_p = 2$ the loop loses the arm beyond 120°, where $2 + 4\cos\theta_0$ turns negative, as the dashed line of Figure 1.3(b) shows. Without the feedforward the same $k_p = 8$ leaves an offset, because the controller must supply the holding input from an error: a 90° setpoint settles at 64.2°.

Two remedies keep a tuning valid at every point. A gain schedule recomputes k_p from the measured angle so that the closed-loop polynomial stays the one at the design point. From $k_p = 8$ at hanging it gives $k_p = 12 - 4\cos\theta_0$, 16 upright, and $s^2 + 4s + 12$ everywhere. Feedback linearization cancels the nonlinearity instead: $u = 4\sin\theta + v$ leaves $\ddot{\theta} = v$, a double integrator at every angle, for an outer law v. Both need the model to be right. Both also trust a linearization whose validity shrinks as the plant moves faster than the schedule.

The upper branch carries one more lesson. Its steady-state gain is negative, yet the loop of Equation (1.3) uses the same signs there as below horizontal. A positive u always accelerates the arm toward larger θ at once, and only the eventual response reverses. A controller fast enough to act on the immediate response must take its sign from it, not from the steady-state gain.

Chapter 2 meets exactly this: its gain is negative on the middle branch, and its reverse-acting PI loop works because warming the jacket always warms the reactor first. Chapter 3 compares a fixed tuning, a schedule that holds the closed-loop polynomial and feedback linearization across 15 to 150°. Chapter 9 schedules on speed, Chapter 10 on speed and friction, and Chapter 6 shows one tuning giving three responses across three gaps.

1.4.4 State-Space Models and Eigenvalues

Near upright, with the tilt $\varphi = \theta - \pi$ and its rate as the state, the arm is the model of Equation (1.4).

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\varphi\\ \dot{\varphi}\end{bmatrix} = \underbrace{\begin{bmatrix}0 & 1\\ 4 & 0\end{bmatrix}}_{A}\begin{bmatrix}\varphi\\ \dot{\varphi}\end{bmatrix} + \underbrace{\begin{bmatrix}0\\ 1\end{bmatrix}}_{B}u, \qquad y = \underbrace{\begin{bmatrix}1 & 0\end{bmatrix}}_{C}\begin{bmatrix}\varphi\\ \dot{\varphi}\end{bmatrix} \tag{1.4}$$

The eigenvalues of A are ± 2, with the eigenvectors $(1, 2)$ and $(1, -2)$. In the coordinates of these eigenvectors the model splits into two first-order modes, as shown in Equation (1.5).

$$\sigma = \dot{\varphi} + 2\varphi, \quad \dot{\sigma} = 2\sigma + u, \qquad \rho = \dot{\varphi} - 2\varphi, \quad \dot{\rho} = -2\rho + u \tag{1.5}$$

The unstable mode σ is the only thing the controller must fight. A tilt that leans back toward upright at twice its size per second has $\sigma = 0$ and returns by itself.

The upright cart has four states, $(x, \dot{x}, \varphi, \dot{\varphi})$, and the linear model $\ddot{x} = \ell u$, $\ddot{\varphi} = 4\varphi + u$. Its eigenvalues are $0, 0, 2$ and -2. One input reaches all four states, so the model is controllable: the controllability matrix $[B, AB, A^2B, A^3B]$ has rank four. Observability is another matter. Because the drive sets the acceleration, the cart never feels the pendulum. The cart's encoder alone therefore says nothing about the pole, and the pole's encoder alone says nothing about where the cart is. Each observability matrix has rank two, and the two together have rank four.

Chapter 9 plots the four eigenvalues of the bicycle against speed and finds a window from 4.292 to 6.024 m/s in which all decay [4]. Chapter 8 reads the shapes of its two unstable modes from the eigenvectors, one folding the links and one leaning them together. Chapter 10 finds two unstable eigenvalues where an analysis at fixed speed finds one, and Chapter 2 tabulates the eigenvalues along its steady-state curve.

1.4.5 Pole Placement and the Linear Quadratic Regulator

A state feedback $u = -k_1\varphi - k_2\dot{\varphi}$ closes the upright arm on $s^2 + k_2 s + (k_1 - 4) = 0$, so any pair of poles can be placed. Poles at -4 and -4 need $k_1 = 20$ and $k_2 = 8$. Placement leaves open which poles to ask for. The LQR answers by cost: it minimizes $J = \int_0^\infty (x^\top Q x + R u^2)\,\mathrm{d}t$ with $K = R^{-1}B^\top P$, where P solves the algebraic Riccati equation of Equation (1.6).

$$A^\top P + PA - PBR^{-1}B^\top P + Q = 0 \tag{1.6}$$

For the arm with $Q = \mathrm{diag}(q_1, q_2)$ and $R = 1$ it has the closed form of Equation (1.7).

$$k_1 = 4 + \sqrt{16 + q_1}, \qquad k_2 = \sqrt{q_2 + 2k_1} \tag{1.7}$$

The weights $Q = \mathrm{diag}(20, 5)$ give $K = [10, 5]$ and poles at -2 and -3. As Q shrinks to zero, the gain tends to $[8, 4]$ and the poles to -2 twice. That limit is the cheapest stabilizing gain: it moves the unstable pole at $+2$ to its mirror image at -2 and leaves the stable one where it was. No controller holds the arm for less input energy, which makes the mirrored pole a floor on what stabilization costs.

Weights are usually set by Bryson's rule, one over the square of the largest acceptable excursion of each state and of the input [3]. Every single-input LQR keeps at least 60° of phase margin and tolerates a gain reduction to one half, a property of the optimal loop [17]. The apps design in discrete time, with the Riccati equation solved for the plant sampled at Δt. At $\Delta t = 0.01$ s the same weights give $K = [9.85, 4.93]$, close to the continuous gain.

Chapter 7 designs a discrete LQR on the cart and pole with Bryson's weights, and Chapter 8 shows that the regulator must push against the upper link's lean, which no pair of intuitive loops does. Chapter 5 compares an observer LQR with PD and PID, Chapter 9 schedules its LQR on speed, and Chapter 10 designs a two-input LQR at the drift equilibrium. Chapter 3 offers an LQR with integral action as a comparison.

1.4.6 Observers

The gain K needs the rate $\dot{\varphi}$, and an encoder reads only the angle. An observer runs a copy of the model and corrects it by the measurement error, as shown in Equation (1.8).

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x}) \tag{1.8}$$

The estimation error obeys $\dot{e} = (A - LC)e$, so L places the observer's poles as K placed the controller's [16]. For the arm the error polynomial is $s^2 + l_1 s + (l_2 - 4)$, and poles at -8 twice need $L = [16, 68]$. With $u = -K\hat{x}$ the closed loop has the controller's poles and the observer's poles together, -2, -3, -8 and -8. This is the separation principle. A Kalman filter is an observer whose gain is chosen from the noise statistics [15].

Figure 1.4 catches the arm from a tilt of 0.1 rad with $K = [10, 5]$ every 0.01 s, on an encoder of 4096 counts per turn. The three estimates of the rate, the true state, the observer and the encoder

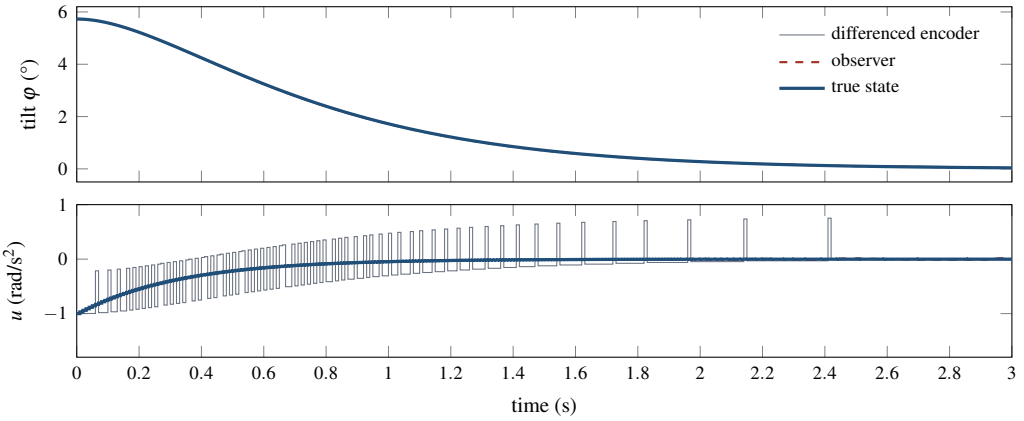


Figure 1.4 Catching the Upright Arm From 0.1 rad With $K = [10, 5]$ Every 0.01 s: the True State (Thick Solid), an Observer With Poles at -8 (Dashed) and the Encoder Differenced Over One Period (Thin Gray). The Tilt Is Nearly the Same, the Input Is Not.

differenced over one period, give nearly the same tilt. All three are within 0.01 rad of upright from 1.62 s. The input tells them apart. One count differenced over 0.01 s is a rate step of 0.15 rad/s, so the differenced law jumps by up to $1.55\,\text{rad/s}^2$ in one period. After 5 s its input has an RMS of $0.487\,\text{rad/s}^2$, 33 times the observer's $0.0146\,\text{rad/s}^2$.

Chapter 8 runs the same comparison on three encoders, with the true state, filtered differences and a Kalman filter side by side. Chapter 5 estimates the ball's speed with an α-β filter, a discrete observer with a double pole, and adds a four-state observer for its torque rig. Chapter 9 blends a gyro and an accelerometer in a complementary filter. Chapter 6, Chapter 12 and Chapter 2 use Kalman filters behind their predictive controllers, and Chapter 2 sets moving horizon estimation beside one, as Section 1.6.3 describes.

1.4.7 Limits Set by Right-Half-Plane Poles and Zeros

An unstable pole sets limits that no gain removes [20, 21]. The arm shows four of them through the mode σ of Equation (1.5).

The first is actuator headroom. With $|u| \le u_{\max}$, the mode $\dot\sigma = 2\sigma + u$ can be turned back only while $|\sigma| < u_{\max}/p$. With $u_{\max} = 2\,\text{rad/s}^2$ the linear model allows a tilt from rest of 0.5 rad and a push to 1 rad/s from upright. The nonlinear arm allows 30° from rest, where gravity's $4\sin\varphi$ equals the limit. This region belongs to the plant and the actuator. All three gains of Table 1.1, clipped at ± 2, catch every tilt up to 30° and none beyond, and no controller can do better.

The second is delay. Proportional feedback on σ through a delay τ_d is stable for some gain only if $p\,\tau_d < 1$. At $\tau_d = 0.1$ s the gain must lie between 2 and 14.5, at 0.25 s between 2 and 5.07, and at 0.5 s no gain works. The window has a floor at p and a ceiling that falls with the delay.

The third is lag. A PD law on the arm through an actuator lag τ gives $\tau s^3 + s^2 + (k_2 - 4\tau)s + (k_1 - 4) = 0$, and the Routh test reduces to Equation (1.9).

$$k_1 > 4, \qquad \tau < \frac{k_2}{k_1} \tag{1.9}$$

The lag must be shorter than the derivative time, whatever the gain. More gain does not help, because it raises both terms together.

The fourth is the RHP zero of the cart. Holding a lean φ needs the input $u = -4\varphi$, so the cart's position follows the lean through $x(s) = \ell\,(s^2 - 4)\,\varphi(s)/s^2$, which has a zero at $z = +2$. To move right, the cart must first move left to set up the lean. A loop around a RHP zero must be slow, and

Table 1.1 Margins of Three Stabilizing Gains on the Upright Arm

	Placement	LQR	Least energy
Gain $K=[k_1,k_2]$	$[20,\ 8]$	$[10,\ 5]$	$[8,\ 4]$
Closed-loop poles	$-4,\ -4$	$-2,\ -3$	$-2,\ -2$
Input for a 0.1 rad tilt (rad/s^2)	2.0	1.0	0.8
Lowest gain factor kept stable, $4/k_1$	0.20	0.40	0.50
Crossover (rad/s)	7.89	4.58	3.46
Phase margin (°)	72.4	66.4	60.0
Delay margin (s)	0.160	0.253	0.302
Longest actuator lag, k_2/k_1 (s)	0.40	0.50	0.50
Tilt caught from rest, $\lvert u\rvert \le 2$ (°)	30	30	30

a loop around a RHP pole must be fast, roughly a bandwidth below $z/2$ and above $2p$ [2]. On the upright cart this asks for an angle loop faster than 4 rad/s inside a cart loop slower than 1 rad/s. For a controller that reads only one output with both a pole p and a zero z, the sensitivity peak is at least $(z+p)/\lvert z-p\rvert$ [22]. With the drive setting the acceleration, $z=p$ and the bound is infinite, which restates that the cart's encoder alone cannot hold the pole.

Table 1.1 puts these limits beside the usual margins for three gains, with the loop broken at the input, $K(sI-A)^{-1}B=(k_2s+k_1)/(s^2-4)$. The margins disagree. The fast placement has the most phase margin and the least delay margin, because its crossover is highest. The least-energy gain tolerates the longest delay, 0.302 s or 0.6 of the 0.5 s time constant of the unstable mode, but it loses the arm at the smallest drop in gain. An unstable loop also has a lower gain margin: below a factor $4/k_1$ the controller no longer outweighs gravity. No single number says how robust a stabilizing loop is.

> **+ DID YOU KNOW? The Most Phase Margin, the Least Delay Margin**
>
> From the simulator: the placement gain $[20,\ 8]$ keeps 72.4° of phase margin, the most of the three, yet a delay of 0.160 s loses the arm. The LQR gain $[10,\ 5]$, with 66.4°, tolerates 0.253 s (Table 1.1).

Chapter 6 derives the lag condition as $\tau_D>\tau_c$ and the gain floor as $K_c>p^2/B$. Chapter 7 maps the window of angle gain against delay, finds it closing at about 0.59 of the e-folding time, and bounds a loop on the cart position alone by a sensitivity peak of 46. Chapter 8 finds two RHP zeros and a sampling margin, Chapter 9 reads countersteering as a RHP zero, and Chapter 5 measures a limit on stopping the ball that no tuning overcomes.

1.4.8 Resonance and Input Shaping

Hanging, the cart's pendulum obeys $\ddot{\theta}+4\theta=-u$. It is a resonance at 2 rad/s with no damping. Any cart move excites it. A single step of the cart's acceleration to 0.5 m/s^2 leaves the load swinging by $0.5/g=0.051$ rad (2.92°) about its tilted rest. The move of Figure 1.5(a) accelerates at 0.5 m/s^2 for 2 s, coasts for 2 s and stops in 2 s, covering 4 m. It leaves a swing of 8.03°.

Input shaping removes the swing from the command instead of the feedback [7, 23]. An impulse leaves a sinusoidal residue, and two impulses half a period apart leave residues that cancel. The zero-vibration (ZV) shaper convolves the command with impulses of 1/2 at $t=0$ and 1/2 at $\pi/\omega_n=1.571$ s. The ZVD shaper also cancels the residue's derivative with respect to the frequency, using 1/4, 1/2 and 1/4 at 0, 1.571 and 3.142 s. Both amplitudes sum to one, so the cart still travels 4 m and never exceeds the unshaped acceleration. The move ends 1.57 s later with ZV and 3.14 s later with ZVD, and both leave less than 0.01° of swing.

The shaper is built from a model, and its weakness is a wrong frequency. For an undamped mode the ZV residue is $\lvert\cos(\pi r/2)\rvert$ of the unshaped one and the ZVD residue $\cos^2(\pi r/2)$, where r is the

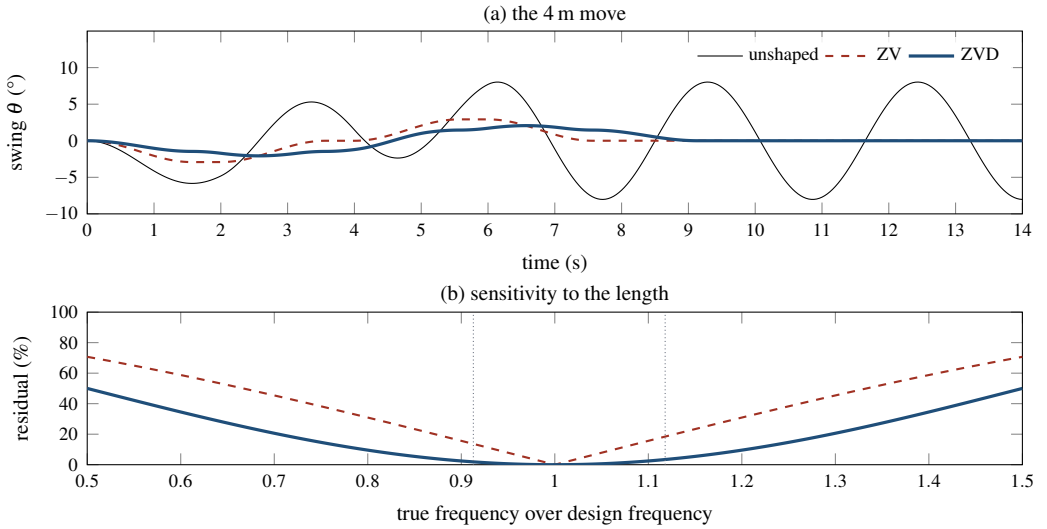


Figure 1.5 Input Shaping on the Hanging Cart. (a) The Swing During and After a 4 m Move, Unshaped (Thin Solid), With the Two-Impulse ZV Shaper (Dashed) and With the Three-Impulse ZVD Shaper (Thick Solid). (b) Residual Swing Against the Ratio of True to Design Frequency for ZV (Dashed) and ZVD (Thick Solid). The Dotted Lines Are a Cable 20 % Longer and 20 % Shorter Than the Design.

ratio of true to design frequency, as Figure 1.5(b) shows. A cable 20 % shorter than the design leaves 18.5 % of the unshaped swing with ZV and 3.4 % with ZVD. Keeping the residue below 5 % needs r within 0.968 to 1.032 for ZV and 0.856 to 1.144 for ZVD. The extra robustness costs one more half period of delay.

Chapter 12 applies both shapers on a crane whose true sway frequency sits 3.3 % above the one its shaper assumes, and measures what an assumed cable length costs. Chapter 11 treats the body bounce and wheel hop as two resonances and shows that at an invariant point near wheel hop no strut force changes the body acceleration. Chapter 13 carries one tool bending mode, and Chapter 5 shapes its reference as a trapezoid with an angle feedforward.

1.4.9 Limit Cycles

A limit cycle is an isolated periodic orbit. Nearby trajectories approach it or leave it, so its amplitude is set by the plant, not by the initial state [1]. Two mechanisms produce the limit cycles of this volume.

The first is a Hopf bifurcation, where a parameter moves a pair of eigenvalues across the imaginary axis. On the arm with $K = [10, 5]$, the actuator lag is such a parameter. At $\tau = 0.45\,\mathrm{s}$ the oscillatory pair is at $-0.111 \pm 2.58\,\mathrm{j}$, at 0.5 s it is at $\pm 2.45\,\mathrm{j}$, and at 0.55 s it is at $+0.091 \pm 2.33\,\mathrm{j}$, exactly as Equation (1.9) predicts. The linear loop then grows until a nonlinearity, such as an actuator limit, bounds it or loses the plant.

The second is a switching input with a delay. Let the arm's motor be a relay that applies $u = \pm 2$ against the sign of σ, measured τ_d late. Without the delay the relay drives σ to zero and holds it there, a sliding mode [11], and the tilt then decays with the mirrored pole at -2. With the delay, σ overshoots each switch, and Equation (1.5) gives the cycle exactly, as shown in Equation (1.10).

$$\sigma_{\max} = \frac{u_{\max}}{p}\left(\mathrm{e}^{p\tau_d} - 1\right), \qquad \frac{T_{\mathrm{cyc}}}{2} = \tau_d - \frac{1}{p}\ln\left(2 - \mathrm{e}^{p\tau_d}\right) \tag{1.10}$$

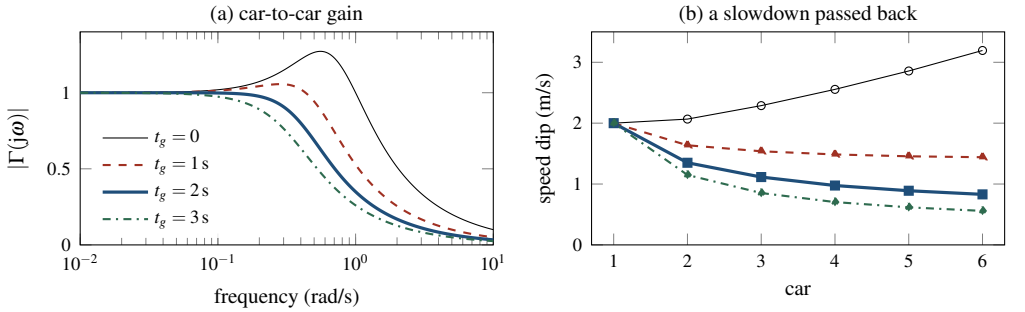


Figure 1.6 String Stability of Six Carts With $k_p = 0.5/\mathrm{s}^2$ and $k_d = 1/\mathrm{s}$. (a) The Car-to-Car Gain for Constant Spacing (Thin Solid) and Time Gaps of 1 (Dashed), 2 (Thick Solid) and 3 s (Dash-Dotted). (b) The Speed Dip of Each Car When the Lead Slows From 10 to 8 m/s and Back, With the Same Line Styles and Markers.

At $\tau_d = 0.1\,\mathrm{s}$ the cycle has $\sigma_{\max} = 0.221\,\mathrm{rad/s}$ and a period of 0.450 s, and the nonlinear arm gives 0.222 and 0.452 s with a tilt swing of 0.72°. The cycle exists only while $p\,\tau_d < \ln 2$, or $\tau_d < 0.347\,\mathrm{s}$. At 0.36 s the arm falls.

The describing function replaces the relay by its gain on the first harmonic of a sinusoid, $4u_{\max}/(\pi a)$ at the amplitude a, and finds the frequency where the linear part's phase is $-180°$ [24]. Here it predicts a period of 0.439 s, 2.6 % short, and an amplitude of σ of 0.176 rad/s, 20 % low, because σ is a train of exponentials rather than a sinusoid.

Chapter 2 locates a Hopf point on its hot branch and measures a relaxation oscillation with the jacket held just above the ignition fold. Chapter 4 finds that its sliding-mode law becomes a relay as its boundary layer shrinks, with a limit cycle the describing function puts at 15.6 Hz. Chapter 9 meets a saturated limit cycle of a fixed PD on a speed ramp. In Chapter 13 a delay turns the tool mode into chatter, and the tool leaving the cut bounds it to a limit cycle [25].

1.4.10 String Stability

Put six of the example's carts on one long rail, pendulums clamped, each following the cart ahead. Each follower sets its acceleration from the gap error and the gap's opening rate. With a time gap t_g the wanted gap grows with the follower's own speed, and the ratio of a follower's acceleration to its predecessor's is Equation (1.11), the form that Chapter 14 derives.

$$\Gamma(s) = \frac{k_d s + k_p}{(1 + k_d t_g)s^2 + (k_d + k_p t_g)s + k_p}, \qquad |\Gamma(j\omega)|^2 \approx 1 + \omega^2\left(\frac{2}{k_p} - t_g^2\right) \tag{1.11}$$

The line is string stable when $|\Gamma(j\omega)| \le 1$ at every frequency, so that no disturbance grows from car to car [19]. Because a lasting speed change passes whole, $\Gamma(0) = 1$, and the low-frequency expansion of Equation (1.11) needs $t_g \ge \sqrt{2/k_p}$.

With $k_p = 0.5/\mathrm{s}^2$ and $k_d = 1/\mathrm{s}$, Figure 1.6 shows the four cases. Constant spacing, $t_g = 0$, gives each follower a damping ratio of 0.707 and the line a peak gain of 1.272 at 0.556 rad/s. Every follower is well tuned and the line is not: when the lead slows by 2 m/s and recovers, the sixth car slows by 3.19 m/s. With $t_g = 1\,\mathrm{s}$ the peak falls to 1.057, a slow wave of 22 s period that grows by 5.7 % a car, too little to show in six cars and a factor of 3 over twenty. The bound $\sqrt{2/k_p} = 2\,\mathrm{s}$ is the shortest string-stable gap, and with it the sixth car slows by only 0.83 m/s. For constant spacing no linear law that sees only the car ahead is string stable [26].

Chapter 14 computes Γ from its cars' own law with lags, delays and a radio, adds the predecessor's broadcast acceleration, and measures the amplification along a nonlinear line of drivers and automated

Table 1.2 The Difficulty of Each Chapter, What Holds the Plant and the Margin Measured

Ch.	What makes it hard	What holds it	Margin measured
2	Two folds, an unstable middle branch, a negative steady gain	Reverse-acting PI, predictive control on the linearization	Largest setpoint step kept below the alarm
3	Stiffness that changes sign at horizontal	Feedforward with a schedule, or feedback linearization	Phase margin and least damping against the setpoint
4	Slip set at the friction peak, where the gain reverses	PI on slip, a sliding-mode law	Phase margin at the setpoint as the car slows
5	Two integrators and a lag, no damping	Filtered derivative, observer LQR	Phase margin, growth per swing
6	One real unstable pole	PID with a derivative longer than the coil lag	Gain window between the floor and the amplifier ceiling
7	Unstable root and a wrong-way zero	Cascade or LQR reading the angle	Gain window against delay, largest push
8	Two unstable modes, two RHP zeros	LQR on a Kalman filter	Longest stable sample interval, catchable tilt
9	Stability that depends on speed	Lean PD or LQR, scheduled on speed	Speed window, growth inside a band
10	Two unstable roots, a saturated tire	Two-input LQR scheduled on speed and friction	Recovery from a kick, a wet patch
11	Two resonances, invariant points	Skyhook and groundhook blend, preview	Tire load swing against liftoff
12	Lightly damped swing	Input shaping on a PID position loop	Residual swing against a wrong length
13	Delay that turns a tool mode into chatter	Adaptive feed, spindle speed selection	Depth against the stability lobe
14	Disturbances that grow along a line	Time gap, cooperative following	Peak car-to-car gain, shortest stable gap

cars. The criterion applies to any cascade of loops that passes a disturbance along, which Volume 6 meets in a supply chain.

1.4.11 What Stabilized Each Plant

The sections above give each plant two questions. What makes it hard: an unstable pole, a RHP zero, a sign change, a resonance, a delay or a chain? What holds it, and by how much: which structure stabilizes it, and which margin runs out first? The example answered both with numbers, and Table 1.1 showed that the answer depends on which margin is asked. Table 1.2 lists the difficulty of each chapter, the structure that holds its plant and the margin the chapter measures, as groundwork for the comparison in Chapter 15.

> ✦ DID YOU KNOW? **More Than a Hundred Controllers in One Car**
>
> A modern car has more than 100 electronic control units. Computer control entered cars in the late 1970s to meet stringent emission requirements, and it then took on suspension control, anti-lock braking, electronic braking and electronic stability control [6], the subjects of Chapter 4, Chapter 11 and Chapter 10.

1.5 Instruments and Final Elements in This Volume

A stabilizing loop acts only on what its sensors report and only as hard as its final element allows. Figure 1.7 lists what the controllers of each chapter read and what they move. An angle or position encoder appears in seven chapters, and an electric motor or servo drive is the final element in seven. Two chapters stand apart. The controller of Chapter 2 reads a temperature in a thermowell and moves a jacket temperature that a slave loop sets through utility valves, and those of Chapter 14 read a radar and a radio and move the throttle and brakes.

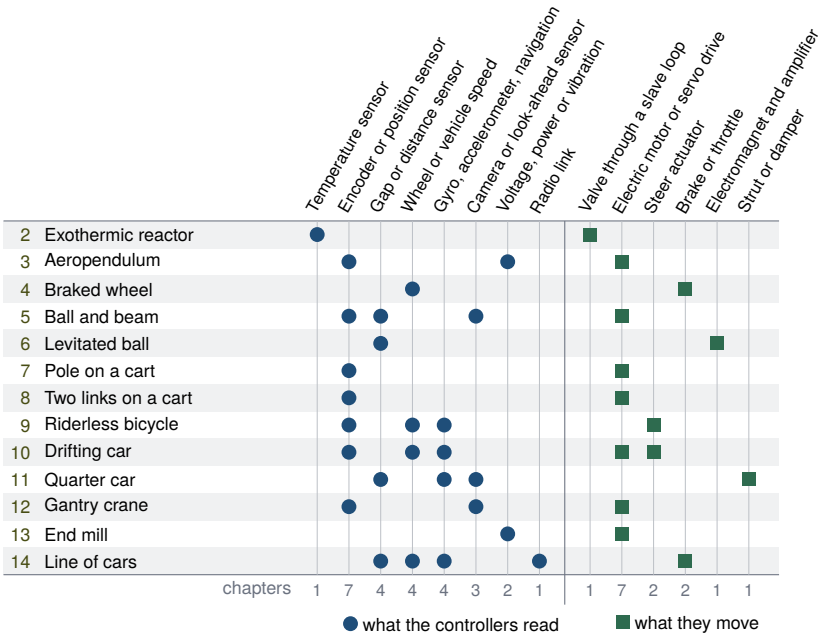


Figure 1.7 What the controllers of each chapter read (round marks) and move (square marks), from each chapter's variables table and technology section. A computed value, such as a slip, is not marked, and the crane's swing sensor serves only its predictive controller. Notice that encoders and electric drives carry the unstable mechanical plants.

An incremental encoder gives two pulse trains about a quarter of a period apart, and the one that leads tells the direction of travel [27]. It reports no speed, so the rate that a state feedback needs is differenced or estimated, the comparison of Figure 1.4.

The drives are mostly electric. A variable frequency drive sets the speed of an alternating-current motor by the frequency of the power it feeds and the torque by the voltage and current, and it can brake the machine by running the motor as a generator [27], as Figure 1.8 shows on a pump. The trolley of Chapter 12 is pulled by such a motor. Most chapters model their drive as a first-order lag with a limit, and several add a slew limit. That limit, not the controller, bounds the tilt that can be caught, as Section 1.4.7 showed on the arm. The one process plant of the volume reads a temperature, top of the figure, and Chapter 2 puts the lag of its thermowell at tens of seconds.

The laws sit in layers. Supervisory and lower-level loops execute with cycle times from milliseconds to minutes, and programmable logic controllers serve as local controllers under a supervisor in larger systems [28]. Figure 1.9 places the loops of this volume on their platforms. The laws of the mechanical plants run every 1 ms to 10 ms, the platoon's spacing law every 50 ms and the reactor's PI loop every 3 s. The drive loops beneath them run faster still, at 1 to 2 kHz in the speed controller of a multirotor's rotor (Chapter 3).

1.6 The Simulator Framework

Every chapter of this book has an app that runs in a web browser, collected with the others at https://apmonitor.com/apps/. About 240 apps exist across the series. All but a few are built on one shared framework, so the engine, the PID algorithm, the trends and the data export behave the same way in every chapter. The exceptions are standalone studios, such as the cascade studio and the two Temperature Control Lab apps, which follow the same ideas in their own code. This section describes the framework once and states where the standalone apps differ. Table 1.3 lists what a reader meets.

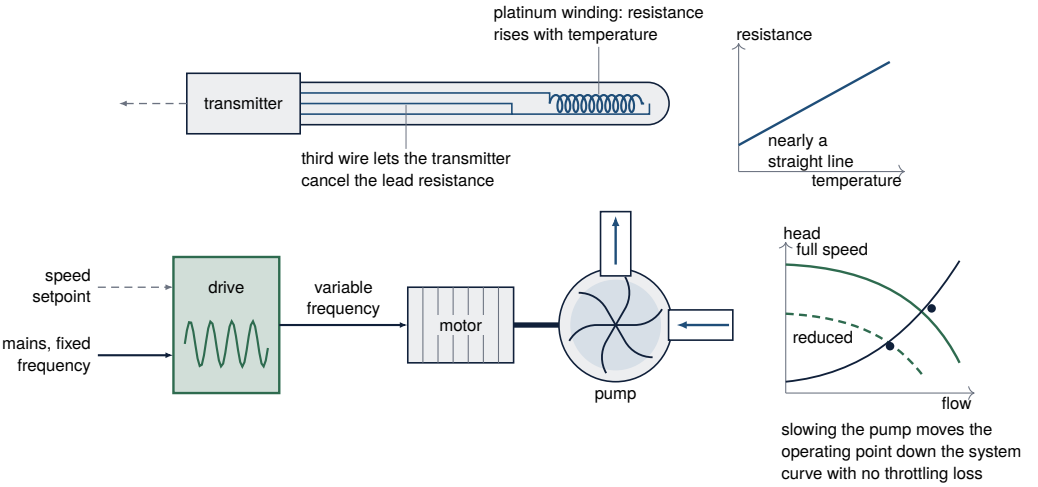


Figure 1.8 Top: principle of a resistance thermometer. Bottom: principle of a variable-speed drive on a centrifugal pump. Notice that the thermometer's resistance follows the temperature nearly in a straight line, and that the drive moves the machine onto a new curve by its speed alone, the way the trolley and spindle drives of this volume take a speed or torque command.

Layer	What it does in this volume	Where it runs	How often
Supervision or planning	a platoon's spacing supervisor, a chatter watch and speed rule, a predictive controller	**Supervisory level** of a machine controller; a **prototyping computer** on the vehicle bus	50 ms in the platoon; a predictive controller in its law's slot
Stabilizing law	state feedback, PID or PD with feedforward, a schedule, a sliding-mode law	**Microcontroller**, **electronic control unit**, **PLC**, **real-time computer**; a **DCS** for the reactor	1 ms to 10 ms on the machines and cars; 3 s on the reactor
Drive and inner loops	a rotor's speed loop, a servo or torque loop, a coil's current, a jacket's slave loop	**Electronic speed controller**, **servo drive**, **frequency inverter**, **current amplifier**	1 ms in the torque beam; 1 to 2 kHz in a rotor's speed controller
Sensors and final elements	encoders, gyros, accelerometers, a radar; motors, a steer, brakes, a strut, a coil	wires, a slip ring, a vehicle bus and a radio link	a camera frame every 33 ms; a radio broadcast every 100 ms proposed

Figure 1.9 The control layers of this volume, the platform that hosts each and how often each acts, from [6, 28] and the chapters' technology sections. Notice that only the reactor and the platoon act more slowly than every 10 ms.

Table 1.3 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.

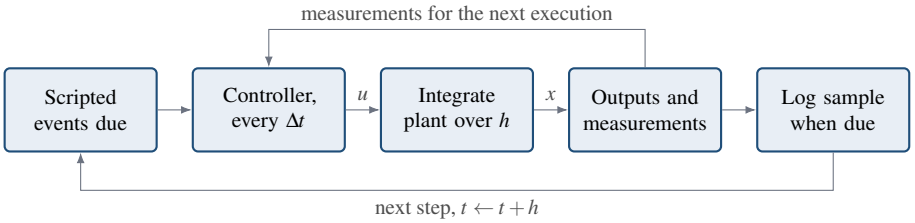


Figure 1.10 Work of the Engine in One Integration Step

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.

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 (1.12) to (1.15) for execution k and control period Δt.

$$e_k = a\,(\mathrm{SP}_k - \mathrm{PV}_k), \qquad P_k = K_c\,e_k, \qquad D_k = -a\,K_c\,\tau_D\,\frac{\mathrm{PV}^f_k - \mathrm{PV}^f_{k-1}}{\Delta t} \tag{1.12}$$

$$S_k = S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \tag{1.13}$$

$$v_k = u_b + S_k + I_{k-1}, \qquad u_k = \mathrm{sat}(v_k) \tag{1.14}$$

$$I_k = I_{k-1} + \left[\frac{K_c}{\tau_I}\,e_k + \frac{u_k - v_k}{T_t}\right]\Delta t \tag{1.15}$$

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 (1.15) drives I toward the value that reproduces the output applied, which is back-calculation [29]. With no integral action, Equation (1.15) 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 (1.16).

$$x_{k+1} = A\,x_k + B\,u_k + E\,w_k + c, \qquad y_k = C\,x_k \tag{1.16}$$

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 (1.17) over a horizon of N steps.

$$J = \sum_{k=1}^{N} (y_k - r_k)^{\top} Q\,(y_k - r_k) + \sum_{j} \Delta u_j^{\top} R\,\Delta u_j + \sum_{i} \rho_i\, s_i^2 \tag{1.17}$$

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.

A gain that changes with the point	An unstable equilibrium held by feedback		A resonance or a delay	A chain of loops
Sections 1.4.2, 1.4.3	Sections 1.4.4 to 1.4.7		Sections 1.4.8, 1.4.9	Section 1.4.10
(2) folds and a negative steady gain	(5) two integrators and a lag	(8) two unstable modes no loop pair holds	(11) two resonances, one strut force	(14) disturbances that grow along a line
(3) a stiffness that changes sign	(6) one unstable pole, a capped gain	(9) stability set by speed	(12) a lightly damped swing	
(4) slip at the friction peak	(7) an unstable root and a wrong-way zero	(10) an unstable drift with a saturated tire	(13) a delay that makes a tool mode chatter	

Figure 1.11 Method map of the volume: each chapter by the kind of difficulty of Section 1.1, with its lesson from Table 1.4 and the subsections it rests on. Notice that six of the thirteen chapters are unstable mechanical plants held by feedback.

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 15.

1.8 Chapter Roadmap

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

Chapters 3, 5, 6 and 12 form a first course: a sign-changing stiffness, a double integrator, one unstable pole and one resonance, each held by a PID-type loop. They rest on Sections 1.4.2, 1.4.3, 1.4.7 and 1.4.8. Chapters 7, 8 and 10 are best read in that order, after Sections 1.4.4 to 1.4.7, since each adds a state or an input to the last. Chapters 2 and 4 rest on Section 1.4.2 and can be read at any

Table 1.4 Roadmap of the Volume

Ch.	Plant	The lesson	Level	Time
2	Exothermic stirred-tank reactor	Folds, an unstable operating point and a negative steady gain held by a PI loop.	intermediate	3 h
3	Propeller arm on a pivot	A stiffness that changes sign. Feedforward, a schedule and feedback linearization.	introductory	2.5 h
4	Braked wheel of a car	Slip at the friction peak where the gain reverses. PI against a sliding-mode law.	intermediate	3 h
5	Ball on a tilting beam	Two integrators and a lag. Derivative as the only damping, an observer LQR.	introductory	2.5 h
6	Ball under an electromagnet	One unstable pole. A derivative longer than the lag, a gain capped by the amplifier.	introductory	2.5 h
7	Pole on a cart	An unstable root and a wrong-way zero. Cascade, LQR and the recoverable push.	intermediate	3 h
8	Two links on a cart	Two unstable modes no loop pair holds. LQR on a Kalman filter, the sampling margin.	advanced	2.5 h
9	Riderless bicycle	Stability set by speed. Countersteering as a RHP zero, a scheduled LQR.	intermediate	2.5 h
10	Drifting car on a skidpad	An unstable equilibrium with a saturated tire. A two-input LQR, scheduled.	advanced	2 h
11	Quarter car	Two resonances and invariant points. Comfort against road holding through one force.	intermediate	2.5 h
12	Gantry crane	A lightly damped swing. Input shaping and its sensitivity to the cable length.	introductory	3 h
13	End mill cutting a pocket	A delay that makes a tool mode chatter. Stability lobes and the spindle speed.	intermediate	2 h
14	Line of cars	String stability. Time gap, cooperation and delay along a line.	intermediate	2.5 h

point. Chapters 9, 11, 13 and 14 each stand alone after the subsection they use. A reader who has taken a course in state-space control can take a shorter path through Chapters 2, 8, 10, 13 and 14.

KEY POINTS

- A plant that would fall, run away or ring is held only near the point its controller was designed for, and only while the actuator can push hard and fast enough.
- A linearization describes the plant near one point. A gain schedule or feedback linearization keeps one tuning valid across points, and both need the model to be right.
- An unstable pole sets limits that no gain removes: actuator headroom, a delay shorter than the pole's time constant, a lag shorter than the derivative time.
- No single margin says how robust a stabilizing loop is, and a line of well-tuned loops can amplify a disturbance.

Where else this applies. A gain schedule on a gain that moves by four orders of magnitude in Volume 2, Chapter 6, a limit cycle past the peak of a compressor characteristic in Volume 4, Chapter 2, and a lander whose position loop meets a right-half-plane zero in Volume 3, Chapter 10.

References

[1] S. H. Strogatz, *Nonlinear Dynamics and Chaos*. CRC Press, 2018.

[2] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. John Wiley & Sons, 2005.

[3] A. E. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. Routledge, 2018.

[4] J. P. Meijaard, J. M. Papadopoulos, A. Ruina, and A. L. Schwab, "Linearized dynamics equations for the balance and steer of a bicycle: A benchmark and review," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2084, pp. 1955–1982, 2007.

[5] M. H. Tuttle, R. A. Kilgore, and R. P. Boyden, "Magnetic suspension and balance systems: A selected, annotated bibliography," NASA Langley Research Center, Hampton, Virginia, NASA Technical Memorandum TM-84661, Jul. 1983. [Online]. Available: https://ntrs.nasa.gov/citations/20080047362

[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] O. J. M. Smith, "Posicast control of damped oscillatory systems," *Proceedings of the IRE*, vol. 45, no. 9, pp. 1249–1255, 1957.

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

[9] W. Levine and M. Athans, "On the optimal error regulation of a string of moving vehicles," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 355–361, 1966.

[10] A. Uppal, W. H. Ray, and A. B. Poore, "On the dynamic behavior of continuous stirred tank reactors," *Chemical Engineering Science*, vol. 29, no. 4, pp. 967–985, 1974.

[11] V. Utkin, "Variable structure systems with sliding modes," *IEEE Transactions on Automatic Control*, vol. 22, no. 2, pp. 212–222, 1977.

[12] J. Hauser, S. Sastry, and P. Kokotović, "Nonlinear control via approximate input-output linearization: the ball and beam example," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 392–398, 1992.

[13] K. J. Åström and K. Furuta, "Swinging up a pendulum by energy control," *Automatica*, vol. 36, no. 2, pp. 287–295, 2000.

[14] R. Tedrake, I. R. Manchester, M. Tobenkin, and J. W. Roberts, "LQR-trees: Feedback motion planning via sums-of-squares verification," *The International Journal of Robotics Research*, vol. 29, no. 8, pp. 1038–1052, 2010.

[15] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[16] D. G. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.

[17] R. E. Kalman, "When is a linear control system optimal?" *Journal of Basic Engineering*, vol. 86, no. 1, pp. 51–60, 1964.

[18] J. T. Bosworth, "Success stories in control: Nonlinear dynamic inversion control," NASA Armstrong Flight Research Center, report DFRC-E-DAA-TN1713, preprint, NASA Technical Reports Server 20110008197, https://ntrs.nasa.gov/citations/20110008197, 2010.

[19] D. Swaroop and J. K. Hedrick, "String stability of interconnected systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 349–357, 1996.

[20] G. Stein, "Respect the unstable," *IEEE Control Systems Magazine*, vol. 23, no. 4, pp. 12–25, 2003.

[21] K. J. Åström, "Limitations on control system performance," *European Journal of Control*, vol. 6, no. 1, pp. 2–20, 2000.

[22] 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.

[23] 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.

[24] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, vol. 20, no. 5, pp. 645–651, 1984.

[25] J. Tlusty and F. Ismail, "Basic non-linearity in machining chatter," *CIRP Annals*, vol. 30, no. 1, pp. 299–304, 1981.

[26] P. Seiler, A. Pant, and K. Hedrick, "Disturbance propagation in vehicle strings," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1835–1841, 2004.

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

[28] 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.

[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 Theory and calculation · 4 min

The arm of Equation (1.1), $\ddot{\theta} = -4\sin\theta + u$ with θ from hanging, is held by the constant input $u_0 = 2\sqrt{3} = 3.46\,\mathrm{rad/s^2}$. Find both steady states between 0 and 180°, the steady-state gain $\mathrm{d}\theta_0/\mathrm{d}u_0$ at each, and the eigenvalues of the linearization at each.

1.2 Concept check · 2 min

The arm is to be held at 150°, where its steady-state gain is -0.289 rad per rad/s^2. The PD law with gravity feedforward of Equation (1.3) is used, with $k_d = 4$. Which statement is right?

(a) k_p and k_d must both change sign, because the steady-state gain is negative

(b) only k_p changes sign, and k_d keeps its sign because the rate response is unchanged

(c) no gains hold the arm there, because the plant is unstable in open loop

(d) both keep their signs, and k_p must exceed 3.46 to keep the total stiffness positive

1.3 Theory and calculation · 6 min

The PD law with gravity feedforward of Equation (1.3) holds the arm with $k_p = 8$ and $k_d = 4$.

(a) Write the closed-loop polynomial at a setpoint of 120° and find its natural frequency and damping ratio. *Checkpoint:* $s^2 + 4s + 6$, 2.45 rad/s, 0.816.

(b) With $k_p = 2$ instead, find the largest setpoint angle the loop holds. *Checkpoint:* 120°.

(c) A schedule keeps the polynomial $s^2 + 4s + 12$ that $k_p = 8$ gives hanging. Find the scheduled k_p at 120°. *Checkpoint:* 14.

1.4 Theory and calculation · 5 min

For the upright arm, $\ddot{\varphi} = 4\varphi + u$, an LQR uses the weights $Q = \mathrm{diag}(65, 10)$ and $R = 1$. Use the closed form of Equation (1.7) to find K and the closed-loop poles. Then find the lowest factor by which K can be scaled down before the arm falls, and the longest actuator lag the unscaled loop tolerates by Equation (1.9).

1.5 Theory and calculation · 4 min

An observer of Equation (1.8) for the upright arm reads the angle only, $C = [1, 0]$. Find the gain $L = [l_1, l_2]$ that places both observer poles at -10, and list the closed-loop eigenvalues when the estimate feeds $K = [10, 5]$.

1.6 Concept check · 3 min

The upright cart of the running example has a drive that sets its acceleration, $\ddot{x} = \ell u$, and an encoder on the cart's position but none on the pole. Which statement is right?

(a) no controller holds the pole, because the pole's motion never reaches the cart's encoder

(b) an observer with poles fast enough recovers the tilt from the cart's position

(c) an LQR with a heavy weight on the cart's position holds it

(d) it can be held, with a sensitivity peak of at least 46

1.7 Theory and calculation · 6 min

The upright arm has the unstable pole $p = 2/\mathrm{s}$.

(a) The placement gain $K = [20,\ 8]$ acts through an actuator lag τ. Find the longest lag it tolerates and the frequency of the oscillation that appears there. *Checkpoint:* 0.4 s, 4 rad/s.

(b) A proportional law acts on the unstable mode σ through a pure delay τ_d. Find the longest delay for which some gain stabilizes it, and the longest delay for which a relay of ± 2 on σ keeps a limit cycle instead of losing the arm. *Checkpoint:* 0.5 s and 0.347 s.

1.8 Theory and calculation · 5 min

A crane load hangs on a cable of 9.81 m, so that $\omega_n = \sqrt{g/\ell} = 1\,\mathrm{rad/s}$, with negligible damping. Give the impulse times and amplitudes of the ZV and ZVD shapers, the delay each adds to a move, and the residual swing each leaves, as a percentage of the unshaped one, if the cable is in fact 10 % longer.

1.9 Theory and calculation · 4 min

Followers in a line use the time-gap law of Equation (1.11) with $k_p = 2/\mathrm{s}^2$ and $k_d = 1/\mathrm{s}$. Find the shortest string-stable time gap from the low-frequency condition, and the sign of the coefficient of ω^2 in $|\Gamma|^2$ at $t_g = 0.6\,\mathrm{s}$. Say what that sign means for a slow wave passing down the line.

1.10 Concept check · 3 min

The motor of the upright arm is replaced by one with $u_{\max} = 3\,\mathrm{rad/s}^2$. Which tilts from rest can a stabilizing state feedback, clipped at the limit, recover?

(a) tilts up to 28.6°, the same as before, because the unstable pole has not changed

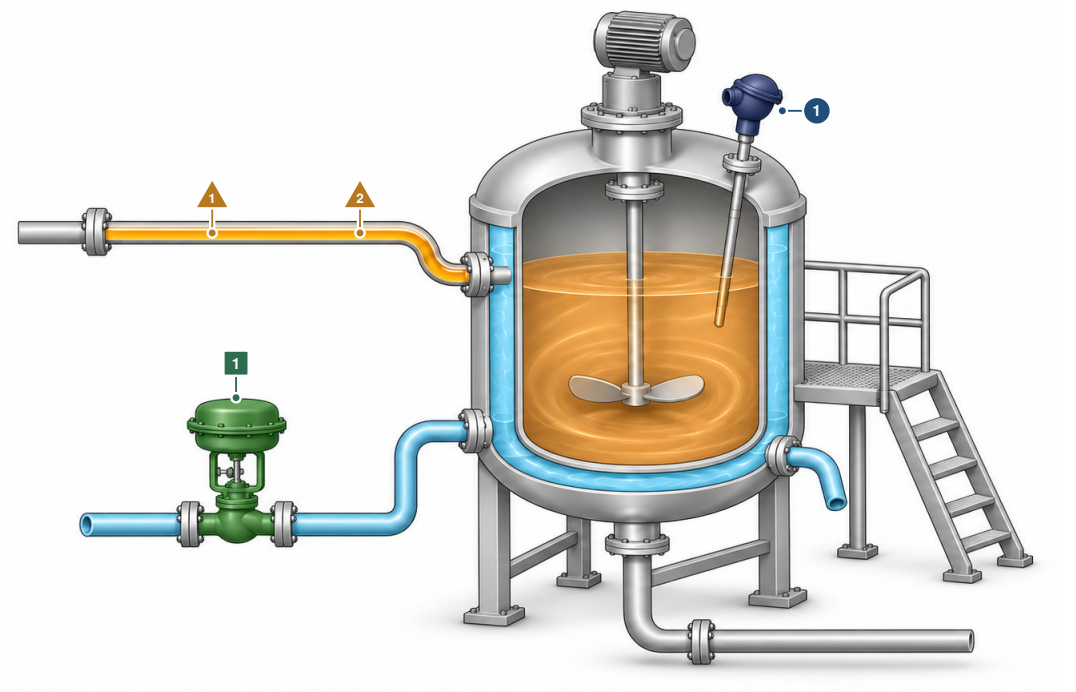
(b) any tilt, once the gain K is large enough

(c) up to 0.75 rad (43.0°) on the linear model and 48.6° on the nonlinear arm, at best

(d) up to 0.75 rad only if the LQR weight on the input is made small

CHAPTER 2

Exothermic Reactor Simulator for Temperature Control Across an Ignition Boundary



The controller moves 1 valve that sets the jacket temperature

The controller reads 1 temperature of the stirred liquid

What disturbs it 1 warmer feed 2 richer feed, more reactant in it

Notice that the jacket around the liquid is the controller's only way to remove the heat the reaction releases, and the probe in the liquid is all it reads. Illustration, not to scale.

IN PLAIN TERMS **A Reactor That Cannot Rest Where It Works**

The main idea. A reactor turns liquids into a chemical, and the reaction gives off heat that a cooling jacket around the tank carries away. The hotter the liquid gets, the faster it reacts and the more heat it gives off.

The challenge. At its usual working temperature the tank cannot stay put on its own, so a controller must keep moving the jacket. A target raised one step too far sends it past its alarm before the controller brings it back.

Why it matters. Real runaway accidents come from the same race between heat made and heat removed.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on reactive incidents (Section 2.1), the history in Section 2.2, the fact on the pendulum (Section 2.4.2), the two try-its in Section 2.7 and the key points
Study	65 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 2.7), Technology in Practice (Section 2.10), Limitations (Section 2.11) and the open challenge

The control question Heat generation in an exothermic reactor grows exponentially with temperature while jacket cooling grows only linearly, so the default setpoint is open-loop unstable and a jacket change of a few kelvin can ignite the vessel. How large a setpoint move can one fixed PI tuning make before the reactor runs past its alarm, and what does a predictive controller built on the linearized balances add?

Learning objectives

- Locate the steady states, the two folds and the Hopf point from the van Heerden construction and the steady-state locus.
- Evaluate the Jacobian at an operating point and classify each steady state from its eigenvalues.
- Explain why a reverse-acting loop holds a branch whose steady-state gain is negative.
- Identify the cold branch from a jacket step test and tell its chord gain from the tangent gain.
- Measure the largest safe setpoint step, and judge detuning, the predictive controller and the moving horizon estimator against it.

Prerequisites Sections 1.4.2 (folds), 1.4.3 (sign from the immediate response), 1.4.4 (eigenvalues), 1.4.9 (Hopf point) and 1.6.2 (velocity form). Section 1.6.3 is needed only for the predictive controller and its estimators.

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

See also Volume 9, Chapter 4 for an offset-free predictive controller whose Kalman filter carries an input bias, and Volume 9, Chapter 3 for a step test that returns a different plant at each operating point. Volume 13, Chapter 4 and Volume 13, Chapter 9 for ignition and branch multiplicity in chemical plants, and Volume 8, Chapter 2 for a runaway held off by dosing.

2.1 Introduction

Most of the plants a first control course meets have a process gain that can be written down once. A tank fills faster when the valve opens further, a heater raises a temperature in proportion to the power it draws, and the constant of proportionality changes slowly enough that a controller commissioned at one operating point still works at another. An exothermic reactor is not that kind of plant. The heat it releases carries the Arrhenius factor $e^{-E/RT}$, so generation climbs exponentially with the very variable that the cooling jacket is trying to hold, while removal through the jacket climbs only in proportion to the temperature difference across the wall [1]. Two curves of that shape can cross once, twice or three times. Where they cross three times the middle crossing is a steady state that no operator can hold without feedback [2, 3], and the region between the outer crossings is a set of operating temperatures that exist only while a controller holds them.

+ DID YOU KNOW? 167 Reactive Incidents, 108 Deaths

From January 1980 to June 2001 the United States Chemical Safety Board counted 167 serious incidents of uncontrolled chemical reactivity, 48 of which killed 108 people. It attributed 35 percent of them to runaway reactions, in which heat is generated faster than cooling media and surroundings can remove it [4], the crossing of the two curves in Figure 2.2(a).

The consequence for control is sharper than the usual warning about nonlinearity. The steady-state gain of the reactor of this chapter passes through $+1.72$ K/K at its nominal operating point, rises without bound at 335.65 K, reappears as -5.56 K/K a few kelvin higher, and returns to $+2.38$ K/K on the hot branch. A proportional gain that gives a well-damped response at low conversion is the same number that produces an excursion of more than 110 K at high conversion, and the excursion happens on the way to a setpoint that the loop then holds perfectly. Runaway in this plant is a property of the transient rather than of the destination.

Exothermic reactors have generated a literature of their own because of that behavior. Steady-state multiplicity and its bifurcation structure were mapped in detail for the single first-order reaction, together with the limit cycles that appear once the ignited state loses stability [5]. Control work has

treated the reactor as the standard hard case for input-output linearization [6], for robust stabilization by classical proportional-integral feedback [7], for gain scheduling across an operating envelope [8], and for constrained predictive control written specifically to keep a batch away from its runaway boundary [9], the problem Volume 8, Chapter 2 takes up for a semi-batch reactor whose dosing is the safety layer. Industrial practice adds a layer the theory usually leaves out, because the jacket temperature that these formulations manipulate is itself the setpoint of a slave loop on coolant flow or coolant mixing [10]. However, a student meeting the reactor for the first time in a course has no way to watch the ignition happen, to see the same controller succeed and then fail on a larger setpoint change, or to find the jacket temperature at which the cold operating point stops existing.

The model rests on two conservation statements over one well-mixed volume, a mole balance on the reactant and an energy balance whose exotherm term is nonlinear in the state that the balance itself determines. The dilution rate, the adiabatic temperature rise and the jacket rate constant are the only three lumped groups in the model, and each is a transport rate divided by a capacity, which is the way a first course builds a dynamic model [11, 12]. What this plant adds to that familiar structure is steady-state multiplicity, an open-loop-unstable branch, a process gain that changes sign, and an actuator that reaches both of its limits inside one transient. The behavior matters industrially because surveys of thermal-runaway incidents keep finding inadequate temperature control and inadequate knowledge of the heat release among the recurring causes [13, 14].

This chapter specifies an interactive exothermic stirred-reactor simulation in enough detail to rebuild it, and reports what its shipped constants actually do rather than what the textbook case is usually drawn as. Section 2.2 sets out the safety record and the cascade that frame the problem, and Section 2.3 gives the balances, the constants and the adaptive explicit integrator. The characterization in Sections 2.4 and 2.5 locates the two fold points, the Hopf point on the hot branch and the open-loop-unstable temperature interval, shows that the ignited branch at the published jacket temperature is not an attractor, and identifies the cold branch from a step test. Section 2.6 describes the PI loop, a linear model predictive controller, its two estimators and a relay experiment. Section 2.7 measures one controller at eight setpoints, where it changes character between two of them, and against two upsets that ignite the reactor in manual. The predictive controller is compared with the shipped loop in Section 2.8, and Section 2.9 hands it a reactor that has already run away, under two estimators of the concentration it cannot measure, while Section 2.10 describes the plant's instruments, valves and control system. The application's eleven guided cases each carry a completion requirement (Section 1.6.4). Sections 2.11 to 2.13 state the limitations and place the results, and Section 2.14 holds the exercises.

2.2 Background

The instrument that matters in an exothermic reactor is the one that reads the liquid temperature, and the history of reactor safety is largely the history of learning what that reading does and does not say. Early continuous processes were run with the operator judging the reaction from the jacket steam or cooling water and from the product, because a thermowell reading inside a stirred vessel lags the hottest part of the contents and gives no warning at all of reactant that has accumulated without reacting. Loss prevention emerged as a discipline in the chemical industry only after a sequence of events made the point that hazard is a property of the process rather than of the care taken by the people running it, and the systematic study of what had actually gone wrong began in the 1960s and 1970s [15].

Runaway reactions have kept their place in that record. A survey of thermal-runaway incidents investigated in the United Kingdom found that the causes clustered in a small number of repeated failures, including inadequate knowledge of the reaction chemistry and its heat release, inadequate temperature control, and inadequate agitation, rather than in exotic mechanisms [13]. A later review across a wider set of incidents reached the same conclusion and added that the majority arose during normal production rather than during commissioning or maintenance [14]. What both surveys

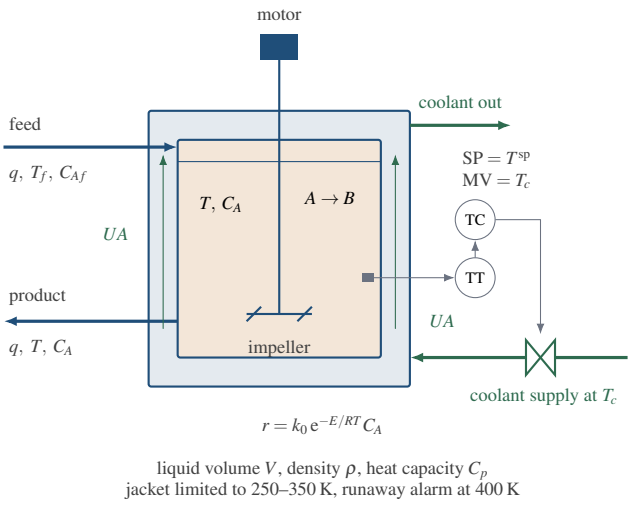


Figure 2.1 Jacketed continuous stirred-tank reactor as simulated. The temperature transmitter reads the well-mixed liquid and the controller sets the jacket temperature directly, so the coolant valve and its slave loop are treated as ideal.

describe is a plant whose heat generation exceeded its heat removal capacity for long enough that no control action could recover it, which is exactly the crossing of the two curves that the theory of autothermic operation had already described [1].

> ⧖ HISTORY **2007: Lost Cooling and a New Course Requirement**
>
> In December 2007 a 2450-gallon batch reactor making a gasoline additive in Florida burst and its contents exploded, killing four people and injuring 32. The United States Chemical Safety Board found that a runaway exothermic reaction had occurred and that a loss of sufficient cooling had most likely caused it. The cooling system lacked the redundancy to survive a single failure, and the root cause was that the company had not recognized the runaway hazard of the chemical it was making [16]. The board asked for reactive hazard awareness to be added to the requirements of chemical engineering degrees. In October 2011 the accreditation board for engineering programs approved changes that require proficiency in all chemical process hazards [17].

The control response to that reading developed alongside it. Reactor stability was placed on an analytical footing by linearizing the coupled mass and energy balances and examining the eigenvalues of the resulting matrix [2], and the extension to a controlled reactor followed immediately [3]. Modern reactor temperature control is structured as a cascade in which the master loop compares the reactor temperature against setpoint and asks for a jacket temperature, while a slave loop delivers that jacket temperature by mixing hot and cold utility or by throttling coolant [10]. The application described here implements the master loop and treats the jacket temperature as though the slave loop were perfect, which is the standard first approximation and the one that makes the thermal behavior visible without a second set of tuning constants.

2.3 Reactor Model

The simulated unit is a perfectly mixed, constant-volume, liquid-phase vessel with a cooling jacket, an agitator, one feed stream and one product stream drawn at the same volumetric rate. A single irreversible first-order reaction $A \to B$ runs in the liquid and releases heat. The arrangement, together with each symbol used in the balances that follow, is shown in Figure 2.1.

Table 2.1 Variable Classification, Units and Ranges

Symbol	Role	Unit	Range or default
T	CV, state	K	200 to 900 (clamped)
C_A	state, shown	mol/m^3	0 to 20 (clamped)
$\hat{T}$	measurement	K	T plus noise
T_c	MV	K	250 to 350
T^{sp}	setpoint	K	280 to 450, default 340
T_f	DV	K	300 to 380, default 350
C_{Af}	DV	mol/m^3	0.5 to 2, default 1
q	parameter	vol/min	50 to 200, default 100
r	derived	mol/m^3-min	Equation (2.1)
X	derived	–	$(C_{Af}-C_A)/C_{Af}$

2.3.1 Variables

The reactor temperature is the only controlled variable and the jacket temperature is the only manipulated variable, which makes this a single-input single-output plant with two states. The classification is given in Table 2.1.

2.3.2 Governing Equations

The reaction rate per unit volume follows the Arrhenius law with a first-order dependence on the reactant concentration, as shown in Equation (2.1).

$$r = k_0\, \mathrm{e}^{-E/(RT)}\, C_A \tag{2.1}$$

where r is the rate of consumption of A in mol/m^3-min, k_0 is the pre-exponential factor in 1/min, E/R is the activation energy divided by the gas constant in K, T is the liquid temperature in K and C_A is the reactant concentration in mol/m^3. The exponential is evaluated at $\max(T, 1\,\mathrm{K})$ so that the argument can never divide by zero.

A mole balance on A over the constant liquid volume gives Equation (2.2), and an energy balance on the same volume gives Equation (2.3).

$$\frac{\mathrm{d}C_A}{\mathrm{d}t} = \frac{q}{V}\left(C_{Af} - C_A\right) - r \tag{2.2}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{q}{V}\left(T_f - T\right) + \frac{m\Delta H}{\rho C_p}\, r + \frac{UA}{V\rho C_p}\left(T_c - T\right) \tag{2.3}$$

where q is the volumetric feed and product flow, V is the liquid volume, C_{Af} and T_f are the feed concentration and feed temperature, $m\Delta H$ is the heat released per mole reacted, ρ and C_p are the liquid density and heat capacity, UA is the product of the overall heat-transfer coefficient and the jacket area and T_c is the jacket temperature. The three terms of Equation (2.3) are the sensible load carried in by the feed, the exotherm and the jacket duty, and all three are reported in K/min.

Only three lumped groups appear in the balances. Their values at the default constants are the dilution rate $q/V = 1.0/\mathrm{min}$, the adiabatic temperature rise per unit of reactant $m\Delta H/(\rho C_p) = 209.2\,\mathrm{K}$ per mol/m^3, and the jacket rate constant $UA/(V\rho C_p) = 2.0921/\mathrm{min}$. Because the individual constants enter only through these groups, the published values are carried literally in the units in which they were published, and consistency of Equation (2.3) requires UA to be read as 5×10^4 J/(min K) rather than in watts.

Table 2.2 Model Parameters, Values and Visibility

Symbol	Meaning	Value and unit	Adjustable
q	feed and product flow	100 vol/min (50–200)	yes
V	liquid volume	100 vol (50–200)	yes
ρ	liquid density	1000 (800–1200)	yes
C_p	liquid heat capacity	0.239 (0.15–0.4)	yes
$m\Delta H$	heat of reaction	5×10^4 J/mol (2×10^4–8×10^4)	yes
E/R	activation temperature	8750 K (7000–10000)	yes
k_0	pre-exponential factor	7.2×10^{10} 1/min (10^{10}–10^{11})	yes
UA	jacket UA	5×10^4 J/min-K (2×10^4–10^5)	yes
σ	transmitter noise	0.15 K (0–1)	yes
T_f	feed temperature	350 K (300–380)	upset
C_{Af}	feed concentration	1 mol/m^3 (0.5–2)	upset
$T_{c,\min}$	jacket lower limit	250 K	no
$T_{c,\max}$	jacket upper limit	350 K	no
T_{alarm}	runaway alarm	400 K	no
h	integration substep	0.005 min	no
T_s	control period	0.05 min	no

The jacket duty and the single-pass conversion follow algebraically from the state, as shown in Equations (2.4) and (2.5).

$$Q = UA\,(T_c - T) \tag{2.4}$$

$$X = \frac{C_{Af} - C_A}{C_{Af}} \tag{2.5}$$

where Q is negative whenever heat leaves through the jacket and X is clipped to the interval 0 to 1. Neither is displayed by the application, which shows and exports the reactant concentration C_A instead.

The temperature transmitter adds a zero-mean Gaussian error to the state, as shown in Equation (2.6).

$$\hat{T}_k = T_k + \sigma\,\varepsilon_k, \qquad \varepsilon_k \sim \mathcal{N}(0,1) \tag{2.6}$$

where $\hat{T}_k$ is the reading the controller acts on and σ is the transmitter standard deviation. One deviate is formed at each controller execution by the Box–Muller transform of two draws from a seeded generator, which restarts from the same seed whenever a case is loaded or reset, so a run repeats exactly at any value of σ. The controller, the recorded trend and the exported file all act on $\hat{T}$, while the noise-free state T is retained separately so that a measured excursion can be separated from its noise. The runaway alarm compares the noise-free T against 400 K and is shown only while the temperature is above it, with no latch, no acknowledgement and no action on the plant.

2.3.3 Parameters and Initial Conditions

Every constant is listed in Table 2.2. Nine of them are adjustable model parameters, the feed temperature and feed concentration are operator-set upsets, and the rest are fixed. A changed parameter or upset reaches the model at the next controller execution, and the feed flow is a parameter rather than an upset.

Every case starts at steady operation. The four manual cases start from the steady state that the default constants produce with the jacket held at 300 K, namely $C_A = 0.87725\,\text{mol/m}^3$ and $T = 324.475\,\text{K}$, with the reaction running at 0.1227 mol/(m^3 min) and a single-pass conversion of 12.27 %. The seven automatic cases start at the steady state of their 340 K setpoint, with $C_A = 0.67598\,\text{mol/m}^3$ and the jacket at the 302.818 K that holds it. Loading a case or pressing Reset solves that start again for the constants, upsets, setpoint and manual jacket temperature then in force, the learner's changes included: in automatic it is the steady state of the setpoint, and in manual the stable steady state that the manual jacket temperature holds. A manual jacket between the ignition fold and the Hopf point of

Section 2.4 holds no stable state, so the reactor then starts from the cold state of the 300 K jacket and ignites into the relaxation cycle that the same jacket gives when it is set after Run, between 332.07 K and 466.7 K at 304 K. A setting chosen before Run is therefore part of the starting state, and it gives the same run as the same setting carried through Reset. Reset keeps every setting the learner has changed, while reloading the case restores the case's own settings.

2.3.4 Numerics

The right-hand side is integrated with explicit Euler on an outer substep of $h = 0.005\,\text{min}$. The rate constant $k_0 \mathrm{e}^{-E/RT}$ grows by three orders of magnitude across the temperature range the model permits, so a fixed step that is comfortable on the cold branch is unstable on the hot one. The integrator therefore subdivides its own step against the fastest decay rate in the system, as shown in Equation (2.7).

$$\begin{aligned} \lambda &= \frac{q}{V} + \frac{UA}{V\rho C_p} + k_0 \mathrm{e}^{-E/(RT)} \\ n &= \operatorname{clamp}\left(\left\lceil \frac{h\lambda}{0.2} \right\rceil,\ 1,\ 8000 \right) \end{aligned} \tag{2.7}$$

where n is the number of inner Euler updates taken over h and 0.2 is the stability margin on $h\lambda$. At the nominal operating point $\lambda = 3.23/\text{min}$ and $n = 1$, so the update is plain Euler. At 450 K the rate constant alone is 258.7/min, so $n = 7$ and the integrator subdivides. The state is clamped to $C_A \in [0,\ 20]\,\text{mol/m}^3$ and $T \in [200,\ 900]\,\text{K}$ after every inner update, which is a numerical guard rather than a physical constraint, because the adiabatic ceiling $T_f + (m\Delta H/\rho C_p)C_{Af}$ is 559 K at the default feed and 768 K at the largest feed concentration the model allows.

Each browser frame advances the simulated clock by the elapsed wall time multiplied by the speed setting, from 0.5 to 10 min of simulated time per second, in whole substeps of h. The controller executes once at $t = 0$ and then at the start of every tenth substep, so the control period is $T_s = 0.05\,\text{min}$ of simulated time, or 3 s. Because the controller counts substeps rather than wall-clock milliseconds, the closed-loop trajectory does not depend on the speed setting or on browser timer jitter. The historian records one row every 0.1 min, every second controller execution, and the export is a comma-separated file of thirty-three columns, eleven holding time, mode, case, setpoint, measured and noise-free reactor temperature, jacket temperature, a saturation flag, the two feed upsets and the reactant concentration, eight more for the predictive controller of Section 2.6.3, and fourteen for the estimator choice and the relay experiment of Sections 2.6.4 and 2.6.5. A row stamped t carries the noise-free temperature at t and the measured temperature that the controller execution at t acts on, while its jacket temperature is the one held over the control period that ended at t, so a jacket move made by an execution first appears on the next row.

2.3.5 Assumptions

Perfect mixing. The liquid is one lumped temperature and one lumped concentration. The assumption fails wherever the agitator is undersized or the viscosity rises with conversion, which is where real runaways start, because a hot pocket reacts faster than the bulk and the thermowell never sees it. The simulator cannot produce that failure mode at all.

Constant volume and constant physical properties. Density and heat capacity do not change with temperature or composition, and the product flow equals the feed flow at all times. Over the 200 K span the model permits, a real liquid changes ρC_p by tens of percent and would change the thermal time constant with it.

Single irreversible first-order reaction. There is no reverse reaction, no side reaction and no decomposition path. Decomposition at elevated temperature is the mechanism that turns many industrial runaways from an excursion into an explosion, and its absence is why the reactor in this simulation always comes back.

Ideal jacket. The jacket temperature is whatever the controller asks for, instantly and uniformly, within its limits. A real jacket has a coolant residence time, a fouling resistance that drifts, and a slave loop with its own dynamics and its own saturation [10]. The consequence is that the manipulated variable here carries no lag of its own, so the only dynamics in the loop are the reactor dynamics.

Ideal measurement. The transmitter is instantaneous and carries white noise only. A thermowell has a time constant of tens of seconds, which is the same order as the reactor time constant identified in Section 2.5, and adding it would change the achievable tuning.

Liquid phase only. There is no vapor space, no pressure state and no relief path. Nothing in the model represents what happens above the boiling point, so temperatures reported above roughly 400 K should be read as a measure of how far the energy balance has run away rather than as a prediction.

2.4 Steady-State Multiplicity and Stability

Setting Equation (2.2) to zero solves for the concentration at any temperature, because the mass balance is linear in C_A, as shown in Equation (2.8).

$$C_A^{ss}(T) = \frac{(q/V)\,C_{Af}}{q/V + k_0 \mathrm{e}^{-E/(RT)}} \tag{2.8}$$

Substituting Equation (2.8) into the energy balance splits Equation (2.3) into a generation term and a removal term, as shown in Equation (2.9).

$$\begin{aligned} Q_{\mathrm{gen}}(T) &= \frac{m\Delta H}{\rho C_p}\, k_0 \mathrm{e}^{-E/(RT)}\, C_A^{ss}(T) \\ Q_{\mathrm{rem}}(T, T_c) &= \frac{q}{V}\left(T - T_f\right) + \frac{UA}{V\rho C_p}\left(T - T_c\right) \end{aligned} \tag{2.9}$$

where both are expressed in K/min so that a steady state is a point at which the two are equal. This is the construction introduced for autothermic operation [1] and it is drawn for the shipped constants in Figure 2.2(a).

2.4.1 Steady States and the Fold Points

The generation curve is sigmoidal rather than purely exponential, because the reactant runs out. Below 330 K the rate constant is small enough that C_A^{ss} stays near C_{Af} and Q_{gen} doubles for roughly every 8 K, while above 380 K conversion exceeds 87 % and the curve flattens against the adiabatic ceiling of 209.2 K/min. The removal line has slope $q/V + UA/(V\rho C_p) = 3.0921\,/\mathrm{min}$ and shifts left as the jacket warms. Three intersections occur wherever the removal line cuts the steep middle part of the generation curve, as shown in Figure 2.2(a).

Solving Equation (2.3) for T_c instead of for T removes the multiplicity, because the jacket temperature that holds a given reactor temperature is unique, as shown in Equation (2.10).

$$\begin{aligned} T_c^{ss}(T) = T - \frac{V\rho C_p}{UA}\Bigg[&\frac{q}{V}\left(T_f - T\right) \\ &+ \frac{m\Delta H}{\rho C_p}\, k_0 \mathrm{e}^{-E/(RT)} C_A^{ss}(T) \Bigg] \end{aligned} \tag{2.10}$$

This expression gives the jacket temperature that the controller must find at each setpoint, 302.818 K at 340 K, and its graph is the S-shaped locus in Figure 2.2(b). The locus turns over twice, the picture of Section 1.4.2 with two folds instead of one. The ignition fold sits at $T_c = 303.229$ K and $T = 335.654$ K, above which the cold branch ceases to exist, and the extinction fold sits at

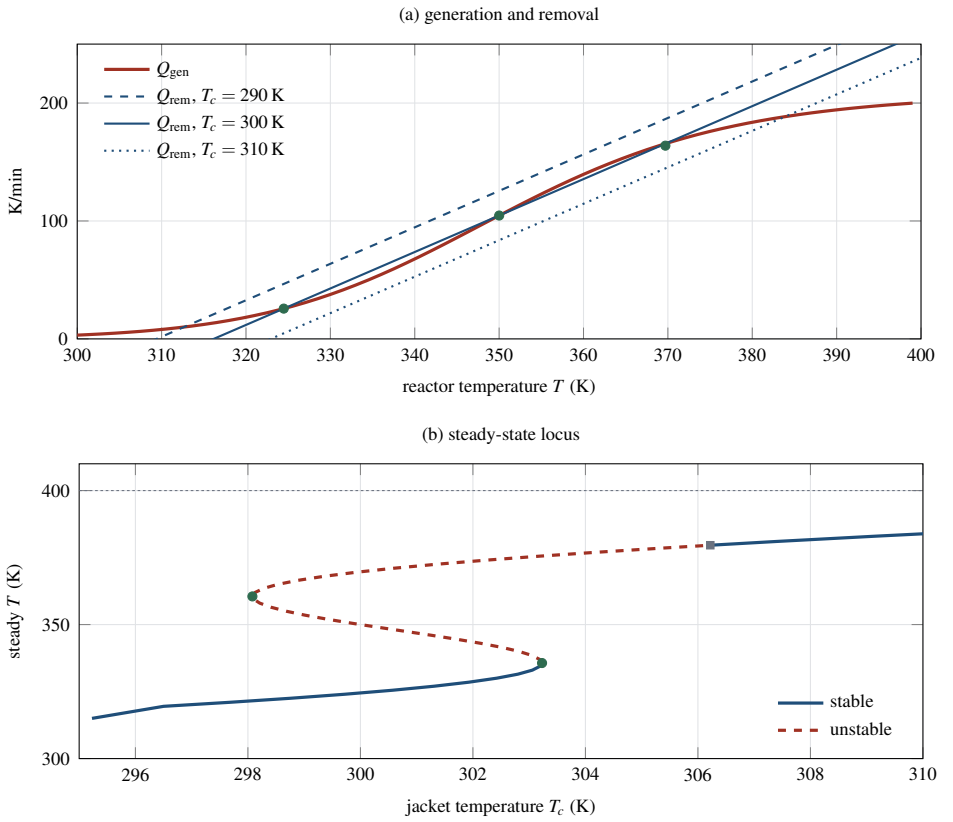


Figure 2.2 Thermal structure of the shipped reactor. Panel (a) is the van Heerden construction at three jacket temperatures, with the three steady states at $T_c = 300$ K marked. Panel (b) is the locus of steady reactor temperature against jacket temperature, with the ignition and extinction folds as circles, the Hopf point at 379.61 K as a square and the 400 K alarm as a horizontal line.

$T_c = 298.080$ K and $T = 360.511$ K, below which the hot branch ceases to exist. Three steady states therefore exist only for jacket temperatures inside a window 5.15 K wide.

At the published jacket temperature of 300 K the reactor has steady states at 324.475 K, 350.006 K and 369.705 K, which is the textbook arrangement. The stability of the third one is not. Steady-state multiplicity of the same kind sets the ignition floor of the catalyst beds in Volume 13, Chapter 4 and the two conversion branches of the reactive column in Volume 13, Chapter 9.

2.4.2 Stability of the Three Steady States

Linearizing Equations (2.2) and (2.3) about a steady state gives the Jacobian in Equation (2.11).

$$A = \begin{bmatrix} -\dfrac{q}{V} - k & -\dfrac{kE}{RT^2} C_A \\ \dfrac{m\Delta H}{\rho C_p} k & -\dfrac{q}{V} - \dfrac{UA}{V\rho C_p} + \dfrac{m\Delta H}{\rho C_p} \dfrac{kE}{RT^2} C_A \end{bmatrix} \tag{2.11}$$

where $k = k_0 \mathrm{e}^{-E/(RT)}$ is the rate constant at the operating temperature. The only positive entry is the lower right one, and it is the derivative of the exotherm with respect to temperature. Instability is therefore always the statement that this single term has outgrown the sum of the dilution rate and the jacket rate constant.

Evaluating Equation (2.11) at the three steady states available at $T_c = 300\,\mathrm{K}$ gives eigenvalues of $-1.049 \pm 0.539\mathrm{j}$ at 324.475 K, -0.454 and $+2.834$ at 350.006 K, and $+1.357 \pm 1.540\mathrm{j}$ at 369.705 K. The middle state is a saddle, as expected. The upper state is an unstable focus, which is not what the textbook drawing of three crossings usually shows, and the application's instructions page says so. The hot branch stays unstable until a Hopf bifurcation at $T_c = 306.220\,\mathrm{K}$ and $T = 379.611\,\mathrm{K}$, where the eigenvalues are $\pm 3.702\mathrm{j}$ and the natural period is 1.70 min. Open-loop, therefore, the reactor is unstable at every temperature between 335.654 K and 379.611 K, an interval 43.96 K wide that contains the default setpoint of 340 K.

> ✦ **DID YOU KNOW? A Reactor and a Pendulum, Both Runaway**
>
> An open textbook of industrial instrumentation files exothermic reactions with the inverted pendulum as runaway processes. Both carry internal positive feedback: gravity pulls an off-center pendulum farther off center, and a hotter reaction runs faster and releases more heat [18]. Chapter 7 holds its pendulum upright by feedback, as this chapter holds 340 K.

Integrating the model confirms the consequence. With the jacket held at 300 K, a reactor released 0.1 K either side of the ignited steady state, or from any temperature between 350.1 and 380 K at the steady-state composition, returns to 324.475 K every time. There is no second stable operating point at the published jacket setting. Holding the jacket instead at 304 K, just above the ignition fold, produces a sustained relaxation oscillation rather than a hot steady state, measured in the application as a cycle between 332.07 K and 466.3 to 466.7 K with a period of 8.39 min. A stable ignited state appears only once the jacket is warmer than 306.220 K, and it is then hotter than 379.6 K, which is inside the alarm margin.

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

2.4.3 Disturbances That Remove the Cold Branch

The same fold analysis run against the two operator-set upsets and the feed flow gives the point at which the cold operating point disappears while the jacket sits at 300 K. Raising the feed temperature removes it at $T_f = 356.8\,\mathrm{K}$, raising the feed concentration removes it at $C_{Af} = 1.141\,\mathrm{mol/m^3}$, and raising the feed flow removes it at $q = 126.3$ vol/min. A warmer feed only shifts the locus of Equation (2.10) by $(q/V)\,\Delta T_f/[UA/(V\rho C_p)]$, so the 15 K feed step scripted in the application is equivalent to warming the jacket by 7.17 K at the design feed. Each of these is a small move from the default. A feed concentration 14 % above design is not an unusual upset, and at the design jacket temperature it removes the cold operating point entirely.

2.5 Dynamic Analysis

2.5.1 Steady-State Gain Along the Locus

The steady-state gain of the temperature loop is the reciprocal of the slope of Equation (2.10), and Figure 2.2(b) shows immediately that this slope passes through zero twice. Table 2.3 lists the gain and the open-loop eigenvalues along the locus. The gain is 1.72 K/K at the nominal point, 3.43 K/K only 5.5 K higher, and unbounded at the ignition fold. Past the fold it is negative, because a warmer jacket on the middle branch pushes the reactor to a lower steady temperature. On the hot branch it recovers a positive value near 1 K/K.

A negative steady-state gain on an open-loop-unstable plant would ordinarily argue for a direct-acting controller, and the application uses a reverse-acting one, with $e_k = T^{\mathrm{sp}} - \hat{T}_k$ and a positive K_c, at every operating point. The reason, as for the arm of Section 1.4.3, is that the sign of the immediate response and the sign of the eventual response are different. The jacket enters only the temperature equation, so the instantaneous sensitivity of the measurement to a jacket move is $UA/(V\rho C_p) = 2.0921/\mathrm{min}$ and is positive everywhere. Warming the jacket always warms the reactor first. Only after the concentration has had time to fall does the loss of reactant pull the steady

Table 2.3 Steady-State Gain and Open-Loop Eigenvalues Along the Locus

T (K)	T_c^{ss} (K)	dT/dT_c (K/K)	eigenvalues (1/min)
324.475	300.000	+1.72	-1.049 ± 0.539j
330.000	302.423	+3.43	-0.642 ± 0.576j
335.654	303.229	$\pm\infty$	−0.281, 0
340.000	302.818	−5.56	−0.517, +1.077
350.000	300.002	−3.25	−0.454, +2.834
360.511	298.080	$\mp\infty$	0, +3.336
369.705	300.000	+2.38	$+1.357 \pm 1.540$j
379.611	306.220	+1.23	± 3.702j
390.000	316.269	+0.91	-2.953 ± 4.833j

temperature back down. The controller acts on the first of these because its 0.05 min period is far shorter than the residence time of 1 min, which is why a reverse-acting loop stabilizes an operating point whose steady-state gain is negative.

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

2.5.2 Identification From a Step Test

Manual mode exists so that the plant can be identified before it is controlled. The application's first case holds the jacket at 300 K with the reactor at its steady state, and stepping the jacket to 303 K at $t = 1$ min moves the reactor from 324.475 K to 332.602 K, so the chord gain over the step is $K_p = 8.127/3 = 2.71$ K/K, the time to 63.2 % of the change is $\tau_p = 1.29$ min, and the noise-free temperature moves on the first 0.005 min substep after the jacket does, so $\theta_p = 0$. The reactor has no transport delay because the jacket acts on the liquid directly.

The chord gain over that 3 K step is 58 % larger than the tangent gain of 1.72 K/K at the starting point, which is a direct measure of how much the plant curves inside a step small enough to look linear on the trend. The application's case 2 makes the smaller step from 300 to 302 K at $t = 1$ min, and the reactor settles at 328.702 K, a chord gain of 2.11 K/K reached 63.2 % of the way 0.94 min after the step. A step of 4 K does not produce a usable fit at all, because 304 K is past the ignition fold and the reactor leaves the cold branch. The test therefore identifies the cold branch only, where Table 2.3 gives a positive steady-state gain, so the 2.71 K/K chord is a property of that branch and not a measurement of the negative gain that the same plant carries a few kelvin higher. A step test that returns a different plant at each operating point is also the lesson of the friction stir weld in Volume 9, Chapter 3.

2.5.3 Tuning From the Fitted Model

Internal model control tuning for a first-order model with negligible dead time gives $K_c = \tau_p/[K_p(\tau_c + \theta_p)]$ and $\tau_I = \tau_p$, where τ_c is the requested closed-loop time constant [11]. Asking for $\tau_c = 0.1\,\tau_p$ cancels the time constant and leaves $K_c = 10/K_p = 3.7$ whatever τ_p turns out to be, and $\tau_I = \tau_p = 1.29$ min then gives $K_I = K_c/\tau_I = 2.9$. The application's default tuning, which is also the suggested tuning of its challenge, is the rounded pair $K_c = 4$ K/K and $\tau_I = 1.333$ min, or $K_I = 3$/min, and it is used for every closed-loop result in Section 2.7 unless a case states otherwise. The identification is performed at 324.5 K and the tuning is then applied across a range in which Table 2.3 shows the gain changing by a factor of three and then changing sign, which is exactly the situation that gain scheduling addresses [8] and that the application makes measurable without providing the schedule.

2.6 Controller Design

2.6.1 Velocity-Form Algorithm

Feedback is applied by the velocity-form PID algorithm of Section 1.6.2, with the proportional term on the error and the derivative on the measurement, executed every $T_s = 0.05\,\text{min}$. In this chapter's symbols each execution forms the increments of Equations (2.12) to (2.14).

$$e_k = T_k^{\text{sp}} - \hat{T}_k \tag{2.12}$$

$$\Delta s_k = K_c\,(e_k - e_{k-1}) - K_c\tau_D\,\frac{\Delta^2 \hat{T}_k}{T_s} \tag{2.13}$$

$$\Delta I_k = \frac{K_c}{\tau_I}\,e_k\,T_s \tag{2.14}$$

where $\Delta^2 x_k = x_k - 2x_{k-1} + x_{k-2}$ is the second backward difference, e_k is the error at execution k, K_c is in kelvin of jacket per kelvin of error, τ_I and τ_D are in minutes, $K_I = K_c/\tau_I$ and $K_D = K_c\tau_D$. The derivative time defaults to zero, so the shipped controller is proportional-integral. Because the derivative acts on the measurement, a setpoint step reaches the jacket only through the proportional term, as a step of $K_c\,\Delta T^{\text{sp}}$ at the next execution. The control period T_s is the Δt of Chapter 1, the running sum of the increments of Equation (2.13) is its S_k, and a gain change acts on the whole error at once.

The controller is reverse acting ($a = +1$ in Equation (1.12)) at every operating point in the ISA sense, because the error of Equation (2.12) is $T^{\text{sp}} - \hat{T}_k$ and K_c is positive, so the jacket demand falls as the measurement rises. Section 2.5 gives the justification for using that action on a plant whose steady-state gain is negative over part of its range.

2.6.2 Limits, Windup and Initialization

The increments are accumulated and the demand is clipped to the jacket range, as shown in Equation (2.15).

$$\begin{aligned} s_k &= s_{k-1} + \Delta s_k \\ c_k &= u_{\text{bias}} + s_k + I_{k-1} \\ u_k &= \min(T_{c,\max},\,\max(T_{c,\min},\,c_k)) \\ I_k &= I_{k-1} + \Delta I_k + \frac{T_s}{\tau_t}\,(u_k - c_k) \end{aligned} \tag{2.15}$$

where s_k accumulates the proportional and derivative increments without limit, I_k is the integral state, c_k is the jacket temperature the algorithm demands, u_{bias} is the bias, fixed when a case is loaded or reset at the manual jacket temperature then in force, and $\tau_t = \tau_I$ is the tracking time of the back-calculation. The integral used at execution k was accumulated through execution $k-1$. The correction is applied only while $K_I \neq 0$. Setting τ_I to the top of its range, 30 min, switches integral action off, and the loop is then proportional only, with neither integral nor correction. Away from the jacket limits the recursion is the positional law, as Section 1.6.2 explains. While the jacket is clipped, substituting the second line into the fourth gives $I_k = I_{k-1} + (T_s/\tau_I)(u_k - u_{\text{bias}} - I_{k-1} - D_k)$, where D_k is the derivative term, so the error cancels and the integral relaxes toward the value that would hold the jacket at its limit with no proportional action, by 3.75 % of the gap per execution at the default $\tau_I = 1.333\,\text{min}$ [19]. After a long saturation the jacket therefore leaves the limit when the error changes sign. There is no rate limit on the jacket temperature and no filter on the derivative term, so derivative action operates directly on the noisy reading and the transmitter noise has to be set to zero before τ_D is useful.

Loading a case or pressing Reset initializes the controller at the case's manual jacket demand u_{man}, as shown in Equation (2.16).

$$u_{\text{bias}} = u_{\text{man}}, \quad I_{-1} = 0, \quad s_{-1} = 0, \quad c_0 = u_{\text{man}} + K_c\,e_0 \tag{2.16}$$

The previous proportional term is zero at that point, so the first execution applies the whole proportional term, which is the positional law about the manual jacket temperature. A case that opens in automatic starts at the steady state of its setpoint with u_{man} at the jacket temperature that holds it, so e_0 is only the transmitter error and the first execution moves the jacket by $K_c \sigma \varepsilon_0$, 0.64 K in the shipped cases at $K_c = 4\,\mathrm{K/K}$. In manual the demand tracks the operator's jacket temperature, and in automatic the manual station follows the jacket, so transfers are bumpless in both directions, as Section 1.6.2 describes. The exception is a setpoint changed together with the transfer, which passes $K_c \Delta T^{\mathrm{sp}}$ to the jacket at the first automatic execution.

The jacket temperature that holds a setpoint matters more on this plant than on a wide-range loop. Only 3.23 K of the 100 K jacket span separates the published setting of 300 K from the ignition fold, so every jacket temperature that holds a cold-branch setpoint above the nominal one lies inside three percent of the actuator span. The learner has no bias to set. When a case starts in automatic the application solves Equation (2.10) itself and seeds the bias with the result, 302.818 K at the 340 K default, so the integral starts at zero, and after a setpoint move or an upset the integral term alone finds the new value. The setpoints the jacket can hold at the design feed run from 282.513 K at its lower limit to 416.427 K at its upper limit.

2.6.3 Model Predictive Control

A linear model predictive controller (MPC) can set the jacket temperature in automatic mode in place of the PI loop, which stays the default everywhere but cases 9 and 10. It works as Section 1.6.3 describes, planning the jacket temperature at every execution, $T_s = 0.05\,\mathrm{min}$ apart, and applying the first move [20, 21]. No step test can supply its model, since the test of Section 2.5 reaches only the cold branch and the reactor leaves its 340 K design point with no input change at all. The model is instead the Jacobian of Equation (2.11) at 340 K with the nominal constants and the design feed, the matrix behind that row of Table 2.3, completed by the jacket term of Equation (2.3), as shown in Equation (2.17).

$$\begin{aligned} \frac{\mathrm{d}\xi}{\mathrm{d}t} &= A_m\,\xi + b_m\,(T_c - \bar{T}_c + \delta), \qquad \tilde{T} = \bar{T} + \xi_2 \\ A_m &= \begin{bmatrix} -1.4793 & -0.024526 \\ 100.279 & 2.0388 \end{bmatrix}, \qquad b_m = \begin{bmatrix} 0 \\ 2.0921 \end{bmatrix} \end{aligned} \tag{2.17}$$

Here $\xi = [\xi_1 \;\; \xi_2]^{\mathsf{T}}$ holds the deviations of C_A and T from $\bar{C}_A = 0.67598\,\mathrm{mol/m^3}$ and $\bar{T} = 340\,\mathrm{K}$, $\bar{T}_c = 302.818\,\mathrm{K}$ holds that state, $\tilde{T}$ is the predicted temperature, δ is an unmeasured disturbance in kelvin of jacket temperature, and the entries keep the units of Equation (2.11), 1/min on the diagonal and in b_m. The eigenvalues of A_m are $+1.0766$ and -0.5171/min, a time constant of 0.929 min for the unstable mode, and its steady-state gain is $G_m = -5.559\,\mathrm{K/K}$. The model thus holds both signs that Section 2.5 separates, the positive first response of 2.0921/min to a jacket move and the negative steady response, so a plan that raises the reactor temperature has to warm the jacket first and then hold it colder than before.

The estimate $\hat{\delta}$ removes offset. A steady-state Kalman filter on Equation (2.17), sampled with a zero-order hold at T_s and extended with δ as a random walk, the input-disturbance model of offset-free predictive control [22], which the boiler of Volume 9, Chapter 4 uses against a fuel bias, runs at every execution in either mode and under either controller, as shown in Equation (2.18).

$$\begin{aligned} \bar{\xi}_k &= \Phi\,\hat{\xi}_{k-1} + \gamma\left(u_{k-1} - \bar{T}_c + \hat{\delta}_{k-1}\right) \\ \eta_k &= \hat{T}_k - \bar{T} - \bar{\xi}_{2,k} \\ \hat{\xi}_k &= \bar{\xi}_k + \ell_\xi\,\eta_k, \qquad \hat{\delta}_k = \hat{\delta}_{k-1} + \ell_\delta\,\eta_k \end{aligned} \tag{2.18}$$

Here $\Phi = \mathrm{e}^{A_m T_s}$ and γ are the sampled model, u_{k-1} is the jacket temperature applied over the period just ended, $\hat{T}_k$ is the reading of Equation (2.6), η_k is the innovation, and $\ell_\xi = [-0.00097 \;\; 0.3595]^{\mathsf{T}}$

and $\ell_\delta = 0.4553$ are the limits of the filter's Riccati recursion when the measurement variance is $r = \sigma^2 = 0.0225\,\mathrm{K}^2$, the state noise variance $10^{-6}r$ and the disturbance noise variance $q_\delta = 10\,r/G_m^2 = 0.00728\,\mathrm{K}^2$. Because δ enters where the jacket does, the 15 K feed step is booked as the 7.17 K of Section 2.4, and a steady load leaves no offset, while $\hat{\xi}_1$ supplies an estimate of C_A that is valid near one operating point.

A quadratic program is posed at every decision on Equation (2.17) sampled at T_s, over a horizon $T_p = 4\,\mathrm{min}$ of $N = 80$ steps, about four time constants of the unstable mode. Twelve jacket temperatures v_j, held over blocks of $(1,1,1,1,2,2,3,4,6,9,15,35)$ steps, minimize Equation (2.19) subject to the hard limits $250\,\mathrm{K} \le v_j \le 350\,\mathrm{K}$.

$$J = \frac{T_s}{\tau_n}\left[\sum_{k=1}^{N}\left(\frac{\tilde{T}_k - T^{\mathrm{sp}}}{y_s}\right)^2 + N\left(\frac{\tilde{T}_N - T^{\mathrm{sp}}}{y_s}\right)^2\right] + \kappa\sum_{j=1}^{12}\left(\frac{v_j - v_{j-1}}{u_s}\right)^2 \tag{2.19}$$

Here $\tilde{T}_k$ is predicted k steps ahead with $\hat{\delta}$ held, v_0 is the jacket temperature applied now, $\tau_n = 1\,\mathrm{min}$, $y_s = 2\,\mathrm{K}$ is the scorecard band, $u_s = 100\,\mathrm{K}$ is the jacket span and $\kappa = 0.3$ is the move suppression, called λ in the application but renamed because λ is the decay rate of Equation (2.7). The weights are 0.0125 per predicted error, 1.0 on the last and 3×10^{-5} per move. No temperature constraint, soft band, rate limit or feed preview enters, so the 400 K alarm and the ignition fold reach the plan only through the model. The decisions are departures from $\bar{T}_c$, a ridge of 10^{-9} relative to the weights makes the Hessian positive definite, and the dual active-set method of [23] finds the optimum. A rule shared by the seven single-loop studios fixed κ and q_δ before any comparison, taking the largest κ within 10 % of the best true-value IAE among settings with no more jacket travel than the PI loop in cases 3 and 4.

The horizon hardly matters here. Case 9 scores 2.694 K min at horizons of 1, 2, 4 and 8 min, with or without the terminal term, and the jacket trajectories differ by at most 0.015 K, because the plan made at the setpoint step already predicts the temperature within 0.04 K of the setpoint 0.25 min ahead. The formulation matters more. With absolute jacket temperatures as decisions and the solver's default ridge, 10^{-9} times one plus the largest diagonal entry of the Hessian, which grows with the square of predictions that the unstable mode multiplies by $\mathrm{e}^{1.0766\times 8} \approx 5500$ over 8 min, the ridge pulls every decision toward 0 K. Rebuilt that way outside the application, case 9 scores 2.712, 13.24 and 469.87 K min at horizons of 4, 6 and 8 min, the last with the reactor falling to 293.58 K on its cold branch while every plan is reported optimal, a defect removed before release.

The PI loop takes over whenever the solver fails to return an optimum or the move it would apply is not finite or lies beyond a jacket limit, and because it tracks the applied jacket at every execution it holds the jacket without a step change until a later plan is accepted. With a rejection forced 0.2 min after the setpoint step of case 9 and the transmitter noise removed, the handover moves the jacket by 0.26 K, within the 0.08 to 0.34 K of the next three PI moves, and the returning plan by 1.22 K. The first plan is computed while a case loads, so the MPC acts from $t = 0$. It reads only the measurement, its own jacket move, the setpoint and the jacket limits, and its model is built once from the nominal constants, so a constant the learner changes reaches the plant of Equations (2.2) and (2.3) and never Equation (2.17).

2.6.4 Moving Horizon Estimation

The Kalman filter of Equation (2.18) is linear in the readings, so nothing keeps its estimate of C_A inside the physical range once the reactor is far from 340 K. The application therefore offers a second source for the state each plan starts from, a moving horizon estimator (MHE) that fits the same sampled model to a window of readings under bounds on the states [24]. The Kalman filter remains the default and keeps running whichever estimator is chosen. With the augmented state $x_j = [\xi_{1,j} \;\; \xi_{2,j} \;\; \delta_j]^{\mathsf{T}}$

and the window $j = s, \dots, k$ with $s = k - N$, every execution solves Equation (2.20) over x_s and $w_s, \dots, w_{k-1}$.

$$\begin{aligned} \min \; & \|x_s - \bar{x}_s\|^2_{\bar{P}^{-1}} + \sum_{j=s}^{k-1} \|w_j\|^2_{Q^{-1}} \\ & + \frac{1}{r} \sum_{j=s}^{k} \left(\hat{T}_j - \bar{T} - x_{2,j}\right)^2 \\ \text{s.t. } & x_{j+1} = \Phi_a x_j + \gamma_a \left(u_j - \bar{T}_c\right) + w_j \\ & 0 \le \bar{C}_A + x_{1,j} \le C_{Af}^{\max} \\ & T_{c,\min} \le \bar{T} + x_{2,j} \le T_f^{\max} + \frac{m\Delta H}{\rho C_p} C_{Af}^{\max} \end{aligned} \tag{2.20}$$

Here $\Phi_a = \left[\begin{smallmatrix} \Phi & \gamma \\ 0 & 1 \end{smallmatrix}\right]$ and $\gamma_a = [\gamma^{\mathsf{T}} \;\; 0]^{\mathsf{T}}$ carry δ as the random walk of Equation (2.18), u_j is the jacket temperature applied from execution j to $j+1$, w_j is the process noise with $Q = \mathrm{diag}(10^{-6} r, 10^{-6} r, q_\delta)$, $\bar{x}_s$ is the estimate the MHE made at $s-1$ carried one step through the model, and $\bar{P}$ is the filter's steady a-priori covariance, held fixed as the arrival cost. The window spans $N = 15$ executions, 0.75 min and 16 readings, so the problem has 48 decisions and 64 bound rows, and the dual active-set method of [23] solves it. The bounds come from declared values only. $C_{Af}^{\max} = 2\,\mathrm{mol/m^3}$ is the richest feed the model allows, $T_{c,\min} = 250\,\mathrm{K}$ is the coldest jacket, and the upper temperature bound of 798.41 K is the adiabatic limit of that feed at the hottest feed temperature, $T_f^{\max} = 380\,\mathrm{K}$. The disturbance δ is not bounded, and a constant the learner changes reaches the MHE only through the reading, as it reaches the filter.

Without an active bound, Equation (2.20) is the filter's own least-squares problem and the two estimates coincide. Over the 599 executions of a 30 min run of case 9 they differ by at most 7.9×10^{-11} of the transmitter's standard deviation, a residue of the filter's Riccati recursion, which stops at a relative change of 10^{-11}, because with the covariance iterated to its fixed point the difference falls to 3.2×10^{-13}. Run under the MPC with each estimator, the ten earlier cases agree to within 6.1×10^{-8} K of jacket temperature. A bound is active in only one of them, case 5, which stays in manual while the filter's concentration estimate falls as far as 0.21 mol/m^3 below zero at 222 executions. The MHE thus adds its bounds and nothing else, and on this reactor only a runaway brings the filter to them (Section 2.9).

A window whose solution is not optimal, not finite or outside the bounds is rejected, and the estimate is then the Kalman update of the MHE's one-step prediction, clipped to the bounds and marked in the log. No case produced a rejection. Forced at the execution at 3.5 min of case 10 with the noise removed and a bound active, a rejection moves the jacket by 23.4 K at that execution, the largest difference from an unfaulted run over the next 32, and because the rejected estimate later serves as the prior of a window, a concentration difference of 2.8×10^{-3} and 1.4×10^{-3} mol/m^3 returns 16 and 32 executions later. An estimate took a median of 142 µs and a 99th percentile of 512 µs under Node.js over the 5470 estimates of those ten cases. In Chrome, on the application's own frame loop during case 9, the estimates took 0.76 and the plans 0.31 percent of one processor core at one minute of process time per second, and 5.23 and 1.71 percent at ten, measured while other work loaded the same computer.

2.6.5 Relay Autotuning

The PI settings can also be measured rather than derived, by the relay experiment of [25], which the application runs as an action from the present state in automatic and never as a mode a case is in. The relay sets the jacket $d = 5\,\mathrm{K}$, five percent of its span, above or below the mean jacket temperature applied over the last 1 min, a tenth of the trend window, and switches each time the reading crosses the setpoint by more than $\varepsilon = 3\sigma = 0.45\,\mathrm{K}$. A cycle counts when each of its halves lasts at least four executions, and once two counted cycles of amplitude a agree within five percent, the describing

function and the Tyreus–Luyben rule give the settings of Equation (2.21).

$$K_u = \frac{4d\sin(\pi D)}{\pi\sqrt{a^2-\varepsilon^2}}, \qquad K_c = \frac{K_u}{3.2}, \qquad \tau_I = 2.2\,P_u \tag{2.21}$$

Here K_u is the ultimate gain in kelvin of jacket per kelvin, D the fraction of each period that the jacket spends at its upper level, which is one half for a symmetric cycle, and P_u the measured period. The experiment ends after 30 cycles or 20 min, and the PI loop then resumes from the applied jacket without a step. Where the band is wide against the swing, the band rather than the loop sets the cycle, which the application flags once $\psi = \arcsin(\varepsilon/a)$, the phase by which the cycle falls short of -180°, reaches 10°.

On this reactor the relay cannot measure the loop. Started from the steady state of case 3 without its setpoint step, its halves last a median of three executions with the transmitter noise and two without it, so 29 of the 30 cycles are rejected and the experiment ends after 4.35 min with the tuning unchanged. Over ten further noise seeds it converged in four, with K_u from 6.13 to 6.90, P_u from 0.389 to 0.413 min and ψ from 23.5° to 26.1°, cycles set by the band. The same relay on the sampled model of Equation (2.17) without noise cycles at $P_u = 0.399$ min, so the fast cycle belongs to the loop and not to the noise. Holding each relay level for at least four executions, 0.2 min, makes the experiment converge for the default seed and all ten others, but the hold then sets the period, $P_u = 0.45$ min or 2.25 holds. The resulting $K_c = 1.65\,\mathrm{K/K}$ and $\tau_I = 0.99$ min score 14.6 and 9.53 K min of measured IAE over the first 30 min of cases 3 and 4, against 5.51 and 6.02 K min for the shipped tuning, and holds of 5, 6 and 8 executions score 24.3, 30.4 and 444 K min on case 3. The application therefore applies no hold and states the reason with the action.

2.7 Results

The application keeps its own scorecard, but its integral of absolute error runs from $t = 0$, so the numbers in this section are computed over fixed windows from the exported columns and from the noise-free state at the 0.005 min substep. Peaks, final values and settling times are taken from the noise-free temperature at the substep, because the export samples every 0.1 min and misses as much as 21 K of an ignition peak in Table 2.5, and the integral of absolute error is taken from the exported measured column that the controller actually sees. Settling time is the last instant, on the 0.005 min grid of the substep, at which the temperature is more than 0.5 K from setpoint, measured from the step, and the integral of absolute error is accumulated over a fixed 12 min window after the step. The alarm entry records whether the noise-free temperature passes 400 K. Each of the eight experiments starts from one of the application's eleven cases with its shipped noise seed, as listed in Table 2.4.

2.7.1 Open-Loop Step Tests

Experiment 1 produces the identification reported in Section 2.5. The reactor moves from 324.475 K to 332.602 K on a 3 K jacket step, rising to 333.219 K 5.14 min after the step, 0.62 K or 7.6 % of the change past its final value, and staying within 2 % of the change from 8.27 min after the step. That single overshoot is the behavior of the underdamped cold branch, whose eigenvalues at 332.6 K are -0.420 ± 0.479j, and the first-order fit with $\tau_p = 1.29$ min captures its dominant part.

Experiment 2 changes the jacket by one kelvin more. The reactor leaves the cold branch, peaks at 469.27 K 3.70 min after the step, and then settles into a sustained oscillation between 332.07 K and 466.3 to 466.7 K with a period of 8.39 min, spending about 0.5 min of each cycle above the 400 K alarm. One kelvin of jacket temperature is the difference between a step response and a relaxation oscillation, which is the whole content of the fold at 303.229 K presented as an experiment. A larger step from case 1, from 300 to 320 K at $t = 1$ min, goes past the Hopf point as well as the fold. The reactor ignites to 505.03 K 0.68 min after the step, stays above 400 K for 0.86 min, and is within 0.5 K of the stable hot steady state at 393.31 K from 1.9 min after the step, with only 0.060 mol/m^3 of reactant left.

Table 2.4 Experiments and the Cases They Start From

#	Case	Mode	Action
1	1	manual	jacket 300 to 303 K at $t = 1$ min, step test
2	1	manual	jacket 300 to 304 K at $t = 1$ min, past the ignition fold
3	1	manual, PI	setpoint 340 K and automatic together at $t = 1$ min
4	1	PI	setpoint 340 K and automatic chosen before Run, then setpoint steps at $t = 1$ min to 345, 350, 351, 352, 355, 360, 365 and 371 K
5	1	PI	as 4 with $K_c = 1.6$, 4 and 8 K/K chosen before Run and a step to 354 K
6	4, 5	PI, manual	feed temperature 350 to 365 K at $t = 1$ min
7	1	PI, manual	feed concentration 1.0 to 1.2 mol/m^3 at $t = 1$ min, in automatic as 4
8	11	PI	as 6 with $K_c = 0.8$ and 4 K/K

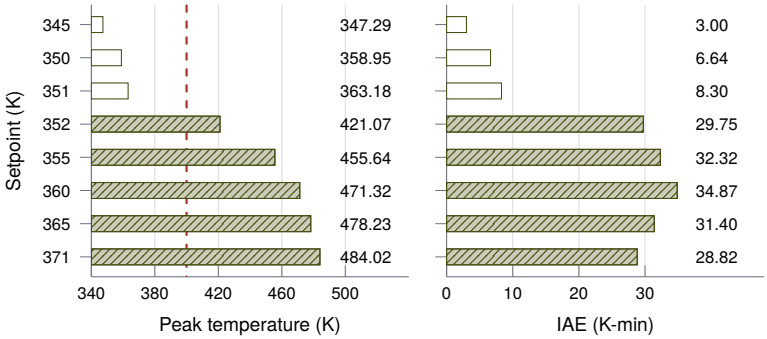


Figure 2.3 One PI tuning on eight setpoint steps from 340 K, the peak temperature (left, dashed line at the 400 K alarm) and the integral of absolute error (right) of Table 2.5. White bars stay below the alarm and hatched bars pass it. Notice the cliff between the 351 and 352 K setpoints, where one kelvin more of setpoint lifts the peak from 363.18 to 421.07 K and the error from 8.30 to 29.75 K min.

2.7.2 One Tuning Across Eight Setpoints

Experiment 3 makes the move from the cold branch that the automatic cases no longer make at their start. Setting the setpoint to 340 K and switching case 1 to automatic together at $t = 1$ min passes $4 \times 15.5 \approx 62$ K to the jacket at the first automatic execution, so the jacket goes to its 350 K limit for one control period, the reactor crosses the ignition fold at 335.654 K under feedback 0.13 min after the transfer, peaks at 342.67 K 0.50 min after it, and stays within 0.5 K of setpoint from 0.97 min after it.

Experiment 4 is the central measurement. Choosing the 340 K setpoint and automatic in case 1 before Run starts the reactor at the steady state of 340 K with the jacket at 302.818 K, the way every automatic case opens. The setpoint is then stepped at $t = 1$ min with $K_c = 4\,\mathrm{K/K}$, $\tau_I = 1.333$ min and $\tau_D = 0$ unchanged, so the 345 K step is case 3 of the application, run for run. The application credits case 3 at $t = 2.0$ min, after the peak, once the temperature has stayed within ± 2 K of 345 K for 0.5 min. Table 2.5 reports the result, and Figure 2.3 draws its peaks and errors against the alarm.

Table 2.5 One PI Tuning, Eight Setpoint Steps From 340 K

T^{sp} (K)	peak (K)	overshoot (K)	IAE (K-min)	settling (min)	alarm
345	347.29	+2.29	3.00	0.845	no
350	358.95	+8.95	6.64	1.505	no
351	363.18	+12.18	8.30	1.545	no
352	421.07	+69.07	29.75	2.355	yes
355	455.64	+100.64	32.32	2.330	yes
360	471.32	+111.32	34.87	2.000	yes
365	478.23	+113.23	31.40	2.105	yes
371	484.02	+113.02	28.82	2.255	yes

The transition is not gradual. An 11 K step overshoots by 12.18 K and never trips the alarm, and a 12 K step overshoots by 69.07 K and trips it, with the integral of absolute error rising from 8.30 to 29.75 K min over the same one-kelvin change of setpoint. In every case the loop reaches the setpoint and holds it, so a metric computed only at the end of the run would report all eight tests as successful. The mechanism is visible in the jacket column. The step moves the jacket at once by $K_c\,\Delta T^{sp}$, to 346.4 K for the 11 K step and to its 350 K limit for the 12 K step, and the reactor, already above the ignition fold, climbs at close to 100 K/min while the controller drives the jacket back down. With the jacket at its 250 K limit, the exotherm exceeds the largest removal the jacket can supply above 362.5 K when 0.50 mol/m^3 of reactant is left and above 364.5 K when 0.45 mol/m^3 is left. The 11 K step turns at 363.18 K with the jacket at 251.1 K. The 12 K step reaches the jacket limit at the execution 0.35 min after the step, with the reactor already at 366.2 K and 0.499 mol/m^3 of reactant, and the reactor then runs to 421.07 K, burning the reactant down to 0.018 mol/m^3 while the jacket sits on its lower limit for 0.6 min.

Where that boundary falls does not depend on the transmitter noise. Repeating Experiment 4 with $\sigma = 0$ and with eight other seeds of the generator puts the transition between the 11 K and the 12 K step every time. The 11 K step peaks between 362.61 and 363.19 K over those seeds, while the 12 K step peaks between 406.17 and 420.34 K, and at 416.67 K without noise, so the noise changes how far the ignited reactor runs but not whether it ignites. What does not move is the abruptness, which is one kelvin of setpoint between a well-damped response and an alarm.

Cases 6 and 7 test the remedy that first suggests itself on the 5 K step of case 3. Reducing the proportional gain from 4 to 1.6 K/K (case 6) raises the peak from 347.29 K to 351.67 K, more than twice the step, and the reactor stays within 0.5 K of setpoint only from 6.095 min after the step, against 0.845 min at the reference gain. Doubling the gain to 8 K/K (case 7) lowers the peak to 346.36 K and settles in 0.51 min. Experiment 5 repeats the comparison on a 14 K step from 340 to 354 K at $t = 1$ min, which ignites the reactor at the reference gain. Reducing the gain to 1.6 K/K raises the peak from 450.78 K to 460.13 K, and the loop never settles. It falls into a relaxation cycle between about 338 and 416 K with a period of 2.10 min, which passes the alarm 15 times in the first 30 min. Doubling the gain to 8 K/K keeps the reactor off the hot branch altogether, with a peak of 363.04 K, or between 361.90 and 363.59 K over five other noise seeds. Linearized about 354 K with $\tau_I = 1.333$ min, the loop is stable only for $K_c > 1.69\,\mathrm{K/K}$, against 0.57 K/K at 340 K and 1.04 K/K at 345 K, because the open-loop instability grows along the middle branch, so the conservative gain cannot hold the new setpoint at all. On the larger 31 K step to 371 K, detuning to $K_c = 1\,\mathrm{K/K}$ and $\tau_I = 1.25$ min, or $K_I = 0.8/\mathrm{min}$, lowers the first peak by 5.95 K, to 478.07 K against 484.02 K, but that gain is again below the stability bound of 1.45 K/K there, and the reactor passes 400 K 21 times in the 30 min after the step. The excursion is governed by how fast the controller can remove heat after ignition, and the operating point itself needs a minimum gain to exist, so detuning buys no safer approach on this plant. Raising the gain does not remove the largest excursions either. Steps to 360 K and above pass the alarm at every K_c from 4 to the 20 K/K top of its range.

Table 2.6 The Same Upset With and Without Feedback

Upset	Mode	peak (K)	final (K)	outcome
T_f to 365	PI	342.15	340.02	returns to setpoint
T_f to 365	manual	484.29	380.75	ignites, alarm
C_{Af} to 1.2	PI	341.06	340.02	returns to setpoint
C_{Af} to 1.2	manual	515.38	392.20	ignites, alarm

Peak after the upset and mean noise-free temperature over the last 5 min of the 30 min after it. PI rows start at the steady state of 340 K, and manual rows hold the jacket at 300 K from the steady state at 324.475 K.

> **TRY IT IN THE APP** **6 · Conservative Tuning**
>
> Load the scenario and press Run. With the gain cut from 4 to 1.6 K/K, the 5 K setpoint step peaks at 351.67 K, more than twice the step, and stays within 0.5 K of setpoint only from 6.095 min after it. Then load 7 · Aggressive tuning. At 8 K/K the peak is 346.36 K and the loop settles in 0.51 min.

2.7.3 Feedback Against Igniting Upsets

Cases 4 and 5 run the same upset twice, a feed temperature step from 350 to 365 K at $t = 1$ min, once with the loop closed at 340 K and once with the jacket held by hand at 300 K and the reactor on its cold branch at 324.475 K. Experiment 7 does the same from case 1 with a feed concentration step from 1.0 to 1.2 mol/m^3 at $t = 1$ min. Table 2.6 gives the outcome.

Both upsets are past the fold values computed in Section 2.4, so in manual the cold branch no longer exists and the reactor has to leave it. Under feedback the same upsets are absorbed with a peak deviation of 2.15 K and 1.06 K respectively, or 1.88 to 2.14 K and 1.06 to 1.17 K over five other noise seeds, because the controller lowers the jacket by 7.2 and 6.5 K, to the 295.65 K and 296.34 K that Equation (2.10) requires, before the exotherm can build. The feed concentration case ends at the hot steady state of 392.20 K with 93.6 % conversion. The feed temperature case ends at 380.75 K after an oscillation that decays with a period of 1.62 min. A feed 15 K warmer acts like a jacket 7.17 K warmer, which places this hot state just past the Hopf point of Section 2.4, where its eigenvalues are -0.238 ± 3.895j and the predicted period is 1.61 min. By hand the same upset is a race. Case 5 asks the learner to keep the cold branch, and a jacket lowered to 293 K, where the cold steady state with the warmer feed is 324.77 K, does so if it is set no later than $t = 2.05$ min, and the reactor then peaks at 339.37 K. Set at 2.1 min, the same jacket is too late and the reactor ignites to 450.86 K before it returns.

> **TRY IT IN THE APP** **5 · Recover by Hand**
>
> Load the scenario and press Run without touching the jacket. At 1 min the feed warms from 350 to 365 K, the cold operating point disappears, and the reactor temperature climbs past the 400 K alarm to a peak of 484.29 K (Table 2.6). Then load 4 · Disturbance rejection, the same upset under the PI loop, and watch it peak at 342.15 K.

Feedback has to be strong enough to act before the exotherm does, and the application's challenge, case 11, makes that its target. It applies the feed temperature step of case 4 with the proportional gain cut to 0.8 K/K, which still stabilizes 340 K in the linear sense, and asks for the measured temperature to stay within ± 2 K of 340 K over the last third of a 4 min run, from $t = 2.64$ min. With the shipped gain the reactor ignites to 445.08 K 0.96 min after the upset, falls back to 325.23 K once the reactant is spent, and is still 15.07 K from setpoint in the judged window, 7.5 times the band. With the suggested $K_c = 4\,\mathrm{K/K}$ the worst measured error in the same window is 0.573 K, inside the band with 1.43 K to spare, and the worst noise-free error is 0.438 K. The suggested tuning gives the same run whether it is set before Run or carried through Reset.



Figure 2.4 Stirred Reactor Control Studio during a setpoint move that ignites the reactor, a composite of the plant view's vessel and readouts (left, scaled down) and the trends (right), with the controls and the Learn panel cut. Case 1 opens in manual at the published steady state of 324.5 K, the setpoint is set to 340 K with the loop transferred to automatic at $t = 4$ min (2), and the setpoint is raised to 355 K at $t = 13$ min. The trends show the reactor igniting above the 400 K alarm (3) while the jacket sits on its 250 K limit (4, shaded) and the reactant concentration collapses (5). The run peaked at 456.0 K (446.9 K in the 0.1 min trend) with the jacket on its limit for 0.5 min, values the screen does not print. At $t = 30$ min the readouts (1) give 354.8 K (355.1 K measured, at the top right of the trends) with the jacket at 299.9 K and 0.414 mol/m^3 of reactant left.

Table 2.7 Five Cases Run Twice, Under the PI Loop and Under the Predictive Controller, With Identical Transmitter Noise

Case	Controller	IAE (K min)	IAE$_T$	OS (K)	$e_{\max}$	$t_{0.5}$ (min)	Travel (K)	Limit (%)
3, setpoint	PI	2.68	1.90	2.29	—	0.90	119.2	0
	MPC	1.86	1.10	0.75	—	0.50	89.2	0
4, feed	PI	3.26	2.75	—	2.15	1.70	93.5	0
	MPC	2.32	1.57	—	2.05	0.70	72.0	0
8, limit	PI	41.89	41.79	—	—	—	71.6	80
	MPC	42.80	42.70	—	—	—	95.7	87
9, unstable	PI	4.38	3.86	2.29	2.33	2.00	127.7	0
	MPC	2.69	2.05	0.75	2.27	0.60	103.5	0
11, challenge	PI, $K_c = 0.8$	57.94	57.80	—	93.44	—	105.8	7.5
	PI, $K_c = 4^a$	2.26	2.30	—	2.15	1.70	39.9	0
	MPC	1.42	1.15	—	2.05	0.70	36.6	0

IAE uses the measured temperature, as the scorecard does, and IAE$_T$ the noise-free one. OS is the largest true overshoot of the new setpoint, $e_{\max}$ the largest true deviation after the feed step, $t_{0.5}$ the time from the last event until the true temperature stays within ± 0.5 K of setpoint (a dash where it never does), Travel the summed absolute jacket change and Limit the percentage of time with the jacket on a bound. [a]Suggested, $\tau_I = 1.333$ min.

One 30 min session carries the whole argument of this chapter, as shown in Figure 2.4. The measured reactor temperature is plotted against its setpoint, with the jacket temperature between its two limits and the reactant concentration on the same time axis below it, so the interval the jacket spends on its lower limit can be read against the ignition peak and the collapse of the reactant concentration.

2.8 Predictive Control at the Unstable Operating Point

Table 2.7 compares the two controllers on cases 3, 4, 8 and 9 over 10 min each and on the 4 min challenge. The controller is selected before the run starts, which gives the two runs of a case identical events and identical transmitter noise. The PI loop keeps $K_c = 4$ K/K, $\tau_I = 1.333$ min and $\tau_D = 0$, except that the challenge loads $K_c = 0.8$ K/K and suggests 4, while every MPC run uses the settings of Section 2.6.3. Unlike Section 2.7, values are read on the 0.1 min log, as the scorecard reads them, with the IAE over the whole case, so case 3 under the PI loop scores 2.68 K min here and 3.00 K min in Table 2.5. An instant finer than the log is named as an execution.

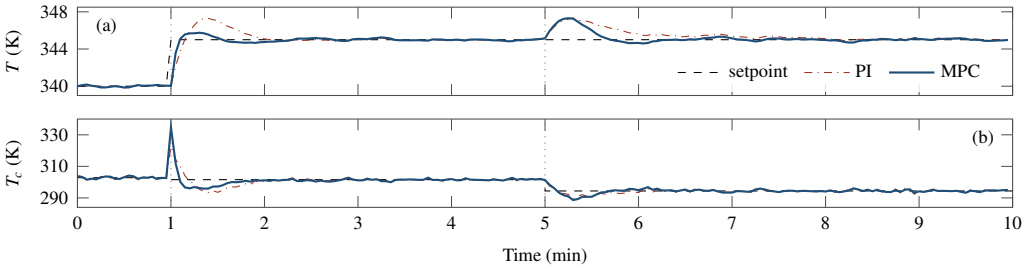


Figure 2.5 Case 9 under the predictive controller and the PI loop with identical transmitter noise, at every controller execution (0.05 min, twice the rate of the trend), the PI loop dash-dotted and the MPC solid. Dotted lines mark the setpoint step to 345 K and the 15 K feed step. (a) Noise-free temperature and setpoint (dashed). (b) Jacket temperature and the steady value that holds the setpoint (dashed).

Case 3 steps the setpoint from 340 to 345 K at 1 min, and the plan made at that execution already holds the maneuver that the negative steady gain requires. Its first block raises the jacket to 334.76 K, its second holds 312.28 K, every later block lies below the 302.82 K that held 340 K, and the last holds 301.84 K, near the 301.92 K at which the model holds 345 K (the balances need 301.57 K). The applied jacket follows, 334.76 K for one execution (311.96 K on the log), below its start from 1.10 min, lowest at 295.90 K. The PI loop makes the same reversal by feedback, from a kick to 322.41 K to a low of 293.46 K, but over the two executions after the kick it lowers the jacket by 8.05 and 4.38 K where the MPC lowers it by 22.80 and 11.77 K. The MPC peaks at 345.75 K against 347.29 K, stays within ± 0.5 K from 0.50 min after the step against 0.90 min, and moves the jacket 25 % less.

In case 4 the estimate $\hat{\delta}$ passes half of the 7.17 K that the 15 K feed step is worth 0.4 min after it and 90 % after 0.6 min, and it ends at 7.08 K, the same under the PI loop, so the difference lies in how the plan uses it. The peaks nearly agree, 342.05 against 342.15 K, but the MPC is back within ± 0.5 K 0.70 min after the upset against 1.70 min, with 23 % less jacket travel.

Case 8 is the one the MPC loses, by 2.2 %. When the setpoint drops to the unreachable 280 K, both controllers put the jacket on its 250 K floor and leave the reactor at 282.51 K, but the MPC lifts the jacket to 255.02 K at 1.55 min and to 261.38 K at 1.95 min. Its model, linearized 43 to 53 K above the reactor, keeps the unstable mode of the middle branch, so the plan made at 1.50 min raises the jacket to 313.69 K within two executions and still predicts a dip to 279.73 K, while the reactor, on its stable cold branch, never falls below 282.51 K. The MPC gathers 0.95 K min more IAE than the PI loop between 1 and 3.3 min, more than it loses by, and its estimate books the error of the model as a load of 63.07 K.

Case 9 repeats case 3 to 5 min, execution for execution, and then warms the feed by 15 K, as shown in Figure 2.5. Both controllers then peak about 2.3 K above the setpoint, 2.27 against 2.33 K, but the MPC is back within ± 0.5 K 0.60 min after the upset against 2.00 min, and the scorecard IAE falls by 39 % with 19 % less jacket travel. Both jackets end near the 294.40 K that holds 345 K against the warmer feed, at 295.04 and 294.92 K. Because the MPC never reads K_c, the challenge under the MPC is the first 4 min of case 4 again, with a worst measured error of 0.41 K in the judged window against 15.07 K for the shipped $K_c = 0.8\,\mathrm{K/K}$ and 0.573 K for the suggested tuning.

Repeating the sweep of Table 2.5 under the MPC measures what the model buys beyond 345 K. On a 0.5 K grid, the largest setpoint step that keeps the noise-free temperature below the alarm is 11.5 K under the PI loop and 13.5 K under the MPC. At 352 K the PI loop peaks at 421.07 K and the MPC at 359.25 K, because from the same first move to the 350 K limit its plan cuts the jacket to 306.27 K at 1.10 min, where the PI loop is at 316.95 K. From a 14 K step up both ignite the reactor, with peaks at 360, 365 and 371 K within 2.1 K of each other. Held at the new setpoint from 10 to 31 min, the MPC moves the jacket a third less but more loosely, a root-mean-square true error of

Table 2.8 Case 10 Under the Two Estimators, Handed Over at the Scripted 3 min and Later

t_h (min)	Estimator	IAE (K min)	IAE_h (K min)	t_2 (min)	Travel (K)	$\min \hat{C}_A$ (mol/m^3)	e_C	e_T (K)
3	Kalman	123.1	54.7	1.60	397	−0.241	0.119	5.37
	MHE	118.0	49.6	1.54	364	0	0.040	5.48
3.5	Kalman	117.5	19.7	1.00	270	−0.212	0.096	2.10
	MHE	112.2	14.7	0.95	170	0	0.041	1.97
4	Kalman	123.1	6.28	0.575	162	−0.212	0.067	0.655
	MHE	123.4	6.47	0.665	165	0	0.042	0.736

IAE is the scorecard's over the 10 min run on the measured temperature, and IAE_h the noise-free one from the hand-over t_h to 10 min. t_2 is the time from the hand-over after which the true temperature stays within ±2 K of 340 K, Travel the summed jacket change, $\min \hat{C}_A$ the lowest concentration estimate of the chosen estimator, and e_C and e_T the root-mean-square errors of its concentration and temperature estimates after the hand-over, on the 0.005 min substep.

0.115 against 0.101 K at 345 K and 0.187 against 0.137 K at 360 K, where the plant is more unstable than the model.

Of 880 plans none was rejected, and a bound was active only in 176 plans of case 8. A plan took 0.05 to 0.22 ms on average under Node.js and 199 and 212 µs in two Chrome runs of case 9, wall-clock values from one desktop computer, at most 0.42 % of a processor core at one minute of process time per second.

2.9 Estimation After a Runaway

Case 10 gives the predictive controller a reactor that has already run away. It opens in manual on the cold branch at 324.475 K with the jacket at 300 K, warms the feed from 350 to 365 K at 1 min as case 5 does, and switches to automatic under the MPC at 3 min. By then the noise-free temperature has peaked at about 484 K and fallen to 418.1 K, with 0.015 mol/m^3 of reactant left. The case loads the MHE and asks for a second run with the Kalman filter, so the two runs share their events and transmitter noise. Table 2.8 reports both, together with runs in which the hand-over is moved to 3.5 and 4 min.

With the hand-over at 3 min, the filter's concentration estimate is below zero from the execution at 2.95 to the one at 4.05 min, 23 executions, down to −0.241 mol/m^3, while the true concentration after the hand-over lies between 0.015 and 0.677 mol/m^3. The MHE holds its estimate on the zero bound instead, with a bound active at 30 executions from 2.95 to 4.40 min, and its concentration error after the hand-over is 0.040 against 0.119 mol/m^3. The plan built on that estimate brings the reactor within ±2 K of setpoint 1.54 min after the hand-over against 1.60 min, lowers the IAE after the hand-over from 54.7 to 49.6 K min, or 9 %, and moves the jacket 8 % less. At the execution at 3.5 min, marked in Figure 2.6, the MHE's plan starts from zero concentration where the filter's starts from −0.237 mol/m^3, and the jacket is set to 273.2 K against 278.2 K, while the true concentration is 0.026 and 0.030 mol/m^3.

The gain is small and depends on when the reactor is handed over. Both estimators feed the same model, linearized 78 K below the temperature at the hand-over, and the MHE's temperature estimate is no better, with an error after the hand-over of 5.48 against 5.37 K. Handed over at 3.5 min, the MHE lowers the IAE after the hand-over from 19.7 to 14.7 K min, but at 4 min it raises it from 6.28 to 6.47 K min. Over ten other noise seeds the MHE gives the lower IAE after the hand-over in all ten runs at 3 and 3.5 min and in none at 4 min, with the 3 min values spread over 54.7 to 55.0 against 49.6 to 49.9 K min. Halving the integration substep gives 53.4 against 48.7 K min at 3 min and 6.34 against 6.52 K min at 4 min, the same ranking. A concentration of the right sign does not repair the model either. In the rows logged at 3.2, 3.3 and 3.4 min the end of the MHE's plan lies below absolute zero, down to −2861 K at 3.3 min, and the filter's plan reaches −2477 K at the same row.

Changing the estimator in the middle of the case moves the plan only where a bound is active. Chosen at 3.5 min of a Kalman run with the noise removed, when the filter places the concentration at −0.238 mol/m^3, the MHE returns zero, moves the temperature estimate from 398.8 to 385.1 K, and

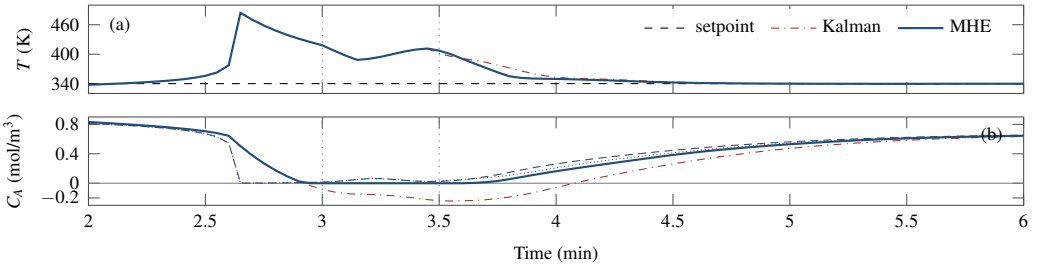


Figure 2.6 Case 10 with the MHE and with the Kalman filter, identical transmitter noise, at every controller execution. Dotted vertical lines mark the hand-over to the predictive controller at 3 min and the execution at 3.5 min discussed in the text. (a) Noise-free temperature with the Kalman filter (dash-dotted) and the MHE (solid), and the setpoint (dashed). (b) The estimate the plan starts from, Kalman dash-dotted and MHE solid, and the true concentration of each run, Kalman dotted and MHE dashed.

sets the jacket 52.3 K away from that of a run that stays on the filter. The difference is the bound acting on the estimate, not a bump in the transfer. Switched back one minute later, the plan again starts from the filter's estimate, which differs from the other run's by 0.086 mol/m^3 only because the plant moved differently in between.

🕮 **Pause.** Try Exercise 2.3 before reading on.

2.10 Technology in Practice

What Measures It

The one controlled variable is the liquid temperature, and in a plant it is read by an element in a thermowell. Resistance thermometers are preferred at lower temperatures, where they are most accurate, and become inaccurate above about 700 °C as their sheath degrades. Thermocouples work from about −200 °C to 2320 °C without a power source, but their insulation wears with use and needs periodic checks to keep their accuracy [26]. The chapter's transmitter is instantaneous and adds white noise of standard deviation 0.15 K, while Section 2.3.5 notes that a real thermowell has a time constant of tens of seconds, the same order as the reactor's own.

> **TECHNOLOGY Sensor: A Temperature Element in a Thermowell**
>
> A reactor's temperature is read by a resistance thermometer or a thermocouple inside a thermowell, a pressure-tight sheath that lets the element be replaced without depressurizing the vessel. A platinum resistance thermometer covers about −200 °C to 800 °C and is chosen for its accuracy and precision [26]. The well's mass adds a first-order lag, an air gap at its tip adds a second one, and a lag severe enough can compromise the stability of feedback. A well in a vessel sees less turbulence than one in a pipe, which slows the heat transfer and leaves more fouling on it. A thermocouple most often fails open, and without burnout detection the trend then shows a wildly varying false temperature, while a burnout mode drives the reading high or low on purpose [18].

What Moves It

The chapter manipulates the jacket temperature itself. In a plant that temperature is the setpoint of a slave loop that mixes hot and cold utility or throttles coolant (Section 2.2), and its final elements are the control valves on the utility lines. A cascade works only if the slave variable responds faster than the master, and the slave controller is configured and tuned first [18]. The ideal jacket of the model, which reaches any temperature from 250 to 350 K at once, is the limit of an infinitely fast slave loop, so the jacket limits are the only property of the final element that the chapter's results feel.

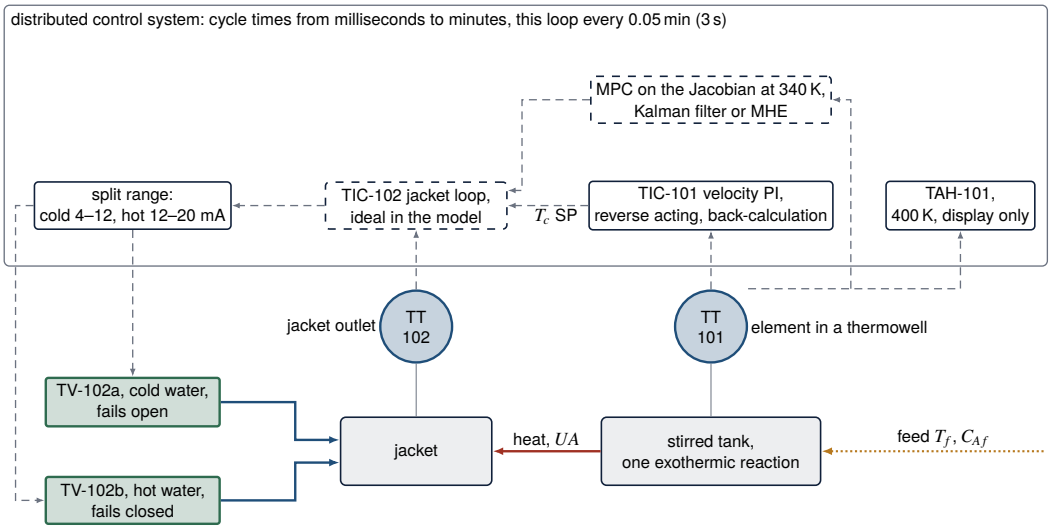


Figure 2.7 The reactor temperature loop on plant hardware. A temperature element in a thermowell (TT-101) feeds the velocity-form PI loop TIC-101 in a distributed control system, or the dashed predictive controller and its estimator when they are selected. Either one sets the jacket temperature, which a plant holds with a slave loop (TIC-102, dashed) driving a cold-water valve that fails open and a hot-water valve that fails closed through a split range. The runaway alarm only displays. Notice that the chapter's ideal jacket removes everything below the jacket setpoint, so the slave loop, its sensor and both valves are the hardware the model does not carry.

TECHNOLOGY **Final Element: Split-Range Jacket Valves**

A jacket that must both heat and cool can be fed by a hot-water valve and a cold-water valve split-ranged on one controller output, the cold valve on the lower half of the 4 to 20 mA signal and the hot valve on the upper half, so the two are never open together. Each fail position is chosen for safety. The hot valve fails closed and the cold valve fails open, so a lost controller signal or a lost air supply gives the reactor full cooling [18]. In the 2007 accident water was injected into the jacket and allowed to boil, and a supply valve failing closed or a drain valve failing open was among the single points of failure of the cooling [16].

Where the Controller Runs

A reactor temperature loop runs in a distributed control system, where supervisory and lower-level loops execute continuously with cycle times from milliseconds to minutes, and where loops can be nested so that the setpoint of one comes from the variable of another [27]. The chapter's PI loop executes every 0.05 min, or 3 s, in the velocity form with back-calculation, and the predictive controller takes the same slot when it is selected. Figure 2.7 draws the loop on plant hardware.

> **IN PRACTICE Controlling the Hottest Reading**
>
> An exothermic reactor can carry several temperature transmitters, with a high-select function passing only the hottest reading to the controller that throttles the coolant. Three transmitters on one variable can instead feed a median select, which keeps a valid measurement when one of them fails [18]. The reading is itself a point of failure. The 2007 investigation listed a faulty temperature indication among the single-point failures of the cooling system, and the procedures then in use gave no emergency instruction for a loss of cooling [16]. The chapter's one perfectly mixed temperature has no hot spot for a selector to find.

What Is Modernizing It

The first development is the predictive controller inside the control system itself. A survey of industrial practice counted the integration of predictive control into distributed control platforms among the major recent changes, and noted that an unstable process cannot in general be represented by the impulse-response models of early products, while state-space or autoregressive models can [20]. That is why the chapter's controller is built on the state-space Jacobian of Equation (2.17) rather than on a step test. The second is the estimator. The same survey found the bias update of industrial controllers the best assumption only for stable plants without disturbance information, and named the Kalman filter as the richer framework for feedback, with better performance shown for a state-space model and an optimal observer [20]. The chapter's filter carries the input disturbance δ of Equation (2.18), and its moving horizon estimator adds bounds on the states.

2.11 Limitations

Nothing in this reactor is intended for design work. The model carries two states and no vapor space, so it cannot represent boiling, pressure rise, relief or the secondary decomposition that turns an industrial excursion into an explosion, and the temperatures it reports above the boiling point of a real liquid are an index of how far the balance has diverged rather than a prediction. The lumped-parameter assumption removes the hot spot, which is the mechanism behind a large share of real runaways. The jacket is ideal, so there is no coolant residence time, no fouling and no slave loop, and the cascade structure that industrial reactors actually use cannot be built or tuned [10]. The transmitter is instantaneous, so the thermowell lag that limits achievable tuning on a real vessel is absent. The noise generator restarts from one fixed seed in every case, so a noisy run replays exactly but a learner cannot sample a second noise realization. The setpoint range runs from 280 to 450 K although the jacket limits hold the reactor only between 282.5 K and 416.4 K at the design feed, and nothing in the application marks a setpoint outside that interval as unreachable. The actuator-limit case asks for 280 K, and the reactor stops at 282.51 K, the coldest the jacket can hold. The reaction is a single first-order step with fixed stoichiometry, so nothing in the model changes when the feed composition changes except the amount of heat available. The four features that were kept, the Arrhenius rate, the two-state balance, the hard jacket limits and the short control period, are there because the loop has to be tuned around every one of them.

The predictive controller adds limits of its own. Its model is Equation (2.11) frozen at 340 K, right about the first response to a jacket move and wrong about what the Arrhenius term changes away from that point, which costs it case 8 and leaves it unable to prevent ignition on steps of 14 K and more. It has no temperature constraint, sees a feed upset only through $\hat{\delta}$, keeps its nominal model when UA or k_0 is changed, and was compared on one noise seed. Its moving horizon estimator bounds the states of that same linear model, so it corrects the sign of the concentration estimate but not the model, and after the 3 min hand-over of case 10 its temperature estimate is no closer to the truth than the filter's, 5.48 against 5.37 K root-mean-square. The relay experiment cannot measure this loop at the 0.05 min control period.

OPEN CHALLENGE When a Nonlinear Predictive Controller Pays

A 2003 survey of industrial predictive control found nonlinear applications mostly in chemicals, polymers and air and gas processing, and much smaller in size and scope than linear ones, likely because of the computing the nonlinear algorithms need. Since nonlinear processes are hard to model from plant data, it expected first-principles models to become necessary, and it named among the open questions of practice when the added expense of a nonlinear controller can be justified [20]. The linear controller of this chapter loses case 8 and cannot prevent ignition on large steps. Progress would be a test, made before commissioning, of how far a linearized model can be trusted on a plant whose gain changes sign.

2.12 Discussion

The gap between a controller that reaches setpoint and a controller that is safe on the way there is not an artifact of a small simulation. Surveys of thermal-runaway incidents report inadequate temperature control and inadequate understanding of the heat release as recurring causes, and report that most events happened during routine production rather than during an unusual operation [13, 14]. Loss prevention as a discipline grew out of exactly this observation, that the hazard is carried by the process and shows up when the operating point moves rather than when someone makes a mistake [15]. The simulation reproduces the structure of that hazard in its smallest form. The plant has an operating region that exists only under feedback, the boundary of that region is crossed by a disturbance of a few percent, and the excursion that follows is over before an operator could act.

The control literature has answered the problem in several ways that the application makes it natural to ask about. Input-output linearization cancels the Arrhenius term explicitly and delivers a loop whose gain no longer moves with the operating point [6]. Robust proportional-integral designs give conditions under which classical feedback stabilizes the reactor over a stated region despite kinetic uncertainty [7]. Gain scheduling interpolates a family of linear controllers across the envelope and is the industrial compromise between the two [8], and constrained predictive control writes the thermal limit into the optimization as a constraint [9]. None of these is implemented here as cited, since the predictive controller of Section 2.6.3 is linear and writes no thermal limit into its plan. What is implemented is the plant that motivates all of them, with the measurement needed to show that a single fixed tuning cannot cover its operating range.

Both balances are differentiable, and the ignition boundary is a property of the model rather than a constraint added to it. A nonlinear predictive controller solving Equations (2.2) and (2.3) over a receding horizon could therefore treat the alarm and the jacket range as explicit constraints instead of meeting them as a saturated actuator [28]. A moving horizon estimator on the same two equations could reconstruct C_A from the temperature history, which is the soft-sensing problem of an industrial reactor with only a thermowell. The bounded linear estimator of Section 2.6.4 marks the limit of the one-point approach, since after a runaway it corrects the sign of C_A while the plan built on it still predicts temperatures below absolute zero.

2.13 Conclusion

The reactor of this chapter is an exothermic jacketed stirred tank simulated in a browser with two states, an Arrhenius first-order reaction, a jacket temperature limited to 250 to 350 K as the manipulated variable, two feed upsets and a 400 K runaway alarm. Analysis of the shipped constants places the ignition fold at $T_c = 303.229\,\mathrm{K}$, the extinction fold at $T_c = 298.080\,\mathrm{K}$ and a Hopf bifurcation on the hot branch at $T_c = 306.220\,\mathrm{K}$, so three steady states exist only inside a jacket window 5.15 K wide and the ignited branch at the published jacket temperature of 300 K is an unstable focus rather than a second place to operate. Every reactor temperature between 335.654 K and 379.611 K is open-loop unstable, including the application's own default setpoint of 340 K. A step test in manual identifies $K_p = 2.71\,\mathrm{K/K}$ and $\tau_p = 1.29\,\mathrm{min}$ with no measurable dead time, and internal model control turns that fit into the shipped tuning of $K_c = 4\,\mathrm{K/K}$ and $\tau_I = 1.333\,\mathrm{min}$. With that tuning held fixed, an 11 K setpoint step overshoots by 12.18 K while a 12 K step overshoots by 69.07 K and trips the alarm

at every noise seed tried, and on a 14 K step reducing the gain to 1.6 K/K raises the peak from 450.78 to 460.13 K and leaves the reactor in a relaxation cycle through the alarm. Feedback absorbs a feed concentration upset that ignites the same reactor in manual, holding the peak deviation to 1.06 K against an excursion to 515.38 K. A linear predictive controller on Equation (2.11) at 340 K plans the warm-then-cold jacket move and raises the largest step kept below the alarm from 11.5 to 13.5 K, but loses at the jacket floor by 2.2 %. Its moving horizon estimator keeps the concentration nonnegative after a runaway and lowers the error after a 3 min hand-over from 54.7 to 49.6 K min, but not after a 4 min one. The natural extensions are a coolant slave loop, a thermowell lag, a choice of noise seed and a gain-scheduled or nonlinear predictive controller for comparison against the fixed tuning. The application is an accessible platform for measuring how far a single linear controller can be trusted on a plant whose gain changes sign.

KEY POINTS

- The reaction's heat climbs faster with temperature than the jacket can remove it, so the default 340 K exists only while a controller holds it.
- The steady-state gain is 1.72 K/K at the nominal point and −5.56 K/K at 340 K, yet a reverse-acting loop holds it, because a warmer jacket always warms the reactor first.
- With one fixed tuning an 11 K setpoint step overshoots by 12.18 K, while a 12 K step overshoots by 69.07 K and trips the alarm.
- A linear predictive controller raises the largest step kept below the alarm from 11.5 to 13.5 K but loses at the jacket floor.

Where else this applies. Chapter 4 controls a wheel at a friction peak where the process gain changes sign, and Volume 8, Chapter 2 holds a runaway off with the feed instead of the jacket.

References

[1] C. van Heerden, "Autothermic processes," *Industrial & Engineering Chemistry*, vol. 45, no. 6, pp. 1242–1247, 1953.

[2] O. Bilous and N. R. Amundson, "Chemical reactor stability and sensitivity," *AIChE Journal*, vol. 1, no. 4, pp. 513–521, 1955.

[3] R. Aris and N. R. Amundson, "An analysis of chemical reactor stability and control—I," *Chemical Engineering Science*, vol. 7, no. 3, pp. 121–131, 1958.

[4] U.S. Chemical Safety and Hazard Investigation Board, "Hazard investigation: Improving Reactive Hazard Management," U.S. Chemical Safety and Hazard Investigation Board, Washington, DC, Tech. Rep. 2001-01-H, 2002. [Online]. Available: https://www.csb.gov/assets/1/20/reactivehazardinvestigationreport.pdf

[5] A. Uppal, W. H. Ray, and A. B. Poore, "On the dynamic behavior of continuous stirred tank reactors," *Chemical Engineering Science*, vol. 29, no. 4, pp. 967–985, 1974.

[6] C. Kravaris and C.-B. Chung, "Nonlinear state feedback synthesis by global input/output linearization," *AIChE Journal*, vol. 33, no. 4, pp. 592–603, 1987.

[7] J. Alvarez-Ramirez and R. Femat, "Robust PI stabilization of a class of chemical reactors," *Systems & Control Letters*, vol. 38, no. 4-5, pp. 219–225, 1999.

[8] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

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

[10] M. G. Balaton, L. Nagy, and F. Szeifert, "Jacket temperature control of a batch reactor using three different temperature levels," *Industrial & Engineering Chemistry Research*, vol. 52, no. 5, pp. 1939–1946, 2013.

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

[13] P. F. Nolan and J. A. Barton, "Some lessons from thermal-runaway incidents," *Journal of Hazardous Materials*, vol. 14, no. 2, pp. 233–239, 1987.

[14] R. Saada, D. Patel, and B. Saha, "Causes and consequences of thermal runaway incidents—will they ever be avoided?" *Process Safety and Environmental Protection*, vol. 97, pp. 109–115, 2015.

[15] T. A. Kletz, "The origins and history of loss prevention," *Process Safety and Environmental Protection*, vol. 77, no. 3, pp. 109–116, 1999.

[16] U.S. Chemical Safety and Hazard Investigation Board, "Investigation report: T2 Laboratories, Inc. runaway reaction, Jacksonville, Florida, December 19, 2007," U.S. Chemical Safety and Hazard Investigation Board, Washington, DC, Tech. Rep. 2008-3-I-FL, 2009. [Online]. Available: https://www.csb.gov/assets/1/20/t2_final_copy_9_17_09.pdf

[17] ——, "CSB applauds AIChE's response to including reactive hazard awareness in college chemical engineering curriculum," News release, 2012. [Online]. Available: https://www.csb.gov/csb-applauds-aiches-response-to-including-reactive-hazard-awareness-in-college-chemical-engineering-curriculum/

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

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

[21] 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] 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.

[24] 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.

[25] K. J. Åström and T. Hägglund, "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, vol. 20, no. 5, pp. 645–651, 1984.

[26] A. Boghossian, J. Brown, and S. Zak, "3.2: Temperature sensors," in *Chemical Process Dynamics and Controls*, P. Woolf, Ed. University of Michigan open textbook, LibreTexts, 2023.

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

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

2.14 Exercises

2.1 ○ Concept check · 2 min

At the default setpoint of 340 K the steady-state gain from the jacket to the reactor temperature is −5.56 K/K. The PI loop is still reverse acting there, with the error $T^{\mathrm{sp}} - \hat{T}$ and a positive K_c, as it is on the cold branch where the gain is +1.72 K/K. Why does that loop hold 340 K?

(a) A warmer jacket always warms the reactor first, and the controller acts on that response long before the reactant runs down

(b) The integral term reverses the sign of the loop on the middle branch, so the proportional term only has to damp the response

(c) The negative gain belongs to the noise-free temperature, while the controller acts on a measurement whose gain stays positive

(d) The jacket limits keep the reactor on the cold branch, where the steady-state gain has the sign the controller assumes

2.2 Concept check · 1 min

With the jacket held at 300 K, the reactor has steady states at 324.475, 350.006 and 369.705 K. At which of them can it rest with the jacket held there and no controller?

(a) At all three, because each of them balances heat generation against heat removal

(b) Only at 324.475 K, because the middle state is a saddle and the upper one an unstable focus

(c) At 324.475 and 369.705 K, the outer two crossings of the van Heerden construction

(d) Only at 369.705 K, because the transmitter noise ignites the cold state sooner or later

2.3 Concept check · 2 min

In case 10 the predictive controller takes over a reactor that has already run away, once with the Kalman filter and once with the moving horizon estimator (MHE). Which statement agrees with Section 2.9?

(a) The MHE estimates the temperature far better, so its plan lowers the IAE after every hand-over time tried

(b) The MHE plans with the nonlinear balances, so its plan no longer predicts temperatures below absolute zero

(c) The MHE holds the concentration estimate at zero where the filter's goes negative, and its gain depends on the hand-over time

(d) The two estimators disagree in every case, because the MHE fits 15 executions while the filter uses only the last

2.4 Theory and calculation · 7 min

Take the default constants: $q/V = 1/\text{min}$, $m\Delta H/(\rho C_p) = 209.205\,\text{K}$ per mol/m^3, $UA/(V\rho C_p) = 2.09205/\text{min}$, $k_0 = 7.2\times10^{10}/\text{min}$, $E/R = 8750\,\text{K}$ and $C_{Af} = 1\,\text{mol/m}^3$.

(a) Find the rate constant k and the steady concentration at $T = 340\,\text{K}$. *Checkpoint:* $k = 0.4793/\text{min}$ and $C_A = 0.6760\,\text{mol/m}^3$.

(b) Evaluate the four entries of the Jacobian of Equation (2.11) at this state. *Checkpoint:* -1.4793, -0.02453, 100.28 and 2.0388.

(c) Find its eigenvalues and the steady-state gain from the jacket to the reactor temperature, $G = -cA^{-1}b$ with $b = [0\ \ 2.0921]^{\mathsf{T}}$ and $c = [0\ \ 1]$. *Checkpoint:* $+1.0766$ and -0.5171/min, and $G = -5.56\,\text{K/K}$.

2.5 Theory and calculation · 8 min

Use the model A_m, b_m of Equation (2.17) at 340 K, with the reactor temperature as the output and the loop closed with the reverse-acting law $u = K_c e$ or the continuous PI law $u = K_c(e + \int e\,\mathrm{d}t/\tau_I)$. Ignore the control period.

(a) Write the transfer function from the jacket to the reactor temperature and find the smallest proportional gain that stabilizes the loop. *Checkpoint:* $2.0921(s+1.4793)/(s^2 - 0.5595s - 0.5567)$, and $K_c > 0.267\,\text{K/K}$.

(b) With $\tau_I = 1.333\,\text{min}$, find the characteristic polynomial and the smallest stabilizing K_c from the Routh test. *Checkpoint:* $K_c > 0.57\,\text{K/K}$.

(c) The challenge loads $K_c = 0.8\,\text{K/K}$. Is that gain stable at 340 K, and why can it still fail after a feed upset? *Checkpoint:* Stable, yet the bound is 1.04 K/K at 345 K.

2.6 Theory and calculation · 5 min

The step test of Section 2.5 gives $K_p = 2.71\,\text{K/K}$, $\tau_p = 1.29\,\text{min}$ and $\theta_p = 0$.

(a) Find the internal model control tuning with $\tau_c = 0.1\,\tau_p$. *Checkpoint:* $K_c = 3.7\,\text{K/K}$, $\tau_I = 1.29\,\text{min}$, $K_I = 2.9/\text{min}$.

(b) The loop runs at steady state at 340 K with the jacket at 302.818 K and $K_c = 4\,\mathrm{K/K}$. Find the smallest setpoint step whose proportional kick puts the jacket on its 350 K limit at the first execution. *Checkpoint:* 11.8 K.

(c) While the jacket is clipped, by what fraction of the gap does the integral relax at each execution with $\tau_I = 1.333\,\mathrm{min}$, and how long does it take to close half the gap? *Checkpoint:* 3.75 %, about 18 executions or 0.9 min.

2.7 App challenge · 8 min

Load scenario 2, Open-loop actuator step, and change no setting. The reactor starts in Manual at 324.5 K with the jacket at 300 K, and the jacket steps to 302 K at 1 min. Run to 20 min at the default seed and export the CSV file.

(a) Read the true temperature, column `true_pv`, at 6 min and at 20 min, and the largest value in the column. *Checkpoint:* 328.79 and 328.70 K, largest 328.85 K near 4.7 min, within 0.01 K.

(b) Compute the chord gain of the step, and find the time after the step at which `true_pv` passes 63.2 % of its change. *Checkpoint:* 2.11 K/K, and between the rows at 1.9 and 2.0 min, about 0.94 min after the step.

(c) Compare the chord gain with the tangent gain of 1.72 K/K at the start and with the 2.71 K/K of a 3 K step, and say what a 4 K step would give. *Checkpoint:* 23 % and 58 % above the tangent; no fit at all.

2.8 App challenge · 8 min

Load scenario 1, Reference operating point. Before Run, set the setpoint to 340 K and switch to Auto, so that the reactor starts at the steady state of 340 K with the jacket at 302.8 K. Use the default seed and the default tuning. Run, pause at any time between 1 and 3 min, set the setpoint to 351 K, run for 4 min more and export the CSV file. Then press the reload button, make the same settings before Run and repeat with a step to 352 K.

(a) For the 351 K run, read the largest `true_pv` after the step and the lowest `mv`. *Checkpoint:* 363 K, within 0.6 K; lowest jacket 250.0 to 254.5 K, on the limit for at most one row.

(b) For the 352 K run, read the same two values and count the rows at which `mv` is 250 K. *Checkpoint:* 405 to 423 K; 5 or 6 rows, 0.5 to 0.6 min on the limit.

(c) Explain why one kelvin of setpoint separates the two outcomes.

2.9 App challenge · 7 min

Load the challenge, Challenge · return to target, the eleventh case. The proportional gain is loaded at 0.8 K/K, the feed temperature steps from 350 to 365 K at 1 min, and the run is judged on the measured temperature from 2.64 to 4 min. Use the default seed, run each case to 4 min, and export the CSV file each time. For each run find the largest $|\mathtt{pv} - \mathtt{sp}|$ in the judged window and read the IAE tile of the scorecard.

(a) Run the case as loaded. *Checkpoint:* 15.07 K and 57.94 K min.

(b) Press Reveal suggested tuning, press Reset and run again. *Checkpoint:* 0.573 K and 2.26 K min.

(c) Press the reload button, set the Controller to MPC and run again. *Checkpoint:* 0.41 K and 1.42 K min.

Accept an error within 0.01 K and an IAE within 0.02 K min.

2.10 Application area · 4 min

The simulated transmitter is instantaneous, while a real thermowell has a time constant of tens of seconds. As a rough guide, treat the thermowell as a delay τ_d and use the limit $p\,\tau_d < 1$ of Section 1.4.7 for a loop around one unstable pole p.

(a) Find the largest such delay at 340 K, where $p = 1.0766/\text{min}$, and at 350 K, where Table 2.3 gives $p = 2.834/\text{min}$. *Checkpoint:* 55.7 s and 21.2 s.

(b) Compare both with the 3 s control period and with a thermowell of 30 s, and state what that means for operating higher on the middle branch. *Checkpoint:* The control period uses about 5 % of the allowance at 340 K.

2.11 Application area · 3 min

The adiabatic ceiling $T_f + (m\Delta H/\rho C_p)\,C_{Af}$ is the temperature the reactor would reach if all of its feed reacted with no cooling.

(a) Find the ceiling at the design feed and after the feed concentration rises to $1.2\,\text{mol/m}^3$, with $T_f = 350\,\text{K}$. *Checkpoint:* 559.2 K and 601.0 K.

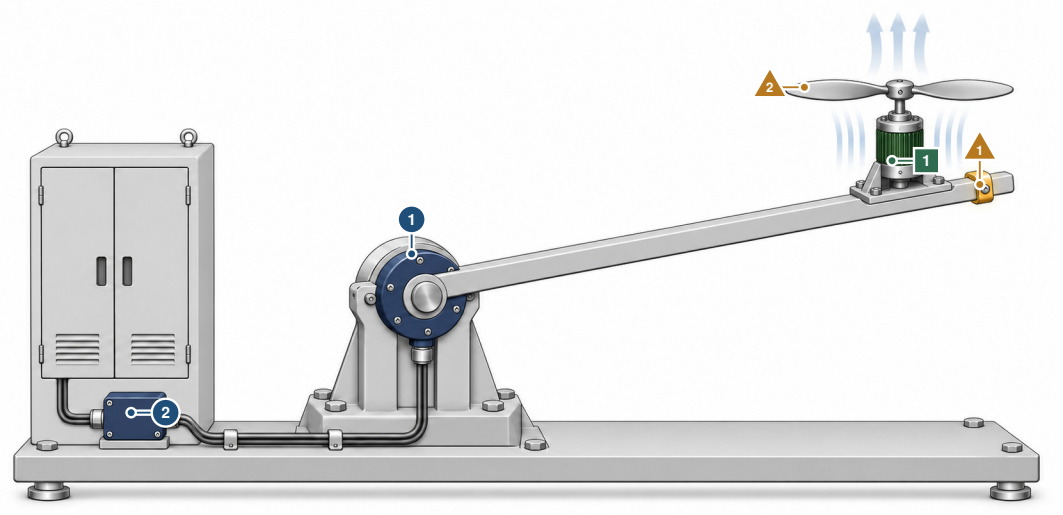
(b) With the jacket at 300 K the cold branch disappears at $C_{Af} = 1.141\,\text{mol/m}^3$. How far is the design feed from that fold, and what does the app's 400 K alarm lack that a plant's protection would need? *Checkpoint:* 14 %.

2.12 Review the agent · 5 min

An AI assistant reviews Table 2.5 and writes: "The 12 K step overshoots by 69 K because $K_c = 4$ is too aggressive near the fold. Before moving the setpoint from 340 to 354 K, lower K_c to 1.6 K/K. The linearized loop at 340 K is stable for any K_c above 0.57 K/K, so 1.6 keeps a margin of almost three, and the approach will be slower and safer." Find the flaw and say how you would check it.

CHAPTER 3

Aeropendulum Simulator in Which the Setpoint Chooses the Plant



The controller moves 1 motor and propeller, throttle sets their speed

The controller reads 1 encoder at the pivot, arm angle 2 battery voltage

What disturbs it 1 clip added at the tip 2 chipped propeller blade

Notice that the propeller sits at the far end of the arm from the encoder that reads its angle, and that the arm here is above level, where gravity pulls it away from the setpoint. Illustration, not to scale.

IN PLAIN TERMS **A Propeller Arm That Changes Its Nature**

The main idea. A small propeller at the end of a light arm lifts it about a pivot to the angle the controller asks for. Below level gravity pulls the arm back like a swing, but above level it pulls the arm farther away.

The challenge. So a controller set up at one angle can fail at another, even while its usual stability check looks fine. The propeller can only push, and once its push runs out no controller can hold the arm.

Why it matters. Aircraft, engines and many plants change their behavior with the point where they run in the same way.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on the abort vehicle (Section 3.1), the history in Section 3.2, the two try-its and the fact on the setpoint program in Section 3.6, 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 3.6), Technology in Practice (Section 3.7), Limitations (Section 3.8) and the open challenge

The control question Gravity gives a propeller arm a stiffness $G\cos\theta$ that is positive below horizontal, zero at it and negative above it, so the setpoint decides whether the plant is a pendulum, an integrator or an unstable equilibrium. What can a PID tuned at one angle do at another, and which remedy repairs which failure?

Learning objectives

- Derive the arm's stiffness, static map and actuator slope at an operating angle, and the stall angle when full thrust runs short.
- Apply the Routh condition of a PID on the linearized arm to find the angle at which a fixed tuning goes unstable.
- Explain why a healthy phase margin misses a slow mode that loses its damping.
- Compute the gains of a schedule that holds the closed-loop polynomial, and compare it with feedback linearization.
- Distinguish windup at an actuator limit from windup with nothing at a limit, and choose between back-calculation and an integral zone.

Prerequisites Sections 1.4.2, 1.4.3 and 1.4.7 (fold, schedule, feedback linearization, limits of an unstable pole), and Sections 1.6.1 and 1.6.2 (engine and velocity-form PID). Chapter 2 for a plant whose gain changes sign.

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

See also Volume 1, Chapter 5 for an air-driven actuator that holds a weight on a feedforward, and Volume 1, Chapter 4 for a microcontroller kit with its own sample time. Volume 2, Chapter 2 for a step test that depends on the operating point, Volume 2, Chapter 6 for gain scheduling and Volume 2, Chapter 3 for back-calculation at an actuator bound.

3.1 Introduction

Small electric multirotor vehicles hold themselves up on fixed-pitch rotors, and the usual way to command their thrust is to command the rotor's speed [1]. A brushless hobby motor driving such a rotor is a plant in its own right, identified from step excitation and bounded by its supply voltage [2]. The smallest machine built on this actuator is the aeropendulum, a light rod on a pivot with a motor and propeller at its free end, which has served as a low-cost experiment for classical and modern control [3] and as an open-source shield for microcontroller boards [4]. Its thrust pushes the arm one way only, and gravity decides what each push is worth.

The setpoint chooses the plant. Linearized at an angle θ^* from hanging, the weight torque on the arm changes by $G\cos\theta^*$ per radian, so in the simulator of this chapter the arm has a natural frequency of 5.95 rad/s hanging and 2.48 rad/s at 80°, behaves as an integrator at horizontal, where holding it takes 72.97 % of the throttle, and runs away from its equilibrium at 120° at +4.06 s^{-1}. A proportional-integral-derivative (PID) loop tuned at 30° shows what that does to a tuning. It settles a 3° step there with 5 % overshoot and carries the arm to 80°, and its linearized loop keeps a phase margin between 49 and 55° from 15 to 105°. The same loop is unstable from 105.8°, and a step to 120° puts the arm on its top stop 1.40 s later. The mode that escapes is a slow one near 1.6 rad/s, far below the crossover near 10 rad/s where the margin is read, and its damping ratio falls from 0.53 at 30° to 0.10 at 100°. The plant corrects a common misreading, that a healthy margin at the crossover certifies a tuning. A gain schedule reverses the picture, holding the damping between 0.38 and 0.53 while its phase margin falls to 19° at 150°.

The difficulty belongs to a general class, plants whose gain, time constants or stability move with the operating point. Gain scheduling, which builds a family of linear designs and indexes them by a measured variable, had its early applications in flight control and automotive engine control before research turned to linearization-based scheduling and linear parameter-varying designs [5], and it has been applied in fields from aerospace to process control [6]. The neutralizer of Volume 2, Chapter 6 schedules its controller on a gain that moves by orders of magnitude across one valve stroke. An

airframe flown across its envelope and an engine run across its speed range share the arm's shape, a nonlinear plant whose linearization has to be redone wherever the loop goes. A model identified at one setpoint describes that setpoint only, which case 2 below measures.

> ✦ DID YOU KNOW? **Unstable for Most of an Abort**
>
> A crew capsule's launch abort vehicle must abort safely at any time during the atmospheric ascent, and its aerodynamics leave it unstable for most of the abort flight. Its controller gains are designed on linear models and scheduled in time, altitude and Mach number, then tested in nonlinear simulation [7].

A study of one designed law on one rig, such as an H_∞ controller about an identified operating point [8] or model reference control of a linear parameter-varying model [9], leaves open the angle at which a simpler tuning stops working and the mode through which it fails, and on hardware finding them ends on the stop. This chapter works out, on a simulated desktop kit whose every equation and constant is given here, how the sign of one stiffness decides what a tuning made at one angle can do at another, and which standard remedy repairs which failure. The model, the PID, the feedforward, the schedule and feedback linearization are all standard methods. Section 3.2 traces the rig and its ideas. Section 3.3 gives the kit as modeled, with its quantized encoder and throttle and a one-sided motor lag. Section 3.4 linearizes the arm at every angle. It replaces the one-sided lag by its mean within 1.3 % in gain and 1.8° in phase, and it derives a closed form, the Routh condition of the PID on the lag-free linearized arm, whose edge lies within 0.3° of the sampled loop's at four proportional gains and which places the escaping mode near $1/\sqrt{\tau_I \tau_D}$. Section 3.5 gives the controller as coded, with a loop view computed from the eigenvalues of the sampled closed loop and checked against the running arm. Section 3.6 reports the nine cases and the graded program as measured, one remedy at a time, with one case that no controller can meet. Section 3.7 describes the sensors, drives and controllers of real rigs, Section 3.8 states what the model leaves out, and Section 3.9 sums up.

3.2 Propeller Arms on the Control Bench

Control courses have used the aeropendulum as a hands-on experiment. A small dc motor at one end of a light rod drove a 2 in propeller, a potentiometer at the pivot measured the angle, and a custom board connected the rig through a virtual serial port to a MATLAB/Simulink module sampling at up to 200 Hz [3]. On it senior mechanical engineering students designed PID and lead-lag controllers, and state-space controllers combined with feedback linearization, in a semester of assignments that needed neither a specialized laboratory nor teaching assistants [3].

The aeropendulum then became an object of dynamics research. Digitization and time delay modify the stability of feedback-controlled systems, and an analytical, numerical and experimental study of the aeropendulum estimated its parameters from a nonlinear model, defining damping coefficients of order zero, one and two, and compared the stability charts of several controllers, finding the larger stable region for an act-and-wait controller [10]. The pivot of the simulator keeps the first two orders of damping and leaves the third out, as Section 3.3.5 states.

Home-made and open-source versions followed. Recommendations for building a home-made prototype from chosen structural materials and electronics, with ideas for identifying a linear model and observing stability, were offered so that a working system could be kept at home [11]. An open-source reference design then packaged the aeropendulum as a swappable shield for Arduino boards, built from off-the-shelf electronics and 3D-printed parts, with an interface for the Arduino IDE, MATLAB and Simulink and classroom examples of both its use and its limitations [4]. The Temperature Control Lab of Volume 1, Chapter 4 is the same idea in heat transfer, a microcontroller benchmark whose cycle time, discrete sampling, communication overhead and model mismatch are its point [12]. The kit simulated here keeps a microcontroller's sample time, its 8-bit throttle and its 12-bit encoder.

Designs for the rig grew past the single PID. A prototype identified with the identification tools of MATLAB was stabilized about an operating point by an H_∞ controller and made to track a square wave of angle [8], and model reference controllers for linear parameter-varying systems, designed with a parameter-dependent Lyapunov function, were evaluated on an aeropendulum in real time [9].

Behind these designs stand two older ideas. Gain scheduling had proven to be a successful methodology before an analysis could guarantee its robustness, performance or even nominal stability, and the analysis published in 1990 treated scheduling on a reference trajectory and scheduling on the plant output, confirming that the scheduling variable should vary slowly [13]. The arm's schedule can follow either the setpoint or the angle the encoder reads, and case 5 measures the difference. Feedback linearization grew from conditions under which a nonlinear system is equivalent to a controllable linear one [14], and within a decade it was a textbook method [15]. On the aeropendulum it was taught on the hands-on rig of 2012 [3], and an integrator against constant disturbances has since been added to it, with the integrator's initial condition chosen so that the closed loop keeps its intended response [16].

> ⧖ HISTORY **Inverting the Airframe, Mid-1980s to Late 1990s**
>
> In aircraft flight control, feedback linearization goes by the name dynamic inversion [17]. It joins the vehicle's nonlinear equations of motion with onboard models of its aerodynamics, mass properties and engine to find the control combination that gives a commanded motion, while outer loops shape the response and give robustness to model errors. In the mid-1980s it was a new theoretical approach, proposed for a research aircraft flying at high angle of attack in the nonlinear, uncertain region past the stall. Fitting its laws into the very limited computers of the time was a significant challenge, and that design never flew. In the late 1990s a crew rescue vehicle program chose it because a change in the vehicle's shape needed only an update of the onboard aerodynamic model. A production aircraft now flies it from conventional flight to hover [18].

3.3 The Kit as Modeled

The simulated kit is a uniform carbon rod of mass m_r and length L hanging from a pivot on a desk stand, with a brushless motor and a three-inch propeller, of mass m_m together, at l_p from the pivot, as shown in Figure 3.1. The propeller blows sideways in the plane of the swing, so its thrust acts perpendicular to the arm and only toward larger angles. The angle θ is measured from hanging, 0 down, 90° horizontal and 180° up, and the three integrated states are θ, the rate $\dot{\theta}$ and the propeller speed W. The rig and every threshold are teaching values sized like the desktop kits that courses build around a microcontroller, not the data sheet of a product, and the application states this in its model panel and on its Instructions page. Table 3.1 sorts the signals by role.

3.3.1 The Arm About Its Pivot

The rod is uniform and the motor and propeller are a point mass, so the moment balance about the pivot is Equation (3.1).

$$J\,\ddot{\theta} = l_p T - G\sin\theta - b_f\,\dot{\theta} - \tau_{c,f}\,\operatorname{sgn}\dot{\theta} + \tau_{fan}(t) \tag{3.1}$$

where T is the thrust and the inertia and the weight torque at horizontal are given by Equation (3.2).

$$\begin{aligned} J &= \tfrac{1}{3}m_r L^2 + m_m l_p^2 + m_c L^2 \\ G &= g\left(\tfrac{1}{2}m_r L + m_m l_p + m_c L\right) \end{aligned} \tag{3.2}$$

Here m_c is a clip at the tip, $b_f = b\,(1 + f/100)$ and $\tau_{c,f} = \tau_c(1 + f/100)$ are the pivot's viscous and Coulomb friction raised by the friction disturbance f, and $\tau_{fan}(t) = \tau_0 + \tau_1 \sin(2\pi t/p_{fan})$ is the desk fan, a steady push and a slow swing of period p_{fan}, positive toward larger angles. Every term is in N m. The defaults give $J = 0.003\,383\,\mathrm{kg\,m^2}$ and $G = 0.1197\,\mathrm{N\,m}$, and the hanging swing has a damping ratio of $b/2\sqrt{GJ} = 0.025$. Within one integration step the Coulomb torque keeps the

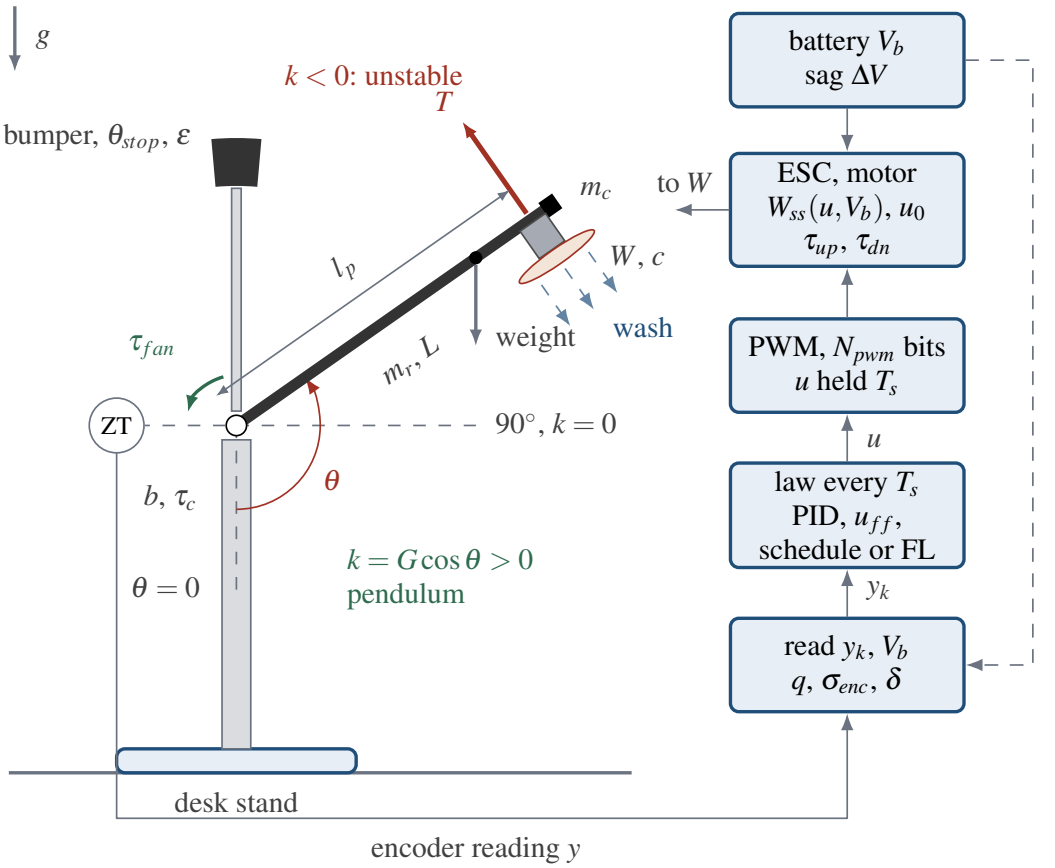


Figure 3.1 The simulated desktop aeropendulum and the kit's chain from the encoder to the propeller, not to scale. The arm is drawn at $\theta = 125°$ from hanging, past horizontal, where the stiffness $k = G\cos\theta$ is negative. A uniform rod (m_r, L) carries the motor and a pusher propeller at l_p, whose thrust T acts perpendicular to the arm toward larger θ while its wash streams the other way. The weight acts at the center of mass, a clip m_c can be added at the tip, and a chip removes the share c of the thrust. The pivot has viscous and Coulomb friction b and τ_c, a desk fan adds the torque τ_{fan}, and a rubber bumper on a bracket stops the arm at θ_{stop} with restitution ε. Every T_s the controller reads the encoder ZT (count q, noise σ_{enc}, offset δ) and the battery V_b and writes an N_{pwm}-bit duty u, which the speed controller and motor turn into the propeller speed W through the dead band u_0 and the lags τ_{up} and τ_{dn}.

sign it had when the step began, which keeps the right-hand side smooth. Between steps a rate that has changed sign is set to zero, and the arm then stays put until the net torque on it exceeds $\tau_{c,f}$. A rubber bumper above the pivot blocks the top 10° of the circle. An arm found at $|\theta| \geq \theta_{stop} = 175°$ is returned to the stop, and an arrival faster than $\omega_{rest} = 0.05\,\mathrm{rad/s}$ is reversed with restitution $\varepsilon = 0.3$ and counted as a hit, while a slower one is held.

Table 3.1 Signals of the Kit by Role

Symbol	Role	Nominal	Range
θ	CV and state, arm angle	per case	SP 0 to 170°
$\dot{\theta}$	state, arm rate	0	—
W	state, propeller speed	holds θ	0 to 20 000 rpm
u	MV, throttle duty	holds θ	0 to 100 %, 255 counts
y	measured, encoder angle	θ	4096 counts a turn
V_b	measured, battery	7.4 V	6.0 to 8.4 V
τ_{fan}	DV, desk fan, push and swing	0	±15, 0 to 15 mN m
m_c	DV, clip at the tip	0	0 to 10 g
ΔV	DV, battery sag	0	0 to 1.4 V
c	DV, thrust lost to a chip	0	0 to 50 %
$\Delta\dot{\theta}$	DV, tap on the arm	—	+30 °/s event
δ	DV, encoder offset	0	±10°
f	DV, pivot friction rise	0	0 to 500 %
T, T^*, T_f	outputs, thrust, held, full	—	—
k	output, stiffness $G\cos\theta$	—	±0.120 N m/rad
ζ, PM	outputs, loop at the setpoint	—	—

3.3.2 A Propeller That Only Pushes

The electronic speed controller drives the motor up and does not brake it, so the propeller speed follows its steady value through the one-sided lag of Equation (3.3).

$$\tau\dot{W} = W_{ss} - W, \qquad \tau = \begin{cases} \tau_{up}, & W_{ss} > W \\ \tau_{dn}, & \text{otherwise} \end{cases}$$
$$W_{ss} = W_{max}\,\frac{V_b}{V_0}\,\frac{\max(0,\ u - u_0)}{100 - u_0} \tag{3.3}$$

where u is the duty in percent, $u_0 = 8\,\%$ the dead band, $\tau_{up} = 30\,\text{ms}$ the spin-up lag, $\tau_{dn} = 75\,\text{ms}$ the coast-down lag, and $W_{max} = 20\,000\,\text{rpm}$ the full-throttle speed at the rated $V_0 = 7.4\,\text{V}$. The battery voltage is $V_b = V_{bat} - \Delta V$. The thrust is the square of the speed, as shown in Equation (3.4).

$$T = k_T(1-c)\,W^2, \qquad k_T = \frac{T_{max}}{W_{max}^2}$$
$$T_f = T_{max}\left(\frac{V_b}{V_0}\right)^2(1-c) \tag{3.4}$$

where c is the share of thrust a chipped propeller has lost, $T_{max} = 0.80\,\text{N}$ is full thrust on a fresh pack, $k_T = 1.824 \times 10^{-7}\,\text{N s}^2/\text{rad}^2$, and T_f is full thrust as the battery and the propeller now allow. The thrust is never negative, and below the dead band the propeller's steady speed is zero.

Holding the arm still at θ^* takes the thrust and the throttle of Equation (3.5), which is the static map of the rig.

$$T^* = \frac{G\sin\theta^*}{l_p}, \qquad u^* = u_0 + (100 - u_0)\sqrt{T^*/T_f}$$
$$\theta_{stall} = \arcsin\frac{T_f\,l_p}{G} \tag{3.5}$$

The held thrust peaks at horizontal, 0.3989 N or 72.97 % of the throttle, so the same throttle holds θ^* and $180° - \theta^*$, 68.46 % for both 60° and 120°. When full thrust falls below the thrust that holds the arm horizontal, the angles around 90° are not equilibria at all, and the highest angle that full throttle holds is the stall angle. Horizontal is the fold of Section 1.4.2, the angle where the holding input is largest, and each throttle below the peak holds two equilibria, one stable and one not.

Table 3.2 Model Constants. Class V Is a Slider in the Model Dialog and H a Hidden Constant

Symbol	Meaning	Value	Range	
m_r	rod mass	0.020 kg	0.005 to 0.08	V
L	rod length	0.32 m	0.2 to 0.5	V
m_m	motor and propeller	0.030 kg	0.01 to 0.08	V
l_p	motor radius	0.30 m	0.15 to 0.5	V
b	pivot viscous friction	1.0 mN m s/rad	0 to 10	V
τ_c	pivot Coulomb friction	0.05 mN m	0 to 1	V
g	gravity	9.81 m/s^2	1.6 to 24.8	V
T_{max}	full thrust at V_0	0.80 N	0.3 to 2	V
u_0	throttle dead band	8 %	0 to 20	V
τ_{up}	spin-up lag	30 ms	5 to 200	V
τ_{dn}	coast-down lag	75 ms	5 to 400	V
W_{max}	full speed at V_0	20 000 rpm	fixed	H
V_0	rated voltage, two cells	7.4 V	fixed	H
V_{bat}	battery voltage	7.4 V	6.0 to 8.4	V
θ_{stop}	top stop	175°	fixed	H
ε	bumper restitution	0.3	0 to 0.9	V
ω_{rest}	resting-contact rate	0.05 rad/s	fixed	H
N_{enc}	encoder resolution	12 bit	8 to 16	V
σ_{enc}	encoder noise	0.5 counts	0 to 3	V
N_{pwm}	throttle resolution	8 bit	6 to 12	V
J, G	inertia, weight torque	0.003383, 0.1197	derived	
T_{90}^*	thrust to hold 90°	0.3989 N	derived	

3.3.3 What the Microcontroller Reads and Writes

The controller never reads the true angle, the propeller speed, the thrust, a clip, a fan or a chip. Every sample interval T_s it reads the magnetic encoder at the pivot and the battery, as shown in Equation (3.6).

$$y_k = q\,\mathrm{rnd}\Big(\frac{\theta(t_k)+\delta}{q} + \sigma_{enc}\,n_k\Big), \qquad q = \frac{2\pi}{2^{N_{enc}}} \tag{3.6}$$

where rnd rounds to the nearest integer, n_k is a standard normal deviate from a generator seeded at 20261001, $\sigma_{enc} = 0.5$ counts, δ is the offset of a slipped magnet, and $N_{enc} = 12$, so one count is 0.0879°. The battery reading is exact. The duty the controller writes is rounded to the throttle's $N_{pwm} = 8$ bits and held until the next sample, as shown in Equation (3.7).

$$u_k = \frac{100}{N}\,\mathrm{rnd}\Big(\frac{N u_{cmd}}{100}\Big), \qquad N = 2^{N_{pwm}} - 1 \tag{3.7}$$

One count is 0.392 % of the throttle. Near horizontal that is a large step, because count 185 holds 80.8°, count 186 holds 87.7° and count 187 holds no angle at all.

3.3.4 Constants, Initial States and Time Stepping

Table 3.2 lists the constants. The framework's engine (Section 1.6.1) integrates the three states with fourth-order Runge–Kutta steps of 1 ms, calls the controller on that clock, which runs its law once T_s has elapsed, and logs at 50 Hz. The step specific to this app is the friction latch and the top stop, applied after each integration and before the outputs are computed, and a run plays at 0.25 to 5 times real time. A case starts with the arm hanging, at rest on its setpoint with the propeller at the speed that holds it, or at rest where the manual throttle, rounded to the throttle's counts, holds it. The encoder noise is seeded, so every run replays exactly, and Reset and the reload button act as in every framework app. The application passes its 40 verification checks, among them energy

conserved to 1.8×10^{-11} of G over 10 s without thrust or friction, swing frequencies within 0.04 % of $\sqrt{G\cos\theta^*/J}$, and, with the encoder noise off, halving the step moving the overshoot of case 3 by 0.0072°.

3.3.5 Assumptions

One rigid arm in a vertical plane. The rod does not bend, the stand does not move, and the motor and propeller are a point mass at l_p.

A pivot with two orders of damping. The pivot's friction is viscous plus Coulomb, with no static level above the sliding one. The quadratic, second-order damping that model-based estimation found on a physical aeropendulum [10] is absent, and a real bearing adds stiction that makes an arm stick and slip near its setpoint.

Thrust from speed alone. The thrust is $k_T W^2$ with a constant coefficient, with no correction for the air that the arm's own motion moves through the rotor, no wash on the arm and no wall effect. On multirotors, thrust control has been proposed in place of speed control because a rotor held at one speed keeps its thrust poorly in updrafts and downdrafts [1]. On a swinging arm the same inflow would add damping beyond the pivot's damping ratio of 0.025.

A first-order motor on an ideal battery. The steady speed is linear in the throttle above the dead band and in the voltage, the speed controller has no current limit or dynamics of its own, the motor does not heat, and the sag is imposed as a disturbance instead of following from the current drawn. The lags are 30 and 75 ms, half the 60 and 150 ms that the application's plan proposed. With the plan's values the best of 360 PID tunings on the application's own loop analysis damps the arm at 30° to only $\zeta = 0.29$, and halved the same search reaches 0.57, so that the cases show a damping that changes with the angle rather than one that is light everywhere.

A fan that pushes the same way at every angle. The fan is a torque about the pivot independent of where the arm points, while a real fan's push depends on the angle and is gusty.

A kit that computes instantly. The encoder adds Gaussian noise of half a count before rounding and has no eccentricity, the throttle only rounds, and the controller computes at the sampling instant with no jitter.

3.4 One Arm Linearized at Every Angle

The linearizations of this section are the analysis model behind the application's Loop view. They are built on the true rig with the controller exactly as it runs, while Equations (3.1) to (3.7) remain the simulation model and the declared rig of Section 3.5 remains the controller's own model, which knows the measured battery but no clip, chip or added friction.

3.4.1 Stiffness, Static Map and Actuator Slope

About an equilibrium θ^* held by T^*, small deviations of the throttle reach the angle through the transfer function of Equation (3.8).

$$P(s) = \frac{\Delta\theta(s)}{\Delta u(s)} = \frac{K_u}{(\bar{\tau}s+1)(Js^2+bs+k)} \tag{3.8}$$

where $\bar{\tau}$ is the motor's lag for small swings (Section 3.4.2), $k = G\cos\theta^*$ in N m/rad is the arm's stiffness and $K_u = 2l_p\sqrt{T^*T_f}/(100-u_0)$ in N m per percent is the slope of the static map, the throttle's grip on the thrust torque. Below horizontal $k > 0$ and the arm is a lightly damped pendulum with natural frequency $\sqrt{k/J}$, 5.95 rad/s hanging, 5.54 rad/s at 30° and 2.48 rad/s at 80°. At horizontal $k = 0$ and the arm is an integrator with a slow pole at $-b/J = -0.30\,\mathrm{s}^{-1}$. Above horizontal $k < 0$ and the equilibrium has a root at $-b/2J + \sqrt{b^2/4J^2 - k/J}$, $+4.06\,\mathrm{s}^{-1}$ at 120° and $+5.39\,\mathrm{s}^{-1}$ at 150° (4.21 and $5.54\,\mathrm{s}^{-1}$ without the pivot's friction), so it is unstable by the linearization alone [19]. The slope K_u grows with the square root of the thrust held, from 2.61 mN m per percent at 30° to 3.68 mN m at 90°, so the throttle has the least grip near hanging. The static gain K_u/k grows without bound as the

arm approaches horizontal, which is why a single count of the throttle moves the arm 6.9° between counts 185 and 186. The application's Map view draws these poles against the setpoint.

3.4.2 The Motor's Two Lags as One

The one-sided lag of Equation (3.3) is positively homogeneous in its input, so its describing function depends on frequency alone. Computed by integrating the lag through twelve periods of a sinusoid, it lies within 1.27 % in gain and 1.79° in phase of a single lag at the mean $\bar{\tau} = (\tau_{up} + \tau_{dn})/2 = 52.5\,\text{ms}$ from 1 to 30 rad/s, the worst at 30 rad/s, which is the lag the analysis uses. The one-sided lag still matters for large moves, where the propeller coasts from 13 065 rpm toward 1244 rpm in case 6 while gravity, not the motor, brings the arm down.

3.4.3 The Edge of a Fixed Tuning

Without the motor lag and the sampling, a PID with gains K_c, $K_i = K_c/\tau_I$ and $K_d = K_c\tau_D$ on the linearized arm closes the loop of Equation (3.9), whose Routh condition is Equation (3.10).

$$J s^3 + (b + K_d K_u)\, s^2 + (k + K_c K_u)\, s + K_i K_u = 0 \tag{3.9}$$

$$(b + K_d K_u)(k + K_c K_u) > J K_i K_u \tag{3.10}$$

Every term is in N m/s per radian. The derivative adds damping and the proportional gain adds stiffness to what gravity supplies, while the integral asks for both. At the edge the cubic has roots at $\pm j\omega_0$ with $\omega_0^2 = (k + K_c K_u)/J$, and because b is small against $K_d K_u$ the condition reads approximately $k + K_c K_u > J/(\tau_I \tau_D)$ with $\omega_0 \approx 1/\sqrt{\tau_I \tau_D}$. The closed loop needs a total stiffness above the inertia divided by the product of the two times.

For the shipped tuning, $K_c = 0.2\,\%/°$, $\tau_I = 0.4\,\text{s}$ and $\tau_D = 1.0\,\text{s}$, the right side of Equation (3.10) divided by $b + K_d K_u$ is 0.0082 to 0.0083 N m/rad from 30 to 120°, close to $J/(\tau_I \tau_D) = 0.0085\,\text{N m/rad}$, while $K_c K_u$ is 0.0299 N m/rad at 30° and 0.0422 N m/rad at 90°. Below horizontal gravity's stiffness, up to 0.120 N m/rad hanging, supplies most of the left side. Past horizontal it subtracts, and the condition fails at 106.07°, against 105.79° for the sampled loop with the motor lag and the hold, with the escaping pair at 1.56 rad/s against 1.61 rad/s, and $1/\sqrt{\tau_I \tau_D} = 1.58\,\text{rad/s}$. More proportional gain with the same times moves the edge, to 124.74, 140.62 and 160.70° for 0.4, 0.6 and 1.0 %/°, against 124.48, 140.40 and 160.57° for the sampled loop, which costs damping at 30° (0.36, 0.27 and 0.16 against 0.53). The lag-free condition places the sampled edge within 0.3° at all four gains, so at the shipped lag and sample time the edge is a property of the PID's structure against a negative stiffness rather than of the lag or the sampling.

3.4.4 Margin and Damping Part Ways

The Loop view decides stability from the eigenvalues of the sampled closed loop, the plant of Equation (3.8) held over each sample with the controller's states, and maps them by $s = \ln z/T_s$ to give each mode's damping ratio. It reports the least damping ratio ζ among the oscillating modes beside the phase margin at the last frequency where $|\mathscr{L}|$ falls through one, for $\mathscr{L}(j\omega) = C(\mathrm{e}^{j\omega T_s})H(j\omega)P(j\omega)$ with the controller as it runs and the hold H. Figure 3.2 draws both against the setpoint. The two answer different questions. Near the crossover, 9.3 to 11.4 rad/s from 15 to 105°, the arm's inertia term $J\omega_c^2$, 0.29 to 0.44 N m/rad, outweighs any stiffness, so the phase there barely depends on the angle, and the fixed tuning's margin stays between 49.2 and 55.4°. The slow mode lives where the stiffness counts. Up to 74° the least damped mode is the fast pair, at 7.6 to 12.6 rad/s with ζ between 0.48 and 0.53 from 15°, and from 75° it is the slow pair at 1.4 to 1.6 rad/s, whose ζ falls almost linearly to zero at 105.8°. From 100° on, the open loop crosses the negative real axis near 1.6 rad/s with a gain of 1.42 at 100°, 1.04 at 105° and 0.99 at 106°, so the loop is stable past horizontal only while it has gain to spare at that low frequency, a margin that no reading at the crossover reports.

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

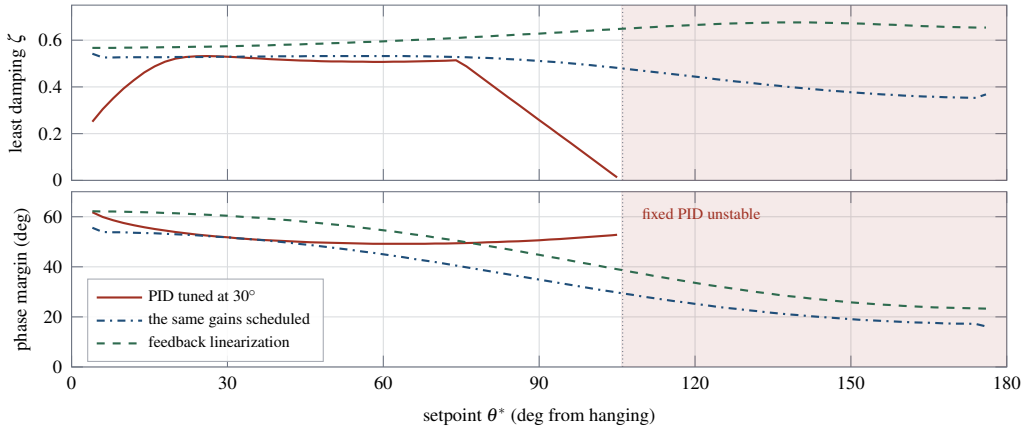


Figure 3.2 Least damping ratio and phase margin of the linearized sampled loop against the setpoint, from the application's own loop analysis at 10 ms samples. The PID tuned at 30° (solid) keeps its margin while its slow mode loses its damping and goes unstable at 105.8° (shaded), near the lag-free Routh edge at 106.07° (dotted vertical line). Scheduled (dash-dotted), the same gains keep ζ between 0.38 and 0.53 while the margin falls to 19.1° at 150°. Feedback linearization is dashed.

The schedule of Section 3.5 moves the curves the other way. It holds the three lower coefficients of Equation (3.9) at their values at the design angle, so the lag-free loop has the same poles at every angle and the sampled loop keeps ζ between 0.377 and 0.532 from 15 to 150°. To do so it raises the proportional gain to supply the stiffness that gravity takes away, and the phase margin at the crossover falls from 53.4 to 19.1°, a smaller delay margin that case 8 meets as a shorter allowable sample time. Feedback linearization keeps ζ between 0.569 and 0.676 with a margin falling from 61.7 to 25.8°.

3.5 The Kit's Controller as Coded

Table 3.3 lists the settings. The controller works from the encoder reading and the battery voltage alone, and its own model is the declared rig, the constants of Table 3.2 at the measured battery, with no clip, no chip and the nominal friction.

3.5.1 Sampling, Throttle and Saturation

The law executes when a whole T_s has passed on the 1 ms clock, reads Equation (3.6) and the battery, and writes a command that is clamped to 0 to 100 % and rounded by Equation (3.7). The saturation flag compares the law's unclamped output with those limits and is reset each sample, which makes a law that asks the propeller to pull read as saturated. The scorecard's throttle chatter is the root mean square change of the applied duty from one sample to the next in automatic mode, and the throttle has no rate limit beyond the motor's own lag. All laws act in the same direction, more throttle when the arm is below its setpoint, which the application's sign checks confirm for every structure.

3.5.2 The Shared PID on a Bias or a Feedforward

The PID is the framework's velocity-form algorithm of Section 1.6.2, written in this chapter's symbols in Equation (3.11), with T_s for the control period and the action positive.

$$
\begin{aligned}
e_k &= \theta_{SP} - y_k, \qquad P_k = K_c e_k \\
D_k &= -K_c \tau_D \left(y_{f,k} - y_{f,k-1}\right)/T_s \\
S_k &= S_{k-1} + \left(P_k - P_{k-1}\right) + \left(D_k - D_{k-1}\right) \\
\hat{u}_k &= u_b + S_k + I_{k-1}, \qquad u_{cmd} = \mathrm{clamp}(\hat{u}_k,\ 0,\ 100) \\
I_k &= I_{k-1} + K_i e_k T_s + \left(u_{cmd} - \hat{u}_k\right) T_s/T_t
\end{aligned}
\tag{3.11}
$$

Table 3.3 Controller Settings, Defaults and Ranges

Symbol	Setting	Default	Range
—	structure	PID	sched., FL
u_b	fixed bias, feedforward off	54 %	0 to 100
—	gravity feedforward	off	on, off
K_c	proportional gain	0.2 %/°	0 to 3
τ_I	integral time, 0 is off	0.4 s	0 to 20
τ_D	derivative time	1.0 s	0 to 2
τ_f	derivative filter	0 ms	0 to 200
T_s	sample time	10 ms	2 to 100
—	anti-windup	back-calc.	zone, off
e_I	integral zone	10°	1 to 30
θ_d	schedule design angle	30°	10 to 160
—	schedule variable	measured angle	setpoint
Δu	step-test size	2 %	0.4 to 10
ω_{FL}	outer frequency of FL	5 rad/s	1 to 15
ζ_{FL}	outer damping of FL	0.9	0.2 to 1.5
$\tau_{I,FL}$	outer integral time	2 s	0 to 20
θ_s	state feedback angle scale	3°	0.5 to 20
u_s	state feedback throttle scale	10 %	1 to 50
σ_q	observer speed	1	0.1 to 10

where y_f is the reading through an exact first-order lag, $y_{f,k} = a\,y_{f,k-1} + (1-a)\,y_k$ with $a = \mathrm{e}^{-T_s/\tau_f}$ ($a = 0$ for $\tau_f = 0$), $K_i = K_c/\tau_I$, and $T_t = \tau_I$ is the tracking time of the back-calculation. The shared PID's optional slew limit is not used, so the clamped value is the output. The running sum, the limit and the back-calculation behave as Section 1.6.2 describes. The gains act on errors in radians, so 0.2 %/° is 11.46 % per radian, and the integral is updated only while K_i is not zero. The bias u_b is either the fixed 54 % that holds 30°, or the gravity feedforward of Equation (3.12), the throttle that holds the setpoint on the inverted static map of the declared rig.

$$u_{ff} = u_0 + (100 - u_0)\sqrt{\frac{G \sin\theta_{SP}}{l_p T_{max} (\hat{V}_b/V_0)^2}} \tag{3.12}$$

where $\hat{V}_b$ is the measured battery, so the feedforward is 53.94 % at 30° and 72.47 % at 80°, and 100 % where the setpoint cannot be held.

3.5.3 Back-Calculation and the Integral Zone

Back-calculation acts only while the output is clamped, and on this plant a feedforward makes every large error a transient during which nothing need be at a limit. The integral zone also stops the integral while $|e_k| > e_I$, by setting K_i to zero for that sample, which clamps nothing, suspends the back-calculation with it and leaves the accumulated I as it was. Switching the integral by a condition on the error is itself a switching defense against windup, and a variable-structure PID method of that kind has been compared with several existing ones [20]. With anti-windup off, T_t is infinite and the integral integrates whatever the throttle does.

3.5.4 Schedule, Feedback Linearization and State Feedback

The schedule rescales the three gains tuned at the design angle θ_d so that the lag-free closed loop of Equation (3.9) keeps the polynomial it had there, as shown in Equation (3.13).

$$\begin{aligned} K_c' &= \max\left(0,\ \frac{K_c K_u(\theta_d) + k(\theta_d) - k}{K_u}\right) \\ K_i' &= K_i \frac{K_u(\theta_d)}{K_u}, \qquad K_d' = K_d \frac{K_u(\theta_d)}{K_u} \end{aligned} \tag{3.13}$$

where k and K_u are evaluated on the declared rig at the angle the encoder reads, or at the setpoint if the reader chooses, held between 5 and 175°, and $k(\theta_d)$ and $K_u(\theta_d)$ are their values at θ_d. The integral and derivative gains scale as $1/K_u$, and the proportional gain also makes up the stiffness that gravity loses, $k(\theta_d) - k$, which past horizontal is an anti-spring. It is the schedule of Section 1.4.3, which holds the closed-loop polynomial of the design point, applied to a PID on the lag-free arm. From the 30° tuning the proportional gain is 0.167, 0.375, 0.632, 0.984 and 1.589 %/° at 15, 60, 90, 120 and 150°. A schedule that only scales all three gains to hold the open loop's magnitude at one frequency cannot supply that stiffness. Applied to the same tuning with the frequency at 6, 8, 10 or 12 rad/s, it is unstable from 146, 124, 110 or 106°.

Feedback linearization asks an outer PID for an angular acceleration, adds the weight torque at the measured angle and inverts the thrust square on the declared rig, as shown in Equation (3.14).

$$\begin{aligned} \alpha &= \omega_{FL}^2 e + \frac{\omega_{FL}^2}{\tau_{I,FL}} \int e\,dt - 2\zeta_{FL}\omega_{FL}\dot{y}_f \\ T_{cmd} &= \frac{J\alpha + G\sin y}{l_p}, \qquad u = u_0 + (100 - u_0)\sqrt{T_{cmd}/T_f} \end{aligned} \tag{3.14}$$

The outer loop is the shared PID of Equation (3.11) with $K_c = \omega_{FL}^2$ and $\tau_D = 2\zeta_{FL}/\omega_{FL}$, and its limits are the accelerations that zero and full thrust give at the measured angle, so its anti-windup sees the throttle's limits. An advanced option replaces the chosen structure by a linear quadratic regulator with integral action and a steady-state Kalman estimator designed at the setpoint from the declared model, a teaching comparison computed by the framework's Riccati module and designed as in Section 1.4.5.

3.5.5 Step-Test Fit

In manual mode a step test moves the throttle by Δu, rounded to the counts, and fits the readings from 2 s before the step to 8 s after it with Equation (3.15) by output error.

$$y(t) = y_0 + K\Delta u\,\phi(t - t_d) + \mathrm{e}^{-\zeta\omega_n t}(A\cos\omega_d t + B\sin\omega_d t) \tag{3.15}$$

where ϕ is the unit step response of $\omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$, zero before the step, $\omega_d = \omega_n\sqrt{1-\zeta^2}$, and the last term is the free swing the arm already had. For each (ω_n, ζ, t_d) the four linear parameters come from a least-squares solve, and the three others from a grid with a coordinate search, t_d limited to 150 ms. A fit whose dead time ends on that limit says so in its note, which then reports the dead time as a bound rather than an estimate.

3.5.6 Manual Mode and the Models Kept Apart

Manual mode hands the throttle to the operator in 0.1 % steps, rounded to the counts, with a helper that sets the throttle the static map gives for the setpoint and two step-test buttons. In automatic mode the manual station shows u_{cmd} smoothed over 0.3 s, a switch to manual hands over the command in force, and every law tracks the throttle in force, so a change of mode or structure is bumpless. The simulation model runs the true rig, the controller's laws run the declared one, the Loop view analyzes the true rig linearized with the declared controller, and the step test identifies a fourth model, second order about

Table 3.4 Scenario Sequence

#	Change from the defaults	Teaching point
1	manual, the throttle ramped to counts 138, 175, toward 186, back to 175	no throttle holds past horizontal
2	manual, step tests at 20, 45 and 80°, feedforward, $\tau_f = 30$ ms	a model of the angle
3	steps 30 to 80 to 120°	a tuning carried too far
4	feedforward on, 30 to 80 to 30°	the integral freed, and loose
5	schedule and feedforward, 15, 60, 105, 150, 15°	one tuning everywhere
6	as 5, 120 to 20 to 120°	windup with no limit
7	as 5 with the zone at 100°, fan, clip, sag	what the model misses
8	as 7 at 140°, $\tau_f = 20$ ms, taps at 10, 40, 80 ms	sampling past horizontal
9	as 5, 6.6 V, chip 40 %, 60 to 90 to 60°	an unreachable setpoint
C	program 20 to 120° with gust, clip and fan	everything at once

Table 3.5 Outcomes Read From the Running Engine

Case and condition	Result
1, counts 138 and 175	29.2 to 31.5°, 59.3 to 62.1°
1, toward count 186	past 90° at 30.94 s, top stop at 32.83 s
2, fits at 20, 45, 80°	1.14, 2.31, 7.96 °/%; 5.72, 4.78, 2.77 rad/s
3, 30 to 80°	4.25° over, 3.57 s, integral 17.94 %
3, 80 to 120°	top stop 1.40 s after the step
4, feedforward on / off	integral −0.48 / 18.01 %, over 16.79 / 4.25°
5, schedule, four steps	3.40, 2.655, 1.01, 3.46 s, no stop
5, with the zone / on the setpoint	0.515 to 1.25 s / top stop at 14.39 s
6, back-calculation / zone	drop 2.94 s, I −8.13 / 0.96 s, −1.81
7, fan, clip, sag	102.7, 93.2, 99.5 to 100.1°
8, 10 / 40 / 80 ms	0.70, 1.42 s outside ±1° / lost
8, filter 0 / 20 ms, first 5 s	chatter 2.35 / 0.74 %
9, any structure	68.2 to 79.8°, stall 73.16°
9, return, aw on / off	0.80 s / not settled after 15 s
C, shipped / suggested	lost at 140° / worst 2.26 s, 9.2°

one angle. The loop view's stability agrees with the running arm on both sides of the edge, stable at 100° and lost at 110° after a 3° step, as the application's own check confirms. The eigenvalues come from the framework's balanced Hessenberg QR routine, `shared/src/framework/core/eig.js`.

3.6 Nine Cases and a Setpoint Program

Table 3.4 sets out the cases, each a departure from the settings of Table 3.3 together with any change the text names, and Table 3.5 holds what the engine measured. The engine was stepped outside the browser at the default seed, one 1 ms step at a time as the page steps it, and times count from the start of the run. A change the text names gives one run whether it is made before Run or carried through Reset. A settling time from the 1 ms record that falls on a half-hundredth of a second is given to the millisecond.

3.6.1 By Hand Past Horizontal

Case 1 hands the operator the throttle and moves it as a hand on the slider would. Ramped to count 138 of 255, 54.12 %, the arm lifts and settles between 29.2 and 31.5° about the 30.26° that count holds, swinging with a period of 1.1375 s, 5.52 rad/s, against the 5.53 rad/s that $\sqrt{G\cos\theta^*/J}$ predicts at that angle. Count 175 settles it between 59.3 and 62.1°. From 26 s the throttle creeps a count at a time toward count 186, the last whose thrust is below the 0.3989 N that holds the arm horizontal. The arm passes 80° at 30.08 s and 90° at 30.94 s with the duty at count 186, and nothing turns it back, because past horizontal the weight torque falls as the arm rises, and it reaches the top stop at 32.83 s.

Brought back to count 175 at 38 s, the throttle that held 60°, it stays on the bumper, where 175° needs only 0.035 N, and only a duty below 27.18 % would let it fall off.

> **TRY IT IN THE APP** **1 · By Hand: 30°, 60°, Then Past Horizontal**
>
> Load the scenario and press Run without touching the throttle. The arm settles near 30° and then 60°. From 26 s the throttle creeps up a count at a time, the arm passes 90° at 30.94 s, and nothing turns it back. It reaches the top stop at 32.83 s (Table 3.5) and stays there when the throttle that held 60° returns.

3.6.2 Step Tests That Find a Different Arm

Case 2 rests the arm at 19.9°, steps the throttle five counts, 1.961 %, and fits Equation (3.15), then lets the PID carry the arm to 45° and back to manual for the same test, and at 80° steps one count down. The fits find 1.14, 2.31 and 7.96 °/% with natural frequencies of 5.72, 4.78 and 2.77 rad/s, against 1.13, 2.24 and 9.01 °/% and 5.75, 4.89 and 2.62 rad/s for the arm linearized at the middle of each step. The gain grows 7.0 times from 20° to 80°, while the stiffness the fit implies, $J\omega_n^2$, falls 4.3 times against 4.8 for $G\cos\theta$ at the middle angles, because at 80° even one count moves the arm 3.12° and the fit averages a stiffness that changes by a third across the step. The 80° fit also runs its dead time to the 150 ms limit of the search, as its note reports, and a search bounded at 300 ms finds 170 ms while moving the gain by 0.08 % and the frequency by 0.5 %. Every fitted damping ratio lies between 0.041 and 0.082, and every model is a model of its angle. The cruise controller of Volume 2, Chapter 2 finds the same with road speed in place of angle.

3.6.3 A Tuning Carried Too Far

Case 3 holds 30° under the shipped PID and steps to 80° at 2 s and to 120° at 14 s. The 50-degree step overshoots 4.25° and settles within ±2° in 3.57 s, with the integral carrying 17.94 % of the throttle that the fixed bias leaves out. The step to 120°, inside the unstable band of Figure 3.2, puts the arm on the stop at 15.40 s, 1.40 s after the step (Table 3.5), and the trend of Figure 3.3 shows it on the bumper. By 30 s it has struck the bumper 12 times. On the running loop a 3° step overshoots 5 % at 30°, 24 % at 80° and 80 % at 100°, where the arm never settles within ±1°, and at 120° it is lost.

The Loop view says where and why. The damping ratio is 0.53 at 30°, 0.42 at 80°, 0.26 at 90° and 0.10 at 100°, and the loop is unstable from 105.8°, while the phase margin reads between 49 and 55° from 15 to 105°. The proportional gain is small because gravity's spring did most of its work at 30°, 0.104 N m/rad, and Equation (3.10) shows that past horizontal the proportional gain alone must outweigh both the negative stiffness and $J/(\tau_I\tau_D)$. Raising it to 0.4, 0.6 or 1.0 %/° moves the edge to 124.5, 140.4 or 160.6° and costs the design point its damping. Over the first 30 s the unfiltered derivative on a 12-bit encoder makes the duty chatter by 2.76 %.

> **TRY IT IN THE APP** **3 · A PID Tuned at 30°, Carried to 80° and 120°**
>
> Load the scenario, open the Loop view and press Run. The step to 80° overshoots 4.25° and settles. The right panel shades the band past 106° where the same tuning is unstable, and the step to 120° lands inside it: the arm reaches the top stop 1.40 s after the step (Table 3.5).

3.6.4 Feedforward, and the Integral It Leaves Loose

Case 4 runs the same PID on the feedforward of Equation (3.12). Just before the step back from 80° the integral holds −0.48 % of the throttle against 18.01 % on the fixed bias, so the bias now follows the setpoint and the integral does not have to. The step itself is worse, 16.79° of overshoot on the way up against 4.25° and 6.88° on the way back against 0.05°, because the integral still integrates the 50-degree error, to 7.94 % at its largest. With the bias already right, that is windup with nothing at a limit.

Pause. Try Exercise 3.2 before reading on.



Figure 3.3 The application in case 3, *A PID tuned at 30°, carried to 80° and 120°*, at $t = 20.00$ s on the Loop view with a 20 s trend window and no control changed, a composite of the Loop view (top) and the arm-angle trend (bottom) cut from one screenshot. (1) The badges read θ 126.6°, setpoint 120°, throttle 62.7 % and thrust 0.333 N, with loop and plant unstable. (2) The open loop at the 120° setpoint, marked unstable, with its crossover at 8.24 rad/s and a distance from −1 of 0.43. (3) The same tuning at every setpoint, the least damping ratio solid and the distance from −1 dashed, and the unstable band shaded and labeled from 106°. (4) The step to 80° settling and the step to 120° at 14 s, with the trend header reading an encoder angle of 125.9°. (5) The arm on the bumper until 16.8 s, then swinging down to 28.1° at 18.2 s and back up, values the screen does not print.

3.6.5 One Tuning From 15 to 150 Degrees

Case 5 schedules the 30° gains on the measured angle and walks the setpoint from 15° to 60, 105 and 150° and back to 15°, 6 s apart, on the feedforward. The four steps settle within ±2° in 3.40, 2.655, 1.01 and 3.46 s with overshoots of 9.0, 9.0, 17.2 and 8.8°, and the arm never touches the stop. The same feedforward under the fixed PID settles the first step in 2.45 s and no other, with 8 hits on the stop. The overshoots are the integral's, as in case 4, and with the integral zone the schedule settles the four steps in 0.65, 0.515, 1.25 and 1.09 s. Scheduled on the setpoint instead, the gains jump to the destination's at the step while the arm is still far from it, and the step to 150° overshoots onto the stop at 14.39 s, 0.39 s after the step, the hazard that scheduling on a variable which jumps invites [13]. Feedback linearization settles the steps in 3.72, 3.18, 2.85 and 5.32 s, and with the zone in 0.955, 0.79, 0.515 and 1.49 s.

3.6.6 A Propeller That Can Only Stop

Case 6 meets the windup of case 4 on the one-sided propeller. The scheduled PID holds 120° on the feedforward, the propeller at 13 065 rpm, until the setpoint drops to 20° at 2 s. The law asks for less than nothing for 0.160 s, the propeller coasts toward 1244 rpm, and gravity takes the arm down, which released with the motor off would pass 20° in 0.369 s. Back-calculation can act only during those 0.160 s, yet the integral reaches −8.13 %, the arm undershoots to 14.4° and settles in 2.94 s, and the rise back to 120° at 10 s overshoots 32.71° and settles in 4.63 s. Without anti-windup the drop takes 3.24 s, little worse, because this windup is not the limit's. With the integral zone the integral

stays within 1.81 %, and the drop and the rise settle in 0.96 and 0.91 s with 0.90 and 1.44° past their setpoints.

3.6.7 What the Feedforward's Model Misses

Case 7 holds 100° with the schedule, the feedforward and the zone. The battery sag of 0.6 V at 21 s moves the arm least, between 99.5 and 100.1°, because the kit measures the battery and the feedforward rises from 72.47 to 78.16 % by itself. The fan's 6 mN m from 3 to 9 s lifts the arm to 102.7° and it is back within ±2° in 0.76 s, the integral settling at −2.04 %. The 5 g clip at 13 s, which adds 0.0157 N m of weight torque at horizontal, drops it to 93.2°, back in 2.405 s, because the feedforward is 4.10 % short of the 76.57 % that now holds 100°, and a 15 % chip would leave it 5.46 % short. Without the feedforward the integral carries 28.19 % by the end and the sag drops the arm to 90.2°, and without the integral the arm stays at 103.21° under the fan and at 93.79° with the clip.

3.6.8 Sample Time and Encoder Past Horizontal

Case 8 holds 140° with a 20 ms derivative filter and taps the arm at 2 s with 10 ms samples, at 8 s with 40 ms and at 16 s with 80 ms. After the tap the arm is outside ±1° for 0.70 s at 10 ms and 1.42 s at 40 ms, the least damping falling from 0.308 to 0.134, and at 80 ms it reaches the stop 2.68 s after the tap. The linearized loop at 140° is unstable from 63.0 ms, and the limit depends on the angle, 91.8, 77.2 and 60.5 ms at 60, 100 and 150°, because the schedule spends phase margin past horizontal, 36.5, 25.3 and 16.7° at 60, 100 and 140° with this filter at 10 ms. An unstable pole leaves a window of delay that narrows as the pole grows (Section 1.4.7), and the hold's half sample comes out of it. One encoder count is 0.0879°, a rate step of 8.79 °/s differenced over 10 ms. Over the first 5 s, the tap at 2 s included, the unfiltered duty changes by 2.35 % root mean square from one sample to the next, six counts, against 0.74 % with the filter and 0.34 % unfiltered on a 16-bit encoder.

3.6.9 Out of Thrust

Case 9 sags the battery to 6.6 V and chips 40 % of the thrust away, which leaves 0.3818 N of full thrust against the 0.3989 N that holds horizontal, 4.3 % short, so the stall angle of Equation (3.5) is 73.16°. On the battery alone full thrust would still be 59.5 % above the thrust that holds horizontal. Asked for 90° at 3 s, the arm swings between 68.2 and 79.8° at full throttle, a pendulum again with no loop left to damp it, and the scheduled PID, feedback linearization and the fixed PID spend 11.3 to 12.0 of the 12 s there. Anti-windup decides only the return. With back-calculation the law asks for 109.1 % just before the setpoint returns at 15 s, and the arm is back within ±2° of 60° 0.80 s later, while without it the integral winds to 89.4 %, the law asks for 179.5 %, and 15 s later the arm still swings between 64.2 and 73.2°. The blending tank of Volume 2, Chapter 3 meets a setpoint beyond its actuator in the same way, with back-calculation holding the stored output at the bound.

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

3.6.10 The Setpoint Program

The graded case starts at rest at 20° under the shipped tuning and runs the program 20, 60, 100, 140, 45 and 120°, a step every 8 s, through a 6 mN m gust from 11 to 14 s, a 5 g clip at 20 s and a fan oscillating by 1.5 mN m from 48 s. Every step and disturbance must be back within ±2° within 3.0 s and stay there, no more than 12° past any new setpoint, with no touch of the stop, the throttle saturated for at most 2 % of the 72 s and a chatter of at most 2 %, at the kit's 10 ms, 8-bit throttle and 12-bit encoder. Table 3.6 compares configurations, and Figure 3.4 draws three of its columns against their limits. As shipped the arm swings between 78.5 and 125.5° at 100°, reaches the stop at 25.06 s and strikes it 34 times, and it never settles at 100, 140 or 120°. The suggested configuration keeps the 30° gains and adds the schedule, the feedforward, the zone and a 20 ms filter. Its slowest segment, the clip, takes 2.26 s, its largest overshoot is 9.2° on the step to 140°, and over seven other seeds of the encoder noise the slowest segment takes 2.17 to 2.30 s. Each part matters, and feedback linearization

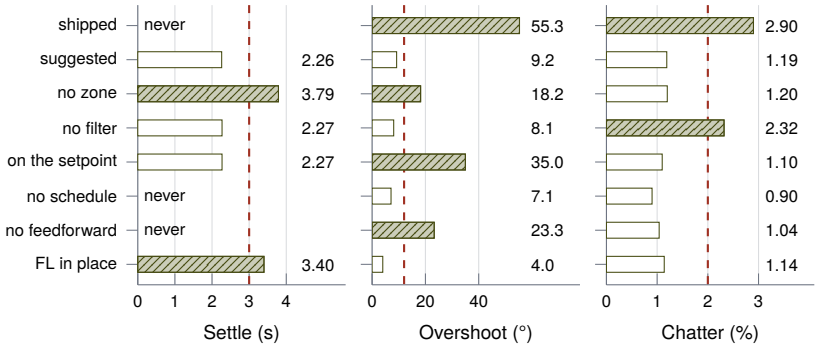


Figure 3.4 The setpoint program in the eight configurations of Table 3.6: the slowest return to ±2° (left), the largest overshoot (middle) and the throttle chatter (right), each with its limit as a dashed line. Hatched bars break that limit and white bars meet it, and three configurations never settle. Notice that the suggested configuration is the only row that is white in all three panels, while the run without the schedule has the smallest overshoot but one and never settles.

Table 3.6 The Setpoint Program in Eight Configurations

Configuration	Settle (s)	Over (°)	Sat. (%)	Chat. (%)	Stop
shipped	never	55.3	0.00	2.90	34
suggested	2.26	9.2	0.43	1.19	0
no zone	3.79	18.2	0.42	1.20	0
no filter	2.27	8.1	0.36	2.32	0
on the setpoint	2.27	35.0	0.42	1.10	1
no schedule	never	7.1	0.00	0.90	0
no feedforward	never	23.3	0.00	1.04	0
FL in place	3.40	4.0	0.17	1.14	0

Settle is the slowest segment's return to ±2°, rounded as the application reports it, Over the largest overshoot past a new setpoint, Sat. the share of the run saturated and Chat. the throttle chatter. The limits are 3.0 s, 12°, 2 %, 2 % and no stop, and only the suggested configuration meets all five.

with the zone and the filter misses only because the clip takes 3.40 s. The case is judged as given, so a change to a model constant or a disturbance voids an attempt.

> ✦ **DID YOU KNOW? A Smaller Overshoot That Never Settles**
>
> From the simulator: in the setpoint program, the suggested configuration without its schedule overshoots only 7.1°, against 9.2° with it, and never saturates the throttle. Yet it never settles, while the full configuration settles in 2.26 s (Table 3.6).

3.7 Technology in Practice

What Measures It

The one controlled variable is the arm's angle, read at the pivot. Laboratory aeropendulums read it with incremental encoders of 600 pulses per revolution [8, 16]. The chapter's kit reads a 12-bit magnetic encoder, 4096 counts a turn, with half a count of noise and an offset when its magnet slips (Section 3.3). One count is 0.0879°, a rate step of 8.79 °/s when differenced over 10 ms, which is why case 8 needs a derivative filter to keep the throttle from chattering.

> **TECHNOLOGY** **Sensor: A Rotary Encoder at the Pivot**
>
> A laboratory aeropendulum reads its rod angle with an incremental encoder of 600 pulses per revolution, read in quadrature to four times that resolution, and runs its law at a sampling period of 0.02 s [16]. Quadrature means two pulse trains a quarter period apart, so the one that leads tells a counter which way the shaft turns. A Hall-effect sensor that reads a magnet on the moving part removes the backlash and wear of a mechanical linkage [21]. Every encoder still rounds the angle to a count. Quantization acts roughly as noise whose variance is the square of one step over twelve, it can cause small oscillations, and a controller passes it on to the actuator unless it filters it [17].

What Moves It

The final element is a motor and propeller: a brushless motor behind an electronic speed controller on one laboratory rig, and a direct-current motor fed a pulse-width-modulated signal through an H-bridge on another [8, 16]. Because the drive is much faster than the swing of the arm, the second rig left its dynamics out and measured the throttle's map instead: it applied a series of throttles, read the angle at each equilibrium, computed the torque from the static balance, and fitted a cubic that converts the controller's torque back to a throttle [16]. This is the static map of Equation (3.5) measured on hardware. The chapter's motor carries that map with its dead band of 8 %, adds the 30 ms spin-up and 75 ms coast-down lags that a static map leaves out, and lets the battery set full thrust.

> **TECHNOLOGY** **Final Element: A Speed-Controlled Rotor**
>
> On small multirotor vehicles the lowest level of control is the motor and rotor, and the usual practice regulates rotor speed, turning a thrust demand into a speed through a static model of a rotor in free air. That model shows significant errors in gusts or when the rotor moves, most of all in updrafts and downdrafts. The electronic speed controller runs its own loop, typically at 1 to 2 kHz, and limits its voltage to prevent overcurrent, while its link to the vehicle's central inertial unit is far slower, typically 50 Hz [1]. The chapter's thrust, a fixed coefficient times the speed squared, is such a static model, and Section 3.3.5 names the inflow it leaves out.

Where the Controller Runs

A teaching rig may run its law on a desktop computer through a data acquisition board, as one rig does at 0.02 s, with an emergency stop button for unexpected behavior [16]. On a multirotor the speed loop runs on the electronic speed controller at 1 to 2 kHz, and the link to the central inertial unit is typically 50 Hz [1]. The chapter's kit computes on a microcontroller every 10 ms of a 1 ms clock, reads the encoder and the battery and writes an 8-bit duty. The velocity-form PID, with its schedule, feedforward and anti-windup, or the outer loop of feedback linearization, is the one function that holds the chapter's structure. Figure 3.5 draws the loop on kit hardware.

What Is Modernizing It

The first development is thrust control on the speed controller itself. A power-based thrust model, computed in real time from the measured current, bus voltage and rotor speed, with an inner current loop and an outer PI loop on thrust, held the thrust setpoint better than speed control in downdrafts and updrafts and reduced a quadrotor's tracking error in a gust [1]. It addresses the inflow that the chapter's thrust law ignores. The second is the parameter-varying design. Newer formulations, linear parameter-varying ones among them, relax classical scheduling's restriction to near-equilibrium operation with a good degree of success [6], and the model reference design cited in Section 3.2 is one tried on an aeropendulum. The third is the generated inverse. Some modeling tools generate the code for an inverse model automatically, though feedback linearization can cancel beneficial terms of the natural dynamics and must be used with care [17].

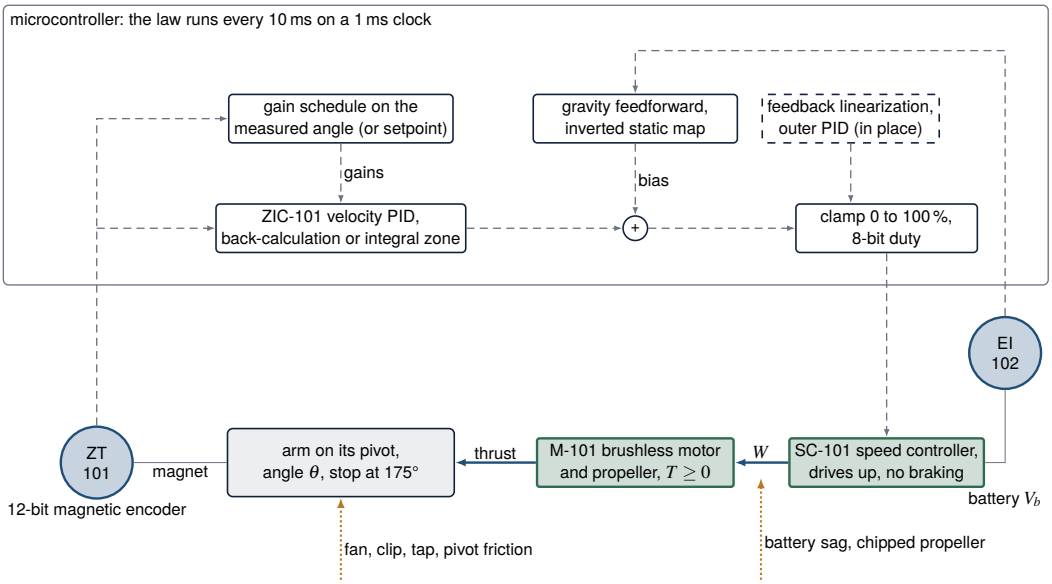


Figure 3.5 The angle loop on kit hardware. The magnetic encoder ZT-101 at the pivot feeds the velocity-form PID ZIC-101, whose gains the schedule sets and whose bias the gravity feedforward supplies from the measured battery (EI-102). Feedback linearization (dashed) takes the PID's place when it is selected. The clamped 8-bit duty goes to the speed controller SC-101, which spins the motor and propeller M-101 up but does not brake it. Notice that the thrust can only push the arm toward larger angles, so every law meets a hard floor at zero thrust as well as a ceiling at full thrust.

> ⚠ **IN PRACTICE** **An Inverse Still Needs Feedback**
>
> A design guide for reentry flight control by dynamic inversion warns that the inversion alone does not guarantee robustness, so a robust outer loop is required. Although advocates assert that no gain scheduling is needed, gain adjustments are still required whenever commands saturate the actuators, because only the vehicle's dynamics are inverted and the actuators are left out [22]. On a laboratory aeropendulum, feedback linearization without an integrator left a steady-state error from disturbances and model uncertainty, which an integrator removed [16]. The chapter's outer PID, with its integral and its limits at zero and full thrust, answers both warnings.

3.8 Limitations

The arm is rigid, the pivot has two orders of damping and no stiction, the thrust ignores the air's motion through the rotor, the motor is a first-order lag on an ideal battery, and the encoder and the throttle only round, so a real kit would damp its swing more and stick near its setpoint, and its fan would push differently at every angle. The lags and every threshold, the 3.0 s, 12°, 2 % and 2 % of the graded case among them, are teaching values rather than any kit's.

The Loop view is linear. It leaves out the rounding, the Coulomb friction and every large-signal effect, so it describes small motions about the setpoint and not a 50-degree step, and the Routh condition of Equation (3.10) is lag-free, a close estimate here because the slow mode lies far below the lags. The schedule and the feedforward trust the declared rig at every instant, and no estimator learns a clip or a chip. The step-test fit assumes a second-order mode about one angle, which fails near horizontal, where the smallest step the throttle can make crosses a third of the stiffness.

OPEN CHALLENGE **Scheduling Faster Than the Theory Allows**

Classical gain scheduling rests on two restrictions: operation near equilibrium, and a plant that moves between operating regions slowly enough. A survey found newer formulations easing the first with a good degree of success, while relaxing the slow variation restriction remained very much an open question. It judged the robustness analysis of gain-scheduled controllers very immature, noted that they often work even when the conditions are violated, and named the choice of scheduling variable as an open need [6]. In case 5 a schedule on the measured angle never touches the stop, and one on the jumping setpoint overshoots onto it. Progress would be a test that tells, before the arm finds out, which scheduling variable is safe.

3.9 Conclusion

The simulator of this chapter carries one tuning across the three plants of a single arm. Gravity's stiffness, $G\cos\theta$, makes the arm a pendulum at 5.95 rad/s hanging, an integrator at horizontal and an equilibrium that diverges at $+4.06\,\mathrm{s}^{-1}$ at 120°. A PID tuned at 30° keeps a phase margin of 49 to 55° up to 105°, yet its slow mode goes unstable at 105.8°, within 0.3° of the lag-free Routh edge. Gravity feedforward takes the integral's share at 80° from 18.01 to −0.48 % but exposes windup with nothing at a limit, which the integral zone cuts from −8.13 % to −1.81 % on the drop of case 6. The schedule holds the damping between 0.38 and 0.53 from 15 to 150° while its phase margin falls to 19.1°, and the graded program passes at 2.26 s only with the schedule, the feedforward, the zone and the filter together. Its arm is rigid, its thrust is deaf to the moving air and its kit computes without delay, and the natural extensions are inflow and stiction, a battery that sags with its current, an estimator for the clip and the chip, and a schedule identified from the reader's own step tests. The kit is a browser laboratory in which the sign of one stiffness decides what a tuning can do.

KEY POINTS

- Gravity makes the arm a pendulum below level, neutral at level and unstable above it, so the setpoint chooses the plant.
- A PID tuned at 30° keeps a phase margin of 49 to 55° up to 105°, yet its slow mode goes unstable at 105.8°, within 0.3° of the lag-free Routh edge.
- Gravity feedforward empties the integral but exposes windup with nothing at a limit, which the integral zone cuts.
- The schedule holds the damping between 0.38 and 0.53 from 15 to 150° while its phase margin falls to 19.1°.

Where else this applies. Volume 2, Chapter 6 schedules its controller on a gain that moves by orders of magnitude across one valve stroke, and Chapter 2 holds an operating point whose steady-state gain changes sign.

References

[1] M. Bangura and R. Mahony, "Thrust control for multirotor aerial vehicles," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 390–405, 2017.

[2] P. Pounds, R. Mahony, and P. Corke, "System identification and control of an aerobot drive system," in *2007 Information, Decision and Control*, 2007, pp. 154–159.

[3] E. T. Enikov and G. Campa, "Mechatronic aeropendulum: Demonstration of linear and nonlinear feedback control principles with MATLAB/Simulink real-time windows target," *IEEE Transactions on Education*, vol. 55, no. 4, pp. 538–545, 2012.

[4] A. Vargová, J. Boldocký, M. Gulan, P. Tibenský, E. Mikuláš, and G. Takács, "AeroShield: An open-source propeller-driven pendulum device for control engineering education," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 9600–9605, 2023.

[5] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

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

[7] S. J. McNamara, C. I. Restrepo, J. M. Madsen, E. A. Medina, R. W. Proud, and R. J. Whitley, "Gain scheduling for the Orion launch abort vehicle controller," in *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, 2011.

[8] R. Breganon, U. N. L. T. Alves, J. P. L. S. De Almeida, F. S. F. Ribeiro, M. Mendonça, R. H. C. Palácios, and M. A. F. Montezuma, "Model identification and H_∞ control of an aeropendulum," *Journal of Applied Research and Technology*, vol. 21, no. 4, pp. 526–534, 2023.

[9] A. Abdullah and N. Borham, "Mixed H_2/H_∞ model reference control of linear parameter-varying discrete-time systems with real-time application to aeropendulum system," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 18, pp. 9920–9938, 2022.

[10] G. Habib, A. Miklos, E. T. Enikov, G. Stepan, and G. Rega, "Nonlinear model-based parameter estimation and stability analysis of an aero-pendulum subject to digital delayed control," *International Journal of Dynamics and Control*, vol. 5, no. 3, pp. 629–643, 2017.

[11] D. M. Ovalle M. and L. F. Combita A., "On the implementation of low-cost home-made aeropendulum prototypes for improving understanding of dynamical systems concepts," in *2021 IEEE 5th Colombian Conference on Automatic Control (CCAC)*, 2021, pp. 297–302.

[12] J. Park, R. A. Martin, J. D. Kelly, and J. D. Hedengren, "Benchmark temperature microcontroller for process dynamics and control," *Computers & Chemical Engineering*, vol. 135, p. 106736, 2020.

[13] J. S. Shamma and M. Athans, "Analysis of gain scheduled control for nonlinear plants," *IEEE Transactions on Automatic Control*, vol. 35, no. 8, pp. 898–907, 1990.

[14] L. Hunt, R. Su, and G. Meyer, "Global transformations of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 28, no. 1, pp. 24–31, 1983.

[15] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[16] U. N. L. T. Alves, R. Breganon, M. Mendonça, and R. H. C. Palácios, "An integral feedback linearization applied to an aeropendulum," *Control and Cybernetics*, vol. 54, no. 1, pp. 49–77, 2025.

[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] J. T. Bosworth, "Success stories in control: Nonlinear dynamic inversion control," NASA Armstrong Flight Research Center, report DFRC-E-DAA-TN1713, preprint, NASA Technical Reports Server 20110008197, https://ntrs.nasa.gov/citations/20110008197, 2010.

[19] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[20] A. Scottedward Hodel and C. E. Hall, "Variable-structure PID control to prevent integrator windup," *IEEE Transactions on Industrial Electronics*, vol. 48, no. 2, pp. 442–451, 2001.

[21] 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.

[22] D. Ito, J. Georgie, J. Valasek, and D. T. Ward, "Reentry vehicle flight controls design guidelines: Dynamic inversion," NASA Johnson Space Center, Houston, TX, Technical Publication NASA/TP-2002-210771, 2002, nASA Technical Reports Server 20020039166.

3.10 Exercises

3.1 ○ Concept check · 2 min

Under the PID tuned at 30°, the Loop view reports a phase margin between 49 and 55° for every setpoint from 15 to 105°, yet the loop is unstable from 105.8°. Why does the margin give no warning?

(a) The crossover moves past the Nyquist frequency of the 10 ms sampling near horizontal

(b) The coast-down lag of the motor adds phase lag only past horizontal, where the thrust falls

(c) The encoder rounds away the error near horizontal, so the loop is open there

(d) The escaping mode is a slow pair near 1.6 rad/s, far below the crossover where the margin is read

3.2 ○ Concept check · 1 min

Case 4 runs the PID tuned at 30° on the gravity feedforward instead of the fixed bias of 54 %, and steps the setpoint from 30 to 80°. What does the feedforward change?

(a) The integral's share at 80° falls to near zero, but the step overshoots more
(b) The integral's share at 80° falls to near zero, and the overshoot falls with it
(c) The integral's share at 80° is unchanged, since a feedforward acts only in transients
(d) The throttle saturates on the step, and back-calculation then winds the integral down

3.3 Concept check · 2 min

Case 9 sags the battery to 6.6 V and chips 40 % of the thrust away, and then asks for 90°. Which statement holds?

(a) Feedback linearization reaches 90°, because it inverts the thrust square on the measured battery
(b) No structure reaches 90°, and anti-windup decides only how the arm returns to 60°
(c) The scheduled PID reaches 90° once its integral has wound up far enough
(d) Every structure reaches 90° slowly, because the static map still has an equilibrium there

3.4 Theory and calculation · 7 min

Take the default constants of Table 3.2: a rod of 0.020 kg and 0.32 m, a motor and propeller of 0.030 kg at $l_p = 0.30$ m, pivot friction $b = 1.0$ mN m s/rad and $g = 9.81\,\mathrm{m/s^2}$, with no clip.

(a) Compute J and G from Equation (3.2). *Checkpoint:* $J = 0.003\,383\,\mathrm{kg\,m^2}$ and $G = 0.1197$ N m.

(b) Compute the natural frequency $\sqrt{G\cos\theta^*/J}$ hanging and at 80°, and the damping ratio $b/2\sqrt{GJ}$ of the hanging swing. *Checkpoint:* 5.95 rad/s, 2.48 rad/s and 0.025.

(c) Compute the root of the equilibrium at 120°, with the pivot's friction and without it. *Checkpoint:* $+4.06\,\mathrm{s^{-1}}$ and $+4.21\,\mathrm{s^{-1}}$.

3.5 Theory and calculation · 6 min

The static map of Equation (3.5) gives the thrust and the throttle that hold the arm still. Use $G = 0.1197$ N m, $l_p = 0.30$ m, $T_{max} = 0.80$ N and $u_0 = 8$ %.

(a) Find the thrust and the throttle that hold the arm horizontal on a fresh battery, and the throttle that holds 60°. Which other angle does that throttle hold? *Checkpoint:* 0.3989 N and 72.97 %; 68.46 %, which also holds 120°.

(b) Case 9 runs at $V_b = 6.6$ V with a chip of $c = 0.40$. Find full thrust, how far it falls short of the horizontal thrust, and the stall angle. *Checkpoint:* 0.3818 N, 4.3 % short, 73.16°.

3.6 Theory and calculation · 10 min

The shipped PID has $K_c = 0.2\,\%/^\circ$ (11.46 % per radian), $\tau_I = 0.4$ s and $\tau_D = 1.0$ s. Use $J = 0.003\,383\,\mathrm{kg\,m^2}$, $G = 0.1197$ N m, $b = 0.001$ N m s/rad, $l_p = 0.30$ m, $T_f = 0.80$ N and $u_0 = 8$ %.

(a) Compute the actuator slope $K_u = 2l_p\sqrt{T^*T_f}/(100 - u_0)$ at 30 and 120°. *Checkpoint:* 2.61 and 3.43 mN m per percent.

(b) Evaluate the coefficients of Equation (3.9) at 120° for the shipped gains, and decide stability with Equation (3.10). *Checkpoint:* The s coefficient is −0.0206 N m/rad, so the loop is unstable.

(c) Compute the scheduled K_c', $\tau_I' = K_c'/K_i'$ and $\tau_D' = K_d'/K_c'$ at 120° from Equation (3.13) with the design angle at 30°. *Checkpoint:* 0.984 %/°, 2.59 s and 0.154 s.

3.7 App challenge · 8 min

Load scenario 3, A PID tuned at 30°, carried to 80° and 120°, and change no setting. The setpoint steps to 80° at 2 s and to 120° at 14 s. Use the default seed.

(a) Run to 20 s and export the CSV file. Find the overshoot past 80° from the largest `th_deg` between 2 and 14 s, and read `i_pct` at 13.90 s. *Checkpoint:* 4.25° within 0.05°, and 17.94 % within 0.1 %.

(b) In the same file, find the first row at which `stop_hits` is above zero. *Checkpoint:* 15.40 s, 1.40 s after the step.

(c) Press Reset, set Proportional gain to 0.60 %/°, and run to 30 s. Read the Top stop and the Loop at the setpoint tiles. *Checkpoint:* 0 hits, and ζ 0.14 with a phase margin of 27°.

Explain the result of (c) with the edge that Equation (3.10) gives for this gain.

3.8 App challenge · 7 min

Load scenario 6, A propeller that can only stop. The scheduled PID holds 120° on the feedforward, and the setpoint drops to 20° at 2 s. Use the default seed.

(a) Run as shipped and pause at about 9.9 s, before the setpoint returns at 10 s. Read the Throttle saturated, Since the last step and Overshoot tiles, and export the file to find the lowest `i_pct`. *Checkpoint:* 0.16 s, settled in 2.94 s, 5.6°; −8.13 %.

(b) Press Reset, set Anti-windup to + zone, and repeat (a). *Checkpoint:* 0.15 s, settled in 0.96 s, 0.9°; −1.81 %.

Explain why back-calculation leaves the integral so large when the throttle is pinned for so short a time.

3.9 App challenge · 8 min

Load the challenge scenario, Challenge, the setpoint program, and press Reveal suggested tuning. Use the default seed and run each case for the full 72 s at speed 5. After each run read the challenge result in the Learn panel: the slowest segment, the overshoot, the share of the run saturated and the chatter.

(a) Run with the suggested settings. *Checkpoint:* 2.26 s, 9.2°, 0.43 %, 1.19 %; the target is met.

(b) Press Reset, set Derivative filter to 0 ms, and run again. *Checkpoint:* 2.27 s, 8.1°, 0.36 %, 2.32 %; not met.

(c) Press Reset, set Derivative filter back to 20 ms and Anti-windup to Back-calc., and run again. *Checkpoint:* 3.79 s, 18.2°, 0.42 %, 1.20 %; not met.

Name the limit each failing run breaks and the cause.

3.10 Application area · 4 min

Case 7 holds 100° on the feedforward of Equation (3.12), which reads the battery but knows nothing of a clip. Use $G = 0.1197\,\mathrm{N\,m}$, $L = 0.32\,\mathrm{m}$, $l_p = 0.30\,\mathrm{m}$, $T_{max} = 0.80\,\mathrm{N}$ and $u_0 = 8\,\%$.

(a) The battery sags from 7.4 to 6.8 V. Find the feedforward before and after the sag. *Checkpoint:* 72.47 % and 78.16 %.

(b) A 5 g clip goes on the tip of a fresh battery's arm. Find the added weight torque at horizontal, the throttle that now holds 100°, and the shortfall of the feedforward. *Checkpoint:* 0.0157 N m, 76.57 %, 4.10 %.

Say which part of the controller must supply the shortfall, and name one measurement or estimate a production design could add.

3.11 Application area · 4 min

The kit's 8-bit throttle moves in counts of 0.392 %. Use the static map of Equation (3.5) with $G = 0.1197\,\mathrm{N\,m}$, $l_p = 0.30\,\mathrm{m}$, $T_f = 0.80\,\mathrm{N}$ and $u_0 = 8\,\%$, and invert it for the angle, $\sin\theta^* = T_f\, l_p\, f^2/G$ with $f = (u - u_0)/(100 - u_0)$.

(a) Find the angles that counts 185 and 186 of 255 hold, and say what count 187 does. *Checkpoint:* 80.8° and 87.7°; count 187 holds no angle.

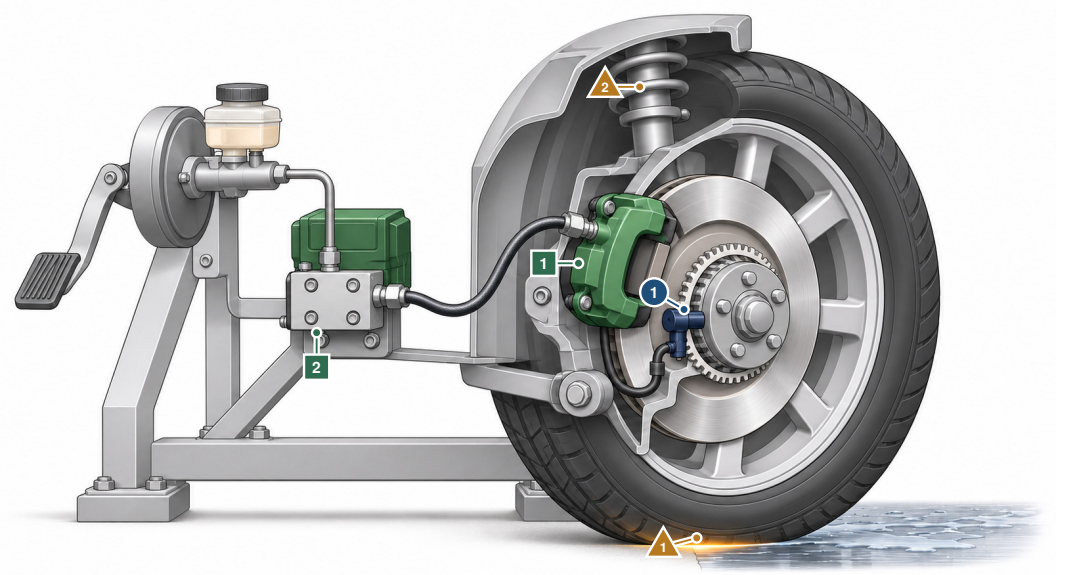
(b) A 10-bit throttle has 1023 counts. Find the last count that holds an angle and the angle it holds, and the angle the count below it holds. *Checkpoint:* Count 746 holds 87.0°, count 745 holds 84.6°.

3.12 Review the agent · 5 min

An AI assistant is asked to extend the PID tuned at 30° to 120° and replies as follows. “The throttle’s grip K_u is 2.61 mN m per percent at 30° and 3.43 mN m per percent at 120°, so the loop gain rises by 32 %. Scale all three gains by $K_u(30^\circ)/K_u(\theta^*)$, which is 0.76 at 120°. The loop gain is then the same at every angle, and the arm will hold 120° as it holds 30°.” Find the flaw and say how you would check it.

CHAPTER 4

Anti-Lock Braking at the Friction Peak Where the Process Gain Changes Sign



The controller moves 1 the brake, squeezing the disc on the wheel
2 the hydraulic modulator, which sets the brake pressure the controller asks for

The controller reads 1 the wheel-speed sensor at the toothed ring, how fast the wheel turns

What disturbs it 1 a change in the road surface under the tire 2 a change in the load the spring puts on the wheel

Notice that only the wheel's speed is read, while the road under the tire, which sets how much braking the wheel can take, is never measured. Illustration, not to scale.

IN PLAIN TERMS Braking Hard Without Locking the Wheel

The main idea. A braked tire grips the road best while it turns a little slower than the car is moving. An anti-lock system keeps each wheel turning just that much slower, at the point where the grip is greatest.

The challenge. Past that point, more brake gives less grip, and the wheel runs away into a skid within a fraction of a second. The best point moves with the road, and the system cannot tell dry pavement from ice before the tire does.

Why it matters. Many machines work best at the top of a curve that falls away beyond it.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history and the fact on heavy trucks in Section 4.2, the two try-its and the fact on the ideal distance in Section 4.6, 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 4.6), Technology in Practice (Section 4.7), Limitations (Section 4.8) and the open challenge

The control question A braked tire delivers its largest force at an intermediate slip, and past that slip more brake torque gives less force, so the process gain changes sign at the setpoint an anti-lock loop wants. How well can a fixed-gain PI loop hold slip at that peak when the road surface, and with it the peak, changes without warning?

Learning objectives

- Derive the hold torque and the linearized slip pole from the two momentum balances, and locate the stability boundary of a constant pedal.
- Compute the phase margin of the slip loop at a given speed and explain why it falls as the car slows.
- Explain why a sudden loss of friction is a windup problem for the PI loop although its output is far from the upper limit.
- Compare detuning the gain with lowering the slip setpoint as remedies for a slow brake actuator, and justify the choice from the phase margin.
- Judge a sliding-mode law against the PI loop from its boundary-layer gain and from measured runs.

Prerequisites Sections 1.4.2 (the fold of the holding input), 1.4.3 (sign from the immediate response), 1.4.7 (a lag ahead of an unstable pole), 1.4.9 (relay and describing function) and 1.6.2 (velocity form and back-calculation). Chapter 2 meets the same fold on an exothermic reactor.

The app https://apmonitor.com/apps/abs-braking/

See also Volume 4, Chapter 2 for a protective loop held off the peak of a compressor characteristic, and Volume 4, Chapter 12 for a slip loop on the same Burckhardt tire. Volume 20, Chapter 9 for speed limited by stopping distance and wheel slip, and Volume 20, Chapter 2 for a vehicle loop whose behavior moves with speed.

4.1 Introduction

Almost every loop in an introductory control course regulates a variable whose plant responds monotonically to the actuator. Opening a valve raises a level and adding heat raises a temperature, and the design questions concern the magnitude and the speed of the response [1]. A braked wheel is different. Pressing the brake pedal harder increases the retarding force up to a point, and past that point pressing harder reduces it, because the friction available at the contact patch is a function of how fast the tire is sliding over the road and that function has a maximum. The operating point that a braking controller wants is that maximum, which means the setpoint is deliberately placed at the one slip value where the derivative of the plant characteristic is zero.

The consequence dominates the control design. Below the maximum the tire is part of the negative feedback, because extra slip raises the friction coefficient, raises the force opposing the brake at the rim, and limits the growth of slip. Above the maximum the same mechanism has changed sign, so a disturbance that pushes slip up keeps pushing until the wheel stops turning, and a stopped wheel has no lateral capacity left either, so the vehicle cannot be steered while it slides. The friction peak is therefore not a flat optimum but a stability boundary that has to be held from the stable side inside a transient lasting two or three seconds [2]. A centrifugal compressor meets a boundary of the same kind at the peak of its pressure-rise characteristic, as Volume 4, Chapter 2 shows.

Two further properties complicate the design. The position of the maximum and the friction available there both depend on the road surface, which the controller cannot measure, so one fixed setpoint has to serve dry asphalt, standing water, snow and ice. The linearized plant gain also grows without bound as the vehicle slows, so a fixed-gain loop tuned with adequate margin at the start of a stop operates near its stability boundary at the end of one.

The physical content of the model is two conservation statements, a momentum balance on the translating quarter car and an angular momentum balance on the wheel, coupled through a single friction force that is the only driving force in the problem [3]. The slip dynamics are one to two orders of magnitude faster than the speed dynamics, and this time-scale separation is what permits a frozen-speed linearization of the loop [4]. The model brings together concepts a process dynamics and control course already carries, including a process gain that changes sign inside the operating

range, an open-loop unstable operating point, integral windup against an actuator limit, actuator lag as a phase budget, and a measurement whose noise grows as the signal shrinks [1, 2].

Textbook treatments of anti-lock braking usually present the friction curve as a static figure and then move directly to a sliding-mode or rule-based controller [5]. However, a static figure does not show the closed-loop consequence of operating on the falling branch, and it does not expose the interaction between the integral term, the actuator limit and a surface change that arrives without warning. This chapter specifies a browser simulation of quarter-car braking with a Burckhardt tire and measures what a fixed-gain slip loop can and cannot do when its setpoint sits on a maximum whose location is unmeasured.

Section 4.2 traces the problem from aircraft antiskid systems to electronic stability control and sets out the tire models and control approaches in use. Section 4.3 gives the four-state simulator, its equations, coefficients, limits, instrument chain and sample periods, at the level of detail a reimplementation needs. Section 4.4 derives the linearized slip dynamics including the vehicle-deceleration term, shows that the constant-torque equilibrium loses stability at the maximum of the hold-torque curve rather than at the maximum of the friction curve, and establishes that this stability boundary is independent of vehicle speed while the loop crossover is not. Section 4.5 describes the three automatic algorithms: the PI slip loop with its low-speed exit, a rule-based pressure cycle and a sliding-mode slip law. Section 4.6 runs thirteen configurations in which one property of the plant or the loop is changed at a time, reads each outcome from the application's own scorecard, and in Section 4.6.6 sets the sliding-mode law against the fixed-gain loop on an ice patch, with counterfactual runs that remove each part of the law in turn. The tire model and all three control laws are standard ones. Section 4.7 describes the sensors, valves and control units of real antilock systems, Section 4.8 states the limitations and Section 4.9 summarizes the measured results.

4.2 Background

Locked-wheel skidding was recognized as a control problem on aircraft long before it was addressed on cars, because an aircraft touching down carries far more kinetic energy per tire than a car and loses directional control completely when a main wheel stops turning. Aircraft antiskid systems modulate brake pressure against a measured wheel deceleration or a slip estimate, and the design of those systems is still organized around a tire-runway friction characteristic with an interior maximum [6]. Rail vehicles adopted the same idea for wheel-slide protection, where a slid wheelset both lengthens the stop and produces flat spots on the tread.

> ⧖ HISTORY **From Wheel Deceleration to Slip on Airliners, 1929 to 1976**
>
> Anti-lock braking was used on aircraft as early as 1929 [2]. A NASA analysis published in 1976 traced how airliner systems grew. In the earlier ones a speed sensor on each braked wheel watched for a deceleration past a preset threshold, and an on-off valve then cut the brake pressure to zero and reapplied it in full. Later systems modulated the pressure by the depth of the previous skid. The latest regulated slip velocity or slip ratio, which needs the airplane's ground speed as well as the wheel speed. Investigators had found flat spots from locked wheels on tires whose brakes and skid-control units all worked within tolerance, which suggested that wrong information was reaching the logic on slippery runways. Normal antiskid action was assured with high wheel spin-up accelerations or a continuous, accurate ground speed reference [7].

Road vehicles adopted the technology when the sensing and the actuation became inexpensive enough. Early automotive systems used analog electronics and mechanical modulators and were confined to expensive cars and to the rear axle. The step that made anti-lock braking a mass-market item was the replacement of analog logic by a digital control unit, which allowed the pressure-modulation rules to be revised in software rather than in hardware and made per-wheel regulation affordable [8]. Electronic brake control then became the platform for traction control and for electronic

stability control, which reuses the same wheel-speed sensors and the same hydraulic modulator to apply an individual brake for yaw regulation.

The safety record has been measured directly. Fleet studies of fatal crashes before and after anti-lock brakes were fitted found the effect mixed, with reductions in some crash types offset by increases in others, and concluded that the benefit is not a simple reduction in stopping distance [9]. Electronic stability control, built on the same actuator, has a much clearer record, with reviews of the empirical evidence reporting large reductions in single-vehicle and rollover crashes [10]. The distinction matters here, because on a low-friction surface anti-lock braking yields little reduction in stopping distance and instead preserves steering.

> **+ DID YOU KNOW? Antilock by Rule on Heavy Trucks**
>
> In the United States, Federal Motor Vehicle Safety Standard No. 121 requires an antilock brake system on each air-braked truck tractor manufactured on or after March 1, 1997, and on each single-unit vehicle from March 1, 1998. It must directly control the wheels of at least one front axle and at least one rear axle [11].

The modeling side developed alongside the hardware. The Magic Formula is a curve-fitting form that reproduces measured longitudinal and lateral tire characteristics from a small number of coefficients and is the standard representation inside vehicle-dynamics simulation [12]. The Burckhardt form used here is simpler, with three coefficients per surface, and is the representation normally taught alongside quarter-car braking models [3]. Both are static maps from slip to friction coefficient. Dynamic alternatives treat the contact patch as a distributed bristle with its own state, which is the LuGre construction originally proposed for friction in mechanisms [13] and later adapted to the tire-road interface, where the lumped and distributed forms are shown to reproduce the static curve at steady sliding while adding a relaxation length during transients [14].

Control of the resulting plant has been approached from several directions. Sliding-mode designs exploit the fact that the controller does not need the friction curve if it only needs to drive a switching surface to zero [15, 16], and a sliding-mode observer can search for the extremum of the friction characteristic directly rather than regulating to an assumed slip value [17]. Gain-scheduled linear designs address the same nonlinearity by scheduling on the operating point and by formulating the wheel-slip loop against a measured or estimated surface [18]. Extremum-seeking control removes the fixed setpoint entirely and dithers the brake pressure to climb the characteristic, which additionally improves the lateral force available during a stop [19]. Regulating a different variable is also an option, and blending wheel slip with vehicle deceleration keeps the loop from being left holding a signal that becomes meaningless at low speed [5]. All of these require some knowledge of where the peak is, and estimating the tire-road friction on line from the slip and the driving force is a recognized identification problem in its own right [20].

4.3 Vehicle Model

One corner of a passenger car is simulated during a straight-line stop. The vertical dynamics of such a corner, a body and a wheel on a spring and a strut, are the plant of Chapter 11. Here the corner moves only along the road. The controlled variable is the brake slip reconstructed from the wheel speed and the vehicle speed, and the manipulated variable is the brake torque commanded to the hydraulic modulator. Road surface, wheel-speed sensor noise and normal-load change remain operator inputs in both modes, and the controller receives none of them, which makes all three unmeasured loads. Figure 4.1 places every symbol in the equations at the point on the vehicle where it applies.

4.3.1 Variables and Signal Classification

Table 4.1 lists what each signal is and where it may be moved from. The single controlled variable is dimensionless, the single manipulated variable is a torque, and the vehicle speed is a state that the controller is allowed to read without error.

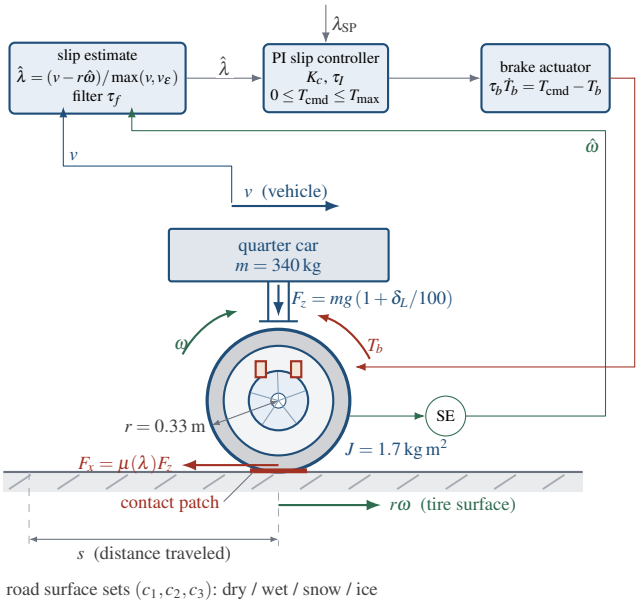


Figure 4.1 Quarter-car braking model with the symbols used in the state equations. Only the wheel speed is instrumented, and the vehicle speed v reaches the slip estimator without error.

Table 4.1 Roles and Ranges of the Signals

Symbol	Role	Unit	Range or default
λ	CV, brake slip	–	$[-1, 1]$
λ_{SP}	setpoint	–	0.02 to 0.60, def. 0.170
T_b	MV, applied torque	N m	0 to 2200
T_{cmd}	controller output	N m	0 to T_{max}
v	state, vehicle speed	m/s	$v_0 = 27$ at reset
ω	state, wheel speed	rad/s	v_0/r at reset
s	state, distance	m	0 at reset
surface	DV, road	–	dry / wet / snow / ice
σ_ω	DV, sensor noise	rad/s	0 to 4, def. 0
δ_L	DV, normal-load change	%	-20 to $+40$, def. 0
$\hat{\omega}$	measurement	rad/s	≥ 0
$\hat{\lambda}$	filtered CV estimate	–	what the loop sees
μ	output, friction coef.	–	0 to 1.17
F_x	output, tire force	N	μF_z

4.3.2 Translation, Rotation and the Slip Definition

Four states are integrated. The vehicle carries a quarter of the mass and is retarded only by the longitudinal tire force, the wheel is accelerated by that force at the rim and retarded by the brake, the brake torque follows its command through a first-order lag, and distance is the integral of speed, as shown in Equations (4.1) to (4.4).

$$m\dot{v} = -F_x \tag{4.1}$$
$$J\dot{\omega} = rF_x - T_b \tag{4.2}$$
$$\tau_b \dot{T}_b = T_{\text{cmd}} - T_b \tag{4.3}$$
$$\dot{s} = v \tag{4.4}$$

where m is the quarter-car mass, v the vehicle speed, F_x the longitudinal tire force, J the wheel rotational inertia, ω the wheel angular speed, r the loaded wheel radius, T_b the applied brake torque, τ_b the hydraulic time constant, T_{cmd} the commanded torque and s the distance traveled. Equation (4.3) runs whether or not the vehicle is moving, because brake pressure continues to build after the car has stopped. The remaining three derivatives are set to zero once $v \le 0$.

The tire force is the product of a friction coefficient and the normal load on the corner, as shown in Equation (4.5).

$$F_x = \mu(\lambda) F_z, \qquad F_z = mg\left(1 + \tfrac{\delta_L}{100}\right) \tag{4.5}$$

where δ_L is the normal-load change in percent and g is gravitational acceleration. With $\delta_L = 0$ the normal load is the static mg, so the deceleration in Equation (4.1) is exactly μg and the ideal stopping distance from speed v is $v^2/(2\mu_{\text{peak}} g)$.

Brake slip is the fraction of the contact-patch velocity that is sliding rather than rolling, and it is regularized in the denominator and clamped, as shown in Equation (4.6).

$$\lambda = \text{clamp}\left(\frac{v - r\omega}{\max(v, v_\varepsilon)}, -1, 1\right) \tag{4.6}$$

where $v_\varepsilon = 0.5$ m/s is the regularization speed. A free-rolling wheel has $r\omega = v$ and $\lambda = 0$, and a locked wheel has $\omega = 0$ and $\lambda = 1$. The exact ratio is singular at the end of a stop, because the numerator and the denominator both approach zero, and the $\max(v, v_\varepsilon)$ guard bounds both the signal and its derivative at the cost of understating the true slip below v_ε.

4.3.3 Burckhardt Tire Model and Surface Coefficients

The friction coefficient is the three-coefficient Burckhardt form, made odd in λ so that a wheel briefly turning faster than the vehicle is pulled back toward free rolling, as shown in Equation (4.7).

$$\mu(\lambda) = \text{sgn}(\lambda)\left[c_1\left(1 - \mathrm{e}^{-c_2|\lambda|_1}\right) - c_3|\lambda|_1\right] \tag{4.7}$$
$$\mu'(\lambda) = c_1 c_2 \mathrm{e}^{-c_2|\lambda|_1} - c_3 \tag{4.8}$$
$$\lambda^* = \min\left(1,\ \tfrac{1}{c_2}\ln\tfrac{c_1 c_2}{c_3}\right), \quad c_3 > 0,\ c_1 c_2 > c_3 \tag{4.9}$$

where $|\lambda|_1 = \min(|\lambda|, 1)$ clips the slip magnitude so the correlation is never extrapolated past locked wheel, c_1 sets the asymptotic level, c_2 the sharpness of the rise out of the origin and c_3 the slope of the falling branch. Every surface gives $\mu(0) = 0$, so a free-rolling wheel transmits no longitudinal force at all. Table 4.2 reports the coefficients that ship with the application together with the quantities they imply, computed from Equations (4.7) to (4.9) rather than quoted from a textbook, and Figure 4.2 draws the four curves. The same curve, with the same dry-asphalt coefficients, drives the wheel of Volume 4, Chapter 12, where wheelspin rather than braking carries the slip past the peak.

Three properties of Table 4.2 shape the control problem. The peak moves by more than a factor of eleven across the four surfaces, from $\lambda^* = 0.1700$ on dry asphalt to 0.0150 on ice, while the friction

Table 4.2 Shipped Burckhardt Coefficients and Derived Peak Quantities

Surface	c_1	c_2	c_3	λ^*	μ^*	$\mu(1)$	$\mu(0.170)$
Dry asphalt	1.2801	23.990	0.5200	0.1700	1.1700	0.7601	1.1700
Wet asphalt	0.8570	33.822	0.3470	0.1308	0.8013	0.5100	0.7953
Snow	0.1946	94.129	0.0646	0.0600	0.1900	0.1300	0.1836
Ice	0.0500	306.39	0	0.0150[a]	0.0500[a]	0.0500	0.0500

[a] Ice has no interior maximum. λ^* is the 99% point of the asymptote and μ^* is the asymptote c_1.

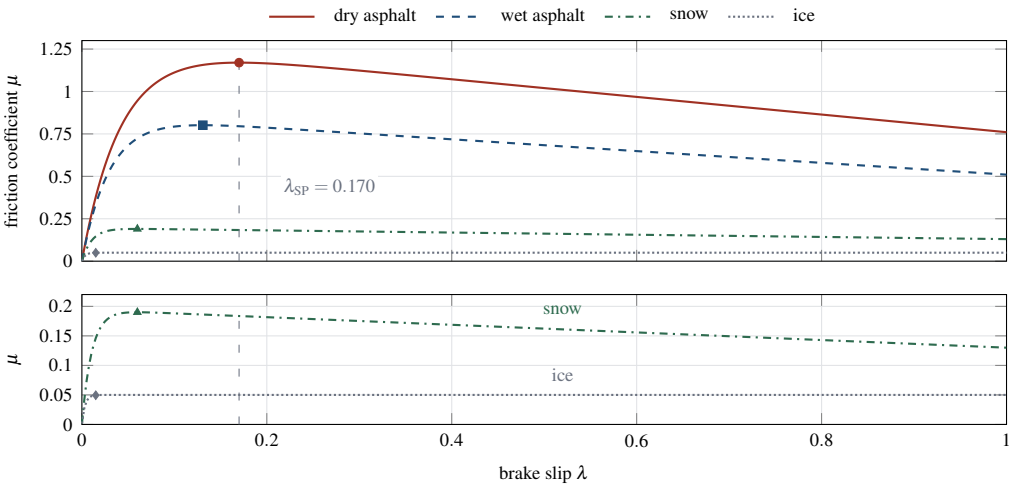


Figure 4.2 Burckhardt friction curves for the four shipped surfaces: dry asphalt solid with a circle at its peak, wet asphalt dashed with a square, snow dash-dotted with a triangle and ice dotted with a diamond (the 99% point of the asymptote for ice). The default slip setpoint of 0.170, the vertical line, is the dry peak and sits on the falling branch of the other three curves. The lower panel repeats snow and ice on an axis that resolves them.

available at the peak moves by a factor of 23.4, from 1.1700 to 0.0500. A locked wheel on dry asphalt gives $\mu(1) = 0.7601$, which is 35.0% below the 1.1700 available at the peak, and the corresponding losses on wet asphalt and snow are 36.4% and 31.6%. Ice has $c_3 = 0$, so its curve never turns over and no interior maximum exists. In that case the application places the displayed peak slip at the 99% point of the asymptote, $\lambda^* = \ln(100)/c_2 = 0.0150$, where $\mu = 0.0495$, which is a convention rather than a maximum. The available friction μ^* is taken as the asymptote $c_1 = 0.0500$ itself, so the ideal stopping distance on ice is the true minimum and cannot be beaten by the closed loop.

4.3.4 Brake Actuator, Wheel-Speed Sensor and the Slip Estimate

The brake modulator is a first-order lag with $\tau_b = 0.020$ s on the commanded torque, with no separate slew limit, and the command is clamped to $[0, T_{\max}]$ with $T_{\max} = 2200$ N m by default. Only the wheel speed is instrumented. The sensor adds Gaussian noise passed through a first-order filter of correlation time 0.01 s and rescaled to a stationary standard deviation of σ_ω, and the sum is clamped at zero, as shown in Equation (4.10).

$$\hat{\omega} = \max(\omega + n(t), 0), \qquad \operatorname{std}[n] = \sigma_\omega \tag{4.10}$$

The vehicle reference speed is handed to the estimator exactly. Slip is then reconstructed with Equation (4.6) using $\hat{\omega}$ in place of ω, and a first-order filter of time constant τ_f smooths the result, as shown in Equation (4.11).

Table 4.3 Constants of the Plant and of the Loop

Symbol	Meaning	Default	Unit	Range
m	quarter-car mass	340	kg	220 to 520
r	wheel radius	0.33	m	0.26 to 0.42
J	wheel inertia	1.7	kg m^2	0.8 to 3.5
g	gravity (hidden)	9.81	m/s^2	9.7 to 9.9
v_0	initial speed	27	m/s	5 to 40
$T_{\max}$	brake torque limit	2200	N m	800 to 2200
τ_b	brake actuator lag	0.020	s	0.005 to 0.200
R^+	rule-based apply rate	12000	N m/s	500 to 30000
R^-	rule-based release rate	15000	N m/s	2000 to 60000
$\Delta\lambda$	rule-based deadband	0.080	–	0.005 to 0.150
v_ε	slip regularization	0.5	m/s	0.2 to 2.0
v_{exit}	slip-control exit	3.0	m/s	1.0 to 6.0
v_{stop}	stop threshold (hidden)	0.05	m/s	0.01 to 0.20
K_c	proportional gain	6000	N m	0 to 30000
τ_I	integral time	0.12	s	0.02 to 2
τ_D	derivative time (adv.)	0	s	0 to 0.020
τ_f	slip filter (adv.)	0.004	s	0 to 0.050
λ_{SP}	slip setpoint	0.170	–	0.02 to 0.60
ρ	boundary-layer scale (adv.)	1	–	0 to 3

$$\hat{\lambda}_k = \hat{\lambda}_{k-1}\, a + \lambda_{\text{raw},k}\,(1-a), \qquad a = \mathrm{e}^{-\Delta t_c/\tau_f} \tag{4.11}$$

where Δt_c is the controller sample period and $\tau_f = 0.004$ s by default, adjustable from 0 to 0.050 s. The exact zero-order-hold form in Equation (4.11) is used rather than an Euler difference, so the filter is stable for any τ_f. There is no transport delay and no sensor bias anywhere in the application. Inferring the vehicle reference speed from the wheel speeds is the hardest part of a production anti-lock system and it is absent here by design.

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

4.3.5 Parameters, Hard Events and Numerics

Table 4.3 lists every constant. All of them except g and v_{stop} are operator-adjustable inputs, and the two exceptions are fixed because they are not quantities a calibration engineer would adjust.

The initial condition is a free-rolling wheel with the brake released, namely $v = v_0$, $\omega = v_0/r$, $T_b = 0$ and $s = 0$, which gives $\lambda = 0$, $\mu = 0$ and $\dot{v} = 0$ exactly. Three hard events are handled in a post-step routine, outside the right-hand side, so the integrator never meets a discontinuity. The wheel angular speed is clamped at zero so the wheel cannot turn backward. The vehicle is declared stopped once $v \le v_{\text{stop}} = 0.05$ m/s, at which point both v and ω are set to zero and frozen. A wheel-lock flag is raised while $v > \max(1.0, v_\varepsilon)$ and $\lambda > 0.92$, and the time spent in that condition is accumulated and reported. The theoretical best distance $v^2/(2\mu^* g)$ is evaluated once, at the instant the applied torque first exceeds 5 N m, and then held. Recomputing it during a stop would make it a moving target, so on a surface transition it continues to quote the figure for the surface the car started on.

The engine of Section 1.6.1 advances the states with the classical fourth-order Runge-Kutta scheme at $\Delta t = 0.002$ s. The controller executes on the same 2 ms grid, which is a 500 Hz sample rate, and the historian logs at 50 Hz. The fastest genuine dynamics in the plant are the 20 ms brake lag and the runaway to lock, which takes of order 100 ms, so the 2 ms step resolves both. The noise generator is seeded with 20260824, so noisy runs replay identically. Time is in seconds throughout, and every logged signal exports at 50 Hz for offline analysis.

4.3.6 Assumptions

The model is one wheel and one quarter of the mass with a single translational degree of freedom. Pitch, roll, yaw, the other three wheels and the suspension are absent, so nothing in the application can show the loss of steering that locking a front wheel actually causes, and that loss has to be asserted in the text rather than demonstrated. The coupling between longitudinal and lateral tire force that causes it is the central constraint of Chapter 10.

The normal load is static unless the normal-load input is changed. A real 1 g stop shifts roughly 30% of the load onto the front axle, which is the reason production systems regulate each wheel separately, and the input is a manual stand-in for that transfer rather than a model of it.

The Burckhardt curve is a steady-state fit. A real tire reaches its force through a relaxation length of a few decimeters, which at 27 m/s is a further 10 to 20 ms of lag that this model does not carry, and that missing lag is part of what makes a production anti-lock system harder to tune than this one. A dynamic friction model would supply it [14].

The friction coefficient depends on slip alone, while speed, temperature, tread wear, water film thickness and vertical load all move the real curve, so the coefficients in Table 4.2 represent one measurement of one tire on one surface. Slip is regularized below $v_\varepsilon = 0.5$ m/s and slip control is abandoned below $v_{\text{exit}} = 3.0$ m/s, which are numerical necessities that together make the last half meter optimistic by about 1 cm of a 30 m stop.

The stop is a hard event with no brake creep, no driveline windup and no rollback on a slope. Aerodynamic drag and rolling resistance are omitted entirely. That is a small error on dry asphalt, where the tire supplies $\mu^* g = 11.5$ m/s^2, but on ice the tire supplies only 0.491 m/s^2 and the omitted resistances are of the same order, so the ice stopping distances in Section 4.6 should be read as upper bounds. It also means the coasting configuration coasts forever.

4.4 Stability Analysis of the Slip Dynamics

4.4.1 Linearized Slip Dynamics

Differentiating Equation (4.6) above the regularization speed and substituting Equations (4.1) and (4.2) gives the slip rate exactly, as shown in Equation (4.12).

$$\dot{\lambda} = \frac{r}{Jv}\left(T_b - r\,\mu(\lambda) F_z\right) - \frac{(1-\lambda)\,\mu(\lambda)\,F_z}{m\,v} \tag{4.12}$$

The first term is the torque imbalance at the wheel and the second is the contribution of the vehicle's own deceleration, which is small but not negligible and is the term a wheel-only linearization omits. Setting $\dot{\lambda} = 0$ and using $F_z = mg$ gives the brake torque that holds a constant slip, as shown in Equation (4.13).

$$T_{\text{hold}}(\lambda) = r\,\mu(\lambda) F_z + \frac{J(1-\lambda)\,\mu(\lambda)\,g}{r} \tag{4.13}$$

Equation (4.13) carries no speed at all, which is why a well-regulated stop shows an almost flat manipulated-variable trace. On dry asphalt at the peak the two terms are 1287.8 N m and 49.1 N m, giving 1336.9 N m in total, and the wheel term is therefore 3.7% of the total.

Linearizing Equation (4.12) about a constant-slip operating point, holding v fixed because the speed dynamics are two orders of magnitude slower, gives an input gain and a pole, as shown in Equations (4.14) and (4.15).

$$k = \frac{\partial \dot{\lambda}}{\partial T_b} = \frac{r}{Jv} \tag{4.14}$$

$$p = \frac{\partial \dot{\lambda}}{\partial \lambda} = -\mu'(\lambda)\left[\frac{r^2 F_z}{Jv} + \frac{(1-\lambda) g}{v}\right] + \frac{\mu(\lambda)\,g}{v} \tag{4.15}$$

where μ' is the tire slope from Equation (4.8). The sign of μ' sets the character of the plant. On dry asphalt at $\lambda = 0.05$ the slope is 8.73 and the pole is $-71.8\ \mathrm{s}^{-1}$, so slip is strongly self-regulating. At $\lambda = 0.12$ the slope has fallen to 1.21 and the pole to $-9.51\ \mathrm{s}^{-1}$. At $\lambda = 0.25$ the slope is -0.444 and the pole is $+4.05\ \mathrm{s}^{-1}$, an unstable mode with a 0.25 s time constant that takes the wheel to lock in a few tenths of a second. The plant is first-order and stable on one side of the setpoint and first-order and unstable on the other, and the same three parameters describe both.

4.4.2 Stability Boundary of the Constant-Torque Equilibrium

Multiplying Equation (4.15) by v removes the speed entirely, so the slip at which the pole crosses zero is a property of the tire, the wheel and the load and not of how fast the car is going. That crossing satisfies Equation (4.16).

$$\mu'(\lambda_c)\left[\frac{r^2 F_z}{J} + (1-\lambda_c)g\right] = \mu(\lambda_c)\, g \tag{4.16}$$

Differentiating Equation (4.13) and setting the result to zero produces the same condition, so the constant-torque equilibrium loses stability exactly where the hold-torque curve is maximized rather than where the friction curve is maximized. This is the picture of Section 1.4.2: the input that holds a steady state has a maximum, the equilibrium loses stability at that maximum, and a larger input has no equilibrium at all. On dry asphalt that is $\lambda_c = 0.1661$ against a friction peak at $\lambda^* = 0.1700$, with $T_{\mathrm{hold}}(\lambda_c) = 1337.0$ N m against 1336.9 N m at the friction peak. The corresponding values are 0.1280 and 917.3 N m on wet asphalt, 0.0587 and 218.2 N m on snow, and 0.0289 and 57.5 N m on ice, where an interior maximum of the hold torque exists even though the friction curve has none.

This boundary is directly checkable in manual mode. A constant pedal of 1337 N m on dry asphalt stops the car in 32.7 m with the wheel never locking, while 1338 N m locks the wheel for 0.27 s and 1340 N m locks it for 0.87 s and costs 1.4 m. The equilibrium does not degrade as the torque is raised past 1337.0 N m but ceases to exist.

4.4.3 Speed Dependence of the Loop Transfer Function

Both the input gain in Equation (4.14) and the pole in Equation (4.15) are inversely proportional to v, so as the car slows the plant keeps its shape and compresses its time scale. What does not compress is the brake actuator, whose 20 ms lag is fixed by the hydraulics. A lag in series with an unstable pole limits the loop in the way Section 1.4.7 shows for the arm. The loop transfer function at the setpoint is given by Equation (4.17).

$$L(s) = K_c\,\frac{\tau_I s + 1}{\tau_I s}\cdot\frac{k}{(s-p)(\tau_b s + 1)} \tag{4.17}$$

Crossover therefore rises roughly as $v^{-1/2}$ while the actuator corner stays where it is, and the phase margin falls monotonically through the stop. Table 4.4 evaluates Equation (4.17) on dry asphalt at the default tuning and at two of the scenario tunings.

Three readings follow from Table 4.4. The default tuning gives a comfortable 40.6° at the start of a stop and 15.2° at the speed where slip control is handed over, which quantifies the need for a low-speed exit. The high-gain tuning is already at 8.4° at 27 m/s and reaches 3.6° at 5 m/s, so the cycling it produces gets faster and less damped as the car slows. The slow-actuator case is not marginal but unstable at the setpoint, at -1.3°, and moving the setpoint down to $\lambda_{\mathrm{SP}} = 0.12$ restores $+28.1^\circ$ by moving the plant pole from $+0.42$ to $-9.51\ \mathrm{s}^{-1}$. Where the actuator prevents an adequate margin at the peak, lowering the setpoint restores the margin.

Table 4.4 Linearized Loop at $\lambda = \lambda_{SP}$ on Dry Asphalt

Condition	k [(N m s)$^{-1}$]	p [s^{-1}]	ω_c [rad/s]	PM [deg]
$v = 27$, default	0.00719	+0.42	36.0	+40.6
$v = 20$, default	0.00971	+0.57	44.3	+37.1
$v = 10$, default	0.01941	+1.15	68.9	+28.1
$v = 5$, default	0.03882	+2.29	102.5	+20.1
$v = 3$, default	0.06471	+3.82	135.0	+15.2
$v = 27$, K_c=25000, τ_I=0.03	0.00719	+0.42	91.6	+8.4
$v = 5$, K_c=25000, τ_I=0.03	0.03882	+2.29	218.8	+3.6
$v = 27$, $\tau_b = 0.12$	0.00719	+0.42	19.0	−1.3
$v = 27$, τ_b=0.12, λ_{SP}=0.12	0.00719	−9.51	17.8	+28.1

4.5 Controller Design

4.5.1 Discrete Form, Direction and Anti-Windup

The primary controller is the velocity-form PID of Section 1.6.2, with the derivative taken on the measurement, executed every $\Delta t_c = 0.002$ s. Written in increments for this loop, it is given by Equations (4.18) to (4.22).

$$e_k = a\left(\lambda_{SP} - \hat{\lambda}_k\right), \qquad a = +1 \tag{4.18}$$

$$\Delta u_k^P = K_c\left(e_k - e_{k-1}\right) \tag{4.19}$$

$$\Delta u_k^I = \frac{K_c}{\tau_I}\, e_k\, \Delta t_c \tag{4.20}$$

$$\Delta u_k^D = -K_c \tau_D\, a\, \frac{\hat{\lambda}_k^f - 2\hat{\lambda}_{k-1}^f + \hat{\lambda}_{k-2}^f}{\Delta t_c} \tag{4.21}$$

$$\begin{aligned} s_k &= s_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= s_k + I_{k-1} \\ T_{\mathrm{cmd},k} &= \mathrm{clamp}(c_k, 0, T_{\max}) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{\tau_t}\left(T_{\mathrm{cmd},k} - c_k\right) \end{aligned} \tag{4.22}$$

where a is the controller direction, I is the integral state, $\hat{\lambda}^f$ is the derivative-path estimate filtered with time constant τ_f, and $\tau_t = \tau_I$ is the tracking time of the back-calculation. These are Equations (1.12) to (1.15) with s_k for the running sum S_k, c_k for the demanded torque v_k, $T_{\mathrm{cmd},k}$ for the clipped output u_k and Δt_c for Δt. The bias term is zero, so the loop starts from no brake pressure.

The controller direction follows from the sign of the process gain. More brake torque slows the wheel relative to the vehicle, which raises slip, so the process gain from T_b to λ is positive, a slip below setpoint produces a positive error and calls for more torque, and $a = +1$. The sign is read from the immediate response, the input gain k of Equation (4.14), which is positive on both branches, and not from the steady-state gain, which reverses at the maximum of the hold-torque curve. Section 1.4.3 makes the same distinction on the pendulum. A shipped numerical test checks that the output is larger below setpoint than on setpoint and smaller above it, so an edit that inverts the loop fails the test.

Anti-windup is the back-calculation term of Equation (4.22). While the command is clipped, $(\Delta t_c/\tau_t)(T_{\mathrm{cmd},k} - c_k)$ drives I toward the value that reproduces the applied torque with a tracking time of τ_I, so the integral alone is corrected and the running sum s_k is never truncated [21]. At the default $\tau_I = 0.12$ s each scan removes $1/60$ of the clipped excess. The tracking time affects the reported results. A recursion that absorbed the whole excess in one scan would release the brake command 0.12 s sooner after the ice patch of Section 4.6. Anti-windup matters more here than in most loops of

this size, because the danger is not an integral that has grown too large against a limit but an integral that is suddenly wrong. When the friction available collapses mid-stop, most of the commanded torque is stored in I, and the proportional term can react within one sample while the integral can only unwind at a rate of order K_c/τ_I. The mechanism is integral windup even though the output is nowhere near its upper limit.

Manual mode and the transfer between modes work as Section 1.6.2 describes, so the pedal and the loop hand over without a bump in both directions. The idle algorithms always track the applied torque, so switching among the three automatic algorithms is bumpless as well.

4.5.2 The Low-Speed Exit

Below $v_{\text{exit}} = 3.0$ m/s the controller stops regulating slip. On the first sample below that speed it latches the torque that holds the present slip against the road force read from the car's own deceleration, as shown in Equation (4.23).

$$T_{\text{exit}} = \operatorname{clamp}\left(m\hat{a}\left(r + \frac{J(1-\hat{\lambda})}{mr} \right), 0, T_{\max} \right) \tag{4.23}$$

where $\hat{a}$ is the fall rate of the vehicle-speed signal filtered with a 0.01 s time constant. The value is commanded for the rest of the stop while the PID continues to track it. Whenever $\hat{a}$ falls below 0.049 m/s^2, a tenth of the peak deceleration on ice, the held torque rises at the modulator's 12000 N m/s, and once that ramp has acted a measured slip above 0.8 with the car above 1 m/s resets the hold to the value of Equation (4.23). The controllers are given the vehicle's mass, wheel radius and inertia, as a production unit is coded with the car and its tire size, and never the road surface or the load. An earlier version held $\max(T_{\text{cmd}}, T_{\text{hold}}(\lambda_{\text{SP}}))$ with T_{hold} from Equation (4.13) on the true surface and load, which a controller cannot know, and the larger of the two also froze a command caught mid-swing. The change moves every run that reaches the exit, and the results below are measured with Equation (4.23). Across 1014 runs (thirteen configurations, 26 settings each and three algorithms, 40 s apiece) every automatic stop either finishes or is still decelerating at what the road carries when the 40 s run ends, with one exception that predates the change, the rule-based cycle at $\lambda_{\text{SP}} = 0.02$ on the wet-to-dry configuration, which coasts at 3.12 m/s with the brake released. Table 4.4 gives the engineering reason for the handover, because the phase margin at 3 m/s is 15.2° and falling, and Section 4.6 gives the measurement reason, because the slip signal itself becomes uninformative at the same time. A production system does the same below roughly 5 km/h.

4.5.3 The Rule-Based Alternative

The second automatic algorithm is the classical increase, hold and decrease pressure cycle, requiring no model and no gains. Its logic is given by Equation (4.24).

$$u_k = \begin{cases} u_{k-1} - R^{-}\Delta t_c, & e_k < 0 \text{ and } \dot{\hat{\lambda}}_k > 0 \\ u_{k-1} + R^{+}\Delta t_c, & e_k > \min(\Delta\lambda, \lambda_{\text{SP}}/2) \\ u_{k-1}, & \text{otherwise} \end{cases} \tag{4.24}$$

followed by a clamp to $[0, T_{\max}]$, where $\dot{\hat{\lambda}}$ is a filtered slip rate with a 0.01 s time constant. Three asymmetries are deliberate. The deadband $\Delta\lambda = 0.080$ sits entirely below the setpoint, so pressure comes off as soon as slip crosses λ_{SP} and goes back on only after slip has fallen a full deadband below it. The deadband is capped at half the setpoint, so a low target such as the 0.060 snow peak can still call for pressure. Release stops as soon as the wheel starts recovering rather than waiting for slip to return, which prevents the cycle from dumping far more pressure than it needs. Apply and release run at fixed rates rather than in proportion to error, exactly as a modulator valve does. All three exist because the plant is asymmetric, since operating slightly below the peak costs a few percent of force while operating past the peak drives the wheel toward lock.

4.5.4 Sliding-Mode Slip Control

The third automatic algorithm is a sliding-mode law, the classical design for a plant whose gain is uncertain but bounded [22]. It predicts nothing and solves no optimization. Equation (4.12) rearranges into the input-affine form of Equation (4.25), in which the brake torque enters with a positive gain that depends only on speed.

$$\dot{\lambda} = b(v)\Big(T_b - F_x\,\kappa(\lambda)\Big), \quad b = \frac{r}{Jv}, \quad \kappa = r + \frac{J(1-\lambda)}{mr} \tag{4.25}$$

The torque that holds any slip is therefore $F_x\kappa$, whatever the speed, which is Equation (4.13) written for an unknown force. The switching surface is the slip error itself, $\sigma = \hat{\lambda} - \lambda_{\mathrm{SP}}$, read from the same filtered estimate the PI loop uses, and the law is given by Equations (4.26) to (4.28).

$$T_{\mathrm{cmd}} = \mathrm{clamp}\Big(\hat{F}_x\,\kappa(\hat{\lambda}) - K\,\mathrm{sat}(\sigma/\phi),\, 0,\, T_{\max}\Big) \tag{4.26}$$

$$K = \bar{F}_x\,\kappa(\hat{\lambda}) + \eta/b(v) \tag{4.27}$$

$$\dot{\hat{F}}_x = -\beta^2\,\sigma/(b\,\kappa), \qquad 0 \le \hat{F}_x \le \bar{F}_x \tag{4.28}$$

where $\hat{F}_x\kappa$ is the equivalent control (the torque that would hold the slip if the friction force were $\hat{F}_x$), sat is the unit saturation, ϕ is the boundary-layer width, $\eta = 1\ \mathrm{s}^{-1}$ is the reaching margin and β is the closed-loop pole defined below. The friction bound $\bar{F}_x = \bar{\mu}\,mg\,(1+0.40) = 5463$ N is the only use of the tire table, with $\bar{\mu} = 1.170$ the largest coefficient any shipped surface reaches (the dry peak) and 1.40 the top of the normal-load input. The estimate $\hat{F}_x$ is an integral of the slip error, projected onto $[0, \bar{F}_x]$, frozen while the command sits on a limit in the direction the error pushes, and updated after the output with a forward Euler step of Δt_c. It is a force rather than a slip, so a constant road needs a constant estimate while the car slows. The law executes every 2 ms with the PI loop tracking its applied torque, it shares the low-speed exit, and a command that is not finite or not inside $[0, T_{\max}]$ hands the brake to the tracking PI loop until the next valid one. Its mean cost is below 1.5 microseconds per execution in Node on every case of Section 4.6.6.

Outside the layer, $|\sigma| \ge \phi$, the gain of Equation (4.27) covers any force and any estimate inside the bound, so the unclamped command gives $\sigma\dot{\sigma} \le -\eta|\sigma|$, the reaching condition [23]. A check over 8.96×10^6 states outside the layer (five vehicles, two torque limits, ten speeds, five setpoints, six layers, three filters, four surfaces, three loads, 51 slips and five estimates) finds no violation for the unclamped command. The clamp breaks it on one side. A fully released brake leaves only the tire to spin the wheel back, at $bF_x\kappa$, which exceeds η only if $F_x \ge \eta Jv/(r\kappa)$, or 412 N at 27 m/s, while ice carries at most 233 N. On ice above about 14 m/s, and on snow near the top of the speed range, the slip error therefore shrinks more slowly than the condition requires.

The brake lag and the sample delay are not in Equation (4.25), and a relay acting through them chatters [24]. The boundary layer replaces the relay by a straight line inside $|\sigma| < \phi$ [23]. With the balance condition $K/\phi = 2\beta/b$ the law inside the layer is a proportional gain $2\beta Jv/r$ on slip plus the integral of Equation (4.28), and the slip error obeys $(d/dt + \beta)^2\sigma = 0$ at every speed. The pole is placed at a third of the bandwidth of the lags between the command and the measured slip, $\beta = 1/(3(\tau_b + \tau_f)) = 13.9$ rad/s with the declared 20 ms brake and the 4 ms filter, and the width follows as $\phi = \rho\,bK/(2\beta)$ with ρ the boundary-layer scale of Table 4.3. At $\rho = 1$ the layer is 0.52 of slip at 27 m/s, 0.97 at 14 m/s and 4.4 at 3 m/s, the proportional gain is 3864, 2003 and 429 N m per unit slip, and the integral time is $2/\beta = 0.144$ s. The loop inside the layer, from the command to the measured slip, is given by Equation (4.29).

$$L_\sigma(s) = \frac{2\beta s + \beta^2}{s^2} \cdot \frac{\mathrm{e}^{-s\Delta t_c/2}}{(\tau_b s + 1)(\tau_f s + 1)} \tag{4.29}$$

Table 4.5 Shipped Configurations

#	Name	Mode	What is changed
1	Rolling wheel	MAN	dry, $T_b = 0$, then constant-pedal trials
2	Dry pavement	AUTO	reference stop, $v_0 = 27$
3	Wet pavement	AUTO	wet, $v_0 = 25$
4	Snow	AUTO	snow, $v_0 = 16$
5	Ice	AUTO	ice, $v_0 = 8$
6	Dry to ice	AUTO	surface to ice at $t = 1.2$ s
7	Wet to dry	AUTO	surface to dry at $t = 1.5$ s, $v_0 = 25$
8	Slow actuator	AUTO	$\tau_b = 0.12$ s
9	Excessive gain	AUTO	$K_c = 25000$, $\tau_I = 0.03$ s
10	Sensor noise	AUTO	$\sigma_\omega = 2.0$ rad/s
11	Manual maximum	MAN	$T_b = 2200$ N m held
12	Sliding mode	AUTO	configuration 6 under the sliding-mode law, $\rho = 1$
13	Challenge	AUTO	wet to ice at 1.5 s, $v_0 = 20$, $K_c = 1200$, $\tau_I = 0.8$ s

It carries no speed, because the law's gain falls as the plant gain $r/(Jv)$ rises. Evaluated exactly in discrete time (zero-order hold at 2 ms, the filter of Equation (4.11), the integral after the output) it crosses over at 25.4 rad/s with 41.8° of phase margin and 18.9 dB of gain margin. A brake lag of 0.12 s, that of configuration 8, leaves 0.6° and 1.7 dB, and the closed-loop spectral radius rises from 0.981 to 0.9998. At $\rho = 0$ the law is a relay between the clamps, 0 and 2200 N m, and a describing function (Section 1.4.9) with the half-sample delay predicts a limit cycle at 15.6 Hz.

Three models are kept apart. The plant is Equations (4.1) to (4.4) with the Burckhardt curve and the run's brake lag and load. The instrument is Equations (4.10) and (4.11). The law's model is Equation (4.25) with the vehicle's coded m, r and J, the declared 20 ms brake and g, and the friction bound. It never reads the road, the load, the true brake lag or the friction curve, so a brake-lag change is model error that the law has to absorb.

4.6 Results

The shipped configurations appear in Table 4.5, and the scorecard at the end of each run appears in Table 4.6, and Figure 4.3 draws three of its columns. The application lists them as numbered scenarios in the same order, with the challenge last and unnumbered, and credits each one by the rules of Section 1.6.4. Every value comes from the running application, with each configuration loaded as shipped, the one change a row names made before the run, and the run taken to the stop. The application gives the same run whether that change is made before the run or carried through Reset. Distance is measured from the instant the applied torque first exceeds 5 N m, IAE is $\int |\lambda_{\mathrm{SP}} - \lambda|\, dt$ over the whole run, MV travel is $\sum_k |T_{b,k} - T_{b,k-1}|$ over the 50 Hz log, and saturation counts only the time the brake sits at $T_{\max}$, because a released brake is resting rather than out of authority. Peak slip, event times and oscillation frequencies are read from the simulated trajectory at the 2 ms step, because the 50 Hz log undersamples the fastest chatter, whose frequency approaches the 157 rad/s Nyquist limit of the log.

4.6.1 Reference Stop and Locked-Wheel Comparison

Configuration 2 is the case every other run is compared against. The loop brings slip onto 0.170 and holds it, the manipulated variable settles at 1336.9 N m and stays there, and the car stops in 33.0 m against a theoretical 31.8 m. The 1.2 m gap is bought in the first third of a second, while the brake pressure is still building through its 20 ms lag and the friction coefficient is still climbing from zero. The friction coefficient reaches 99% of its peak 0.24 s after the brake is applied, and from 0.33 s to

Table 4.6 Measured Outcome of Each Configuration, Read From the Application Scorecard

Configuration	Distance [m]	Ideal [m]	Locked [s]	Peak λ	IAE	MV travel [N m]	Saturated [%]	Note
2 Dry, PI	33.0	31.8	0.00	0.170	0.025	730	0.0	reference stop
2 Dry, rule-based	36.7	31.8	0.00	0.552	0.256	17910	0.0	3.7 m longer, 25× the MV travel
3 Wet	40.5	39.8	0.00	0.170	0.022	448	0.0	λ_{SP} 0.039 past the wet peak
4 Snow	71.1	68.7	0.00	0.236	0.037	1144	0.0	λ_{SP} nearly 3× the snow peak
4 Snow, rule-based	77.7	68.7	1.42	1.000	3.221	38313	0.0	fixed rates do not transfer
5 Ice	65.2	65.2	0.00	0.370	0.100	955	0.0	ideal from the μ asymptote
6 Dry to ice	218.1	31.8	0.00	0.655	0.280	2267	0.0	ideal frozen at the dry figure
7 Wet to dry	36.9	39.8	0.00	0.170	0.025	1235	0.0	no excursion above the setpoint
8 Slow actuator	35.0	31.8	0.00	0.264	0.153	6368	0.0	unstable at the setpoint
9 Excessive gain	33.3	31.8	0.00	0.344	0.161	68023	0.0	93× the MV travel, 0.3 m worse
10 Sensor noise	34.0	31.8	0.00	0.906	0.205	27477	0.0	no lock at the exit
11 Manual maximum	47.9	31.8	3.30	1.000	2.854	809	92.7	14.9 m lost to a locked wheel
12 Sliding mode	239.3	31.8	0.00	0.849	0.488	2334	0.0	ice patch under the law
13 Challenge, as loaded	269.7	25.4	0.00	0.410	0.503	913	0.0	never reaches the wet peak
13 Challenge, retuned	94.3	25.4	0.00	0.629	0.197	1704	0.0	$K_c = 6000$, $\tau_I = 0.12$ s

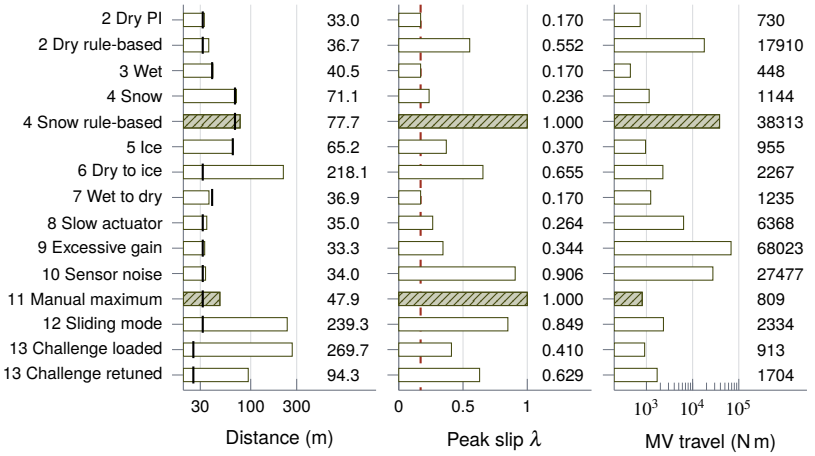


Figure 4.3 The fifteen runs of Table 4.6: stopping distance with the ideal as a black tick (left, logarithmic), peak slip with the 0.170 setpoint dashed (middle) and the travel of the brake torque (right, logarithmic). Hatched bars are the two runs that lock the wheel. Notice that configurations 2 and 9 stop within 0.3 m of each other while their torque travel differs by a factor of 93, and that both runs that lock the wheel carry a peak slip of 1.000.

the end of the stop the car is 0.55 m/s faster than an ideal stop, so no tuning removes the gap and tunings differ only in how quickly they stop paying it.

> **TRY IT IN THE APP** **2 · Dry Pavement · ABS Holding the Peak**
>
> Load the scenario and press Run. Watch the slip trace climb onto the 0.170 setpoint and stay there while the brake torque settles at 1336.9 N m and stays flat. The car stops in 33.0 m against a theoretical 31.8 m (Table 4.6).

Configuration 11 quantifies the loss that anti-lock braking exists to prevent. Holding the pedal at the 2200 N m limit locks the wheel 0.15 s after the brake is applied and keeps it locked for 3.30 s of a 3.57 s stop, at which point the tire is delivering $\mu(1) = 0.7601$ instead of the 1.1700 available at the peak. The stop takes 47.9 m against 33.0 m for the regulated case, so braking harder lengthens the stop by 14.9 m. Handing the sliding wheel to the PI loop at $t = 1.5$ s, at 15.5 m/s, unlocks it 0.07 s later, brings slip back within ± 0.03 of the setpoint by $t = 1.83$ s and finishes the stop in 42.8 m. In the all-manual stop the manipulated variable is saturated for 92.7% of the run while its total travel is 809 N m, barely more than the well-behaved reference stop. A loop can be quiet, saturated and badly wrong at the same time, and no single scorecard number detects the failure.

Pause. Try Exercise 4.2 before reading on.

4.6.2 Fixed Setpoint Across Four Surfaces

Configurations 3, 4 and 5 keep the setpoint at 0.170 and change the road. A fixed setpoint costs little braking force and a great deal of stability margin. On wet asphalt $\mu(0.170) = 0.7953$ against a peak of 0.8013, which is 99.2% of what is available, and the stop takes 40.5 m against an ideal 39.8 m. Aiming at the wet peak itself, $\lambda_{\mathrm{SP}} = 0.131$, gives 40.7 m, and aiming far past it at 0.30 still stops in 42.5 m without a lock. On snow the setpoint is nearly three times past the peak of 0.0600, yet $\mu(0.170) = 0.1836$ against 0.1900, which is 96.6%, and the stop is 71.1 m against an ideal 68.7 m, or 68.9 m with the setpoint moved to the snow peak. The falling branch is gentle on both surfaces because the $c_3\lambda$ term is small, which is why production systems can use a single target slip.

What the fixed setpoint costs is the stabilizing term. On wet asphalt at $\lambda = 0.170$ the tire slope is -0.255 and the open-loop pole is $+2.38\ \mathrm{s}^{-1}$ at 27 m/s, and on snow it is -0.0646 and $+0.60\ \mathrm{s}^{-1}$, so the loop runs on the runaway branch of both surfaces and only feedback keeps the wheel turning. Trading 0.8% of the available friction for the whole restoring term is why a calibration engineer aims below a peak rather than at it.

On ice the regulated stop takes 65.2 m against an ideal of 65.2 m, because the controller operates at $\lambda = 0.170$ on the flat asymptote, where $\mu = 0.0500$ is exactly the μ^* the ideal is computed from. A pedal held at 2200 N m keeps the wheel locked for 14.2 s of the 16.2 s stop and also stops in 65.2 m, so the whole benefit of anti-lock braking on ice is the lateral force that a rolling tire still has.

> **DID YOU KNOW?** **A Stop That Beats Its Own Ideal**
>
> From the simulator: configuration 7 stops in 36.9 m against an ideal of 39.8 m (Table 4.6). The ideal is computed once, for the wet asphalt the stop starts on, and the dry patch at 1.5 s then offers more grip than that figure assumes.

4.6.3 Surface Transition From Dry Asphalt to Ice

Configuration 6 is the central demonstration of the application, and it is reproduced in Figure 4.4. The loop has settled at 1336.9 N m holding $\lambda = 0.170$ on dry asphalt when the road becomes ice at $t = 1.2$ s, where the torque that holds the same slip is 57.1 N m. That is a factor of 23.4 applied in one step, and Equation (4.13) makes it a step in the plant and not in the setpoint.

Slip reaches 0.655 at $t = 1.27$ s, which is far out on what would be the runaway branch on any surface with $c_3 > 0$, and the wheel does not lock. The proportional term reacts within one sample but the stored integral has to be unwound, the command reaches zero at $t = 1.22$ s and stays clamped there until $t = 1.90$ s, and slip returns to within ± 0.03 of setpoint at $t = 1.86$ s, which is 0.66 s after

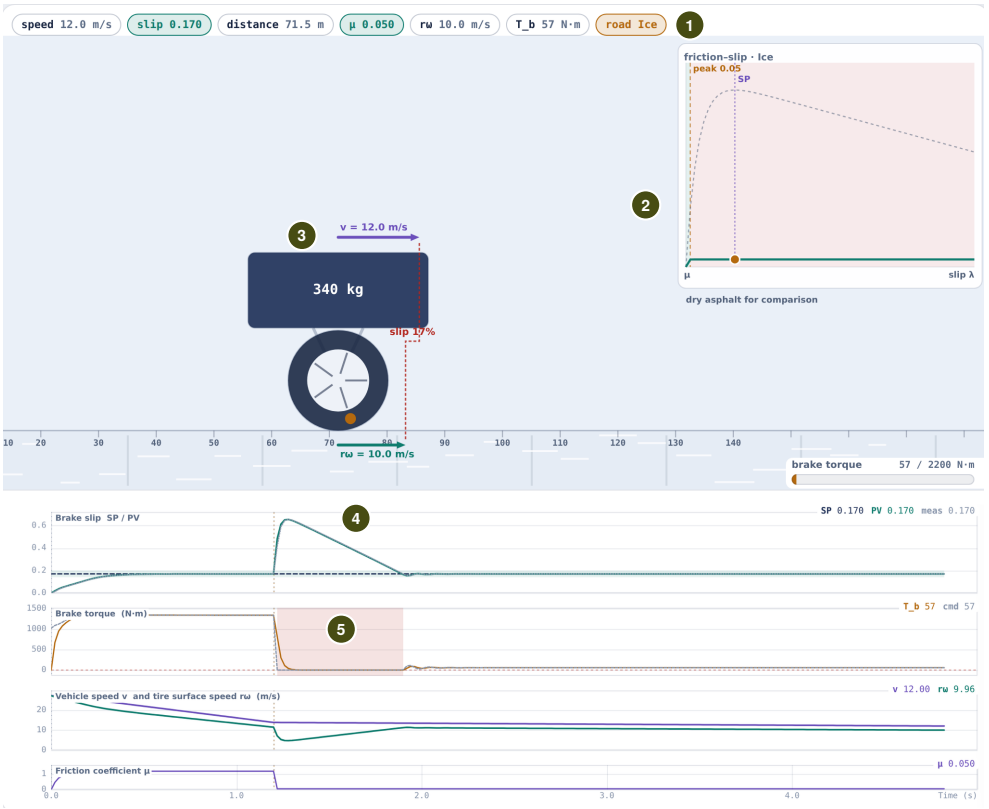


Figure 4.4 Dry asphalt turning to ice 1.2 s into a stop from 27 m/s, paused at $t = 4.83$ s, where the trends of the 5 s window end. The picture is a composite of the plant view (top) and the four trends (bottom) cut from one screenshot. (1) The badges: at the instant shown the car has covered 71.5 m and still travels at 12.0 m/s, with slip on its 0.170 setpoint and μ at 0.050 on ice. (2) The friction-slip view of ice, its operating point on the flat curve. (3) The vehicle speed above the car and the tire surface speed below the wheel. (4) Slip spikes from the setpoint to 0.655 (Table 4.6). (5) The brake command is held against its lower limit (shaded), from 1.22 s to 1.90 s, times the screen does not print, while the friction coefficient collapses from 1.170 to 0.050 (Table 4.2) and the vehicle and tire surface speeds separate and rejoin.

the event. Raising τ_I to 1 s slows the unwind without locking the wheel, because the slower integral has built only 1254.5 N m by the time the ice arrives. Slip then peaks at 0.498, is still 0.016 above setpoint at $t = 3.2$ s and does not return within ± 0.03 until $t = 2.57$ s. No integral time in the 0.02 to 2 s range locks the wheel on this transition, while a gain between 600 and 1500 N m does, for 0.21 s at $K_c = 1000$ N m. This is the clearest demonstration in the application that the sign of a disturbance matters more than its size for this plant, because configuration 7 applies the mirror-image change and produces no excursion above the setpoint, with a peak slip of 0.170 and an IAE of 0.025 that are indistinguishable from an undisturbed stop. Its only excursion is toward free rolling, a dip to 0.074 at $t = 1.53$ s that is back within ± 0.03 by $t = 1.57$ s.

Pause. Try Exercise 4.3 before reading on.

Reducing the torque limit makes the recovery better rather than worse, and the integral state explains why. With $T_{\max}$ lowered to 900 N m the loop spends the entire dry phase asking for more torque than the brake can deliver, back-calculation holds the integral near the value that reproduces the applied output instead of letting it grow, and there is therefore less stored torque to unwind when

the ice arrives. Peak slip after the transition falls from 0.655 to 0.430, and slip is back within ± 0.03 of setpoint 0.47 s after the event instead of 0.66 s. The dry phase costs distance, because the car meets the ice at 18.0 m/s instead of 13.8 m/s, and the whole stop grows from 218.1 m to 356.7 m, while the ice transition costs less.

> **TRY IT IN THE APP** **6 · Dry → Ice, One Second Into the Stop**
>
> Load the scenario and run it to 3.5 s: slip spikes to 0.655 and is back within ± 0.03 of the setpoint 0.66 s after the ice. Reload the scenario, set Brake torque limit to 900 N m under Parameters in Show Model, and run again. Slip now peaks at 0.430 and recovers in 0.47 s.

4.6.4 Actuator Lag, Excessive Gain and Sensor Noise

Configurations 8, 9 and 10 each attack the loop in a different way and produce three different failures. The slow actuator at $\tau_b = 0.12$ s adds phase without adding gain, and Table 4.4 puts the loop at $-1.3°$, which is unstable at the setpoint. The simulated trajectory agrees. Slip never settles and instead runs a sustained oscillation between 0.084 and 0.264 for the whole regulated part of the stop. Its period shortens from 0.32 s in the first cycle to 0.16 s in the eighth as the car slows, a frequency rising from 20 to 41 rad/s with a mean of 26.7 rad/s, against a predicted crossover of 19.0 rad/s at 27 m/s rising to 31.1 rad/s at 10 m/s. The wheel never locks. At the low-speed exit at $t = 2.24$ s the hold is the 1268 N m the road is carrying, not the 1710 N m the oscillation was demanding at that instant. The stop takes 35.0 m instead of 33.0 m.

Two repairs behave very differently. Halving the gain to $K_c = 3000$ lengthens the stop to 37.1 m and leaves the slip IAE at 0.153, because the oscillation is driven by pressure that was already commanded arriving late rather than by loop gain as such. Moving the setpoint to $\lambda_{SP} = 0.120$ gives 36.1 m and drops the slip IAE from 0.153 to 0.031, a factor of five, and Table 4.4 says why, because the plant pole moves from $+0.42$ to -9.51 s^{-1} and the margin returns to $+28.1°$. Lowering the setpoint restores the stability margin that the actuator lag consumed.

The excessive-gain case demonstrates that stopping distance alone does not discriminate between tunings. At $K_c = 25000$ and $\tau_I = 0.03$ s the car stops in 33.3 m against 33.0 m, a difference the driver would never feel, while the total variation of the brake torque rises from 730 to 68023 N m, a factor of 93, and the slip IAE gets six times worse at 0.161 against 0.025. Every one of those cycles is a real modulator valve opening and closing. The cycling is present from the start, with 22% of the travel spent in the first half second, and what changes through the stop is its frequency. Slip cycles at 75 rad/s over the first half second and at 123 rad/s over the last half second before the low-speed exit, rising with the predicted crossover of 91.6 rad/s at 27 m/s and 218.8 rad/s at 5 m/s while the phase margin falls from $8.4°$ to $3.6°$.

Sensor noise of $\sigma_\omega = 2.0$ rad/s is 0.66 m/s at the tire surface, which is 0.024 of slip at 27 m/s and 0.13 of slip at 5 m/s, because Equation (4.6) divides by the speed. The measurement therefore degrades in exactly the same proportion as the plant gain grows, and the actuator travel records it, because 10% of the total 27477 N m is spent in the first half second and 68% in the last 1.3 s. The stop grows by 1.1 m. At the low-speed exit at $t = 2.26$ s the noisy command is 2200 N m, the actuator limit, but the exit holds the 895 N m the road is carrying and the wheel does not lock. Raising the slip filter to $\tau_f = 0.02$ s removes only 16% of the actuator travel, because the noise already has a 0.01 s correlation time. Halving K_c to 3000 N m removes 48% of the travel but lengthens the stop by a further 1.8 m. The absent lock is partly a property of the seed. Over noise seeds 1 to 10 the default gain locks the wheel in four runs, three of them after the exit latch, and the halved gain in five, two of them after the latch. Neither repair makes the slip signal informative at 3 m/s, which is the argument for the exit.

4.6.5 Rule-Based Comparison and the Graded Challenge

The rule-based cycle reaches the peak faster than the PI loop because it ramps pressure on at a fixed 12000 N m/s regardless of the error, and it then hunts. On dry asphalt it stops in 36.7 m against 33.0 m and spends 17910 N m of actuator travel against 730, and that pressure cycling is the pulsation a driver feels at an anti-lock pedal. On snow the same fixed rates fail, because 12000 N m/s is an enormous step next to the 218 N m the surface can hold, and the cycle overshoots into lock for 1.42 s and stops in 77.7 m against 71.1 m. Moving its setpoint to the 0.060 snow peak does not help, giving 80.6 m, and although the wheel no longer locks, slip still peaks at 0.32. Fixed modulation rates do not transfer between surfaces, which parallels the fixed-setpoint result of configurations 3 to 5.

The graded challenge starts on wet asphalt at 20 m/s with $K_c = 1200$ and $\tau_I = 0.8$ s and turns the road to black ice at $t = 1.5$ s. The run is scored at $t = 3.6$ s, and it passes when slip stays within ± 0.03 of setpoint over the final third, from $t = 2.4$ s, and the wheel spends no more than 0.02 s locked in total, with the model parameters and the road as given. As loaded, the loop is so slow that slip has reached only 0.023 by the time the ice arrives, far short of the wet peak. The ice excursion then peaks at 0.410 and does not return within ± 0.03 until $t = 3.17$ s, so the worst error over the scored window is 0.086, nearly three times the band, although the wheel never locks. Restoring $K_c = 6000$ and $\tau_I = 0.12$ s brings slip back within the band 0.41 s after the ice and leaves a worst error of 0.0007 over the scored window, inside the band with 0.029 to spare. Run on to the stop, the retuned loop cuts the distance from 269.7 m to 94.3 m and the slip IAE from 0.503 to 0.197, mostly by braking properly during the 1.5 s of wet asphalt rather than by doing anything clever on the ice, because the car meets the ice at 8.5 m/s instead of 15.4 m/s. Raising K_c helps twice, because it enlarges the proportional response and it also speeds the integral unwind for a given τ_I, while shortening τ_I helps once and spends stability margin.

Configuration 1 has no controller and precedes the closed-loop configurations. The wheel rolls freely at 27 m/s with $\lambda = 0$ and $\mu = 0$, so the car does not decelerate at all. Separate stops at constant pedal torques, each restarted from 27 m/s, move the operating point along the stable left branch of Figure 4.2, from a slip of 0.092 at 1250 N m to 0.155 at 1337 N m one second into the stop, until the equilibrium disappears at 1337.0 N m. Walking the pedal up within a single run does not reproduce this, because with 100 N m steps every 0.5 s the car stops after 5.4 s with the pedal at only 1100 N m. The scorecard is not meaningful for this configuration, accumulating the full 0.170 setpoint error as IAE (0.51 over 3 s of coasting) with no loop closed, while it correctly reports 0.0% saturation for a released brake.

4.6.6 Sliding-Mode Law Against the PI Loop

Configuration 12 is configuration 6 with the sliding-mode law of Section 4.5.4 selected at $\rho = 1$. The comparison in Table 4.7 runs each case to the stop under four structures with the same seed, instruments and noise: the shipped PI loop (and, on the challenge, its suggested tuning), the rule-based cycle, the law, and the PI loop retuned once by the law's own rule, with $K_c = 2\beta J v / r$ set every execution from the speed it reads and $\tau_I = 2/\beta = 0.144$ s. Nothing is retuned per case. Torque chatter is the root-mean-square difference between the 2 ms command and its 42 ms centered mean over the regulated part of the stop (automatic, $t \geq 0.1$ s, $v \geq v_{\text{exit}}$). Peak slip error is the largest $|\lambda - \lambda_{\text{SP}}|$ while regulating after 0.5 s, and recovery is the time after a surface change until slip stays within ± 0.03 of the setpoint for 0.2 s.

The law stops later than the PI loop on the two high-friction roads, 34.58 m against 32.97 m on dry asphalt and 41.37 m against 40.54 m on wet asphalt, and ties it on snow (71.20 m against 71.13 m) and ice (65.28 m against 65.24 m), as listed in Table 4.7. At 27 m/s its proportional gain is 3864 N m per unit slip against the PI loop's fixed 6000, so slip reaches the setpoint later. The same falling gain lowers its peaks on snow and ice, 0.170 and 0.193 against 0.235 and 0.370. On the ice patch of configurations 6 and 12 the law's slip peaks at 0.849 against 0.655 and recovers in 1.00 s against 0.66 s, because at the peak (14.4 m/s, $t = 1.29$ s) its gain is 2066 N m per unit slip, a third of the PI

Table 4.7 Sliding-Mode Law Against the Shipped Structures, Each Case Run to the Stop

Case	Structure	Distance [m]	Locked [s]	Peak λ	Peak error	Recovery [s]	IAE	MV travel [N m]	Chatter [N m]
2 Dry	PI	32.97	0.000	0.170	0.003	–	0.025	730	0.4
	rule-based	36.66	0.000	0.552	0.356	–	0.256	17910	30.7
	sliding mode	34.58	0.000	0.170	0.032	–	0.052	937	0.2
	scheduled PI	34.58	0.000	0.170	0.031	–	0.052	938	0.2
3 Wet	PI	40.54	0.000	0.170	0.001	–	0.022	448	0.2
	rule-based	47.61	0.414	1.000	0.830	–	0.952	27748	29.2
	sliding mode	41.37	0.000	0.170	0.006	–	0.039	626	0.2
	scheduled PI	41.37	0.000	0.170	0.006	–	0.039	627	0.2
4 Snow	PI	71.13	0.000	0.235	0.000	–	0.037	1144	3.2
	rule-based	77.66	1.424	1.000	0.631	–	3.221	38313	29.7
	sliding mode	71.20	0.000	0.170	0.001	–	0.040	297	0.3
	scheduled PI	71.20	0.000	0.170	0.001	–	0.040	297	0.3
5 Ice	PI	65.24	0.000	0.370	0.003	–	0.100	955	2.7
	rule-based	65.31	0.000	0.851	0.422	–	4.311	20538	26.4
	sliding mode	65.28	0.000	0.193	0.000	–	0.078	189	0.2
	scheduled PI	65.28	0.000	0.194	0.000	–	0.078	189	0.2
6, 12 Dry to ice	PI	218.09	0.000	0.655	0.485	0.66	0.280	2267	11.0
	rule-based	257.70	0.214	1.000	0.830	–	3.914	55154	23.8
	sliding mode	239.29	0.000	0.849	0.679	1.00	0.488	2334	4.6
	scheduled PI	239.23	0.000	0.857	0.687	0.97	0.495	2373	4.5
7 Wet to dry	PI	36.88	0.000	0.170	0.096	0.13	0.025	1235	13.2
	rule-based	41.57	0.230	1.000	0.231	–	0.570	21178	29.4
	sliding mode	37.90	0.000	0.170	0.114	0.50	0.077	1057	3.8
	scheduled PI	37.90	0.000	0.170	0.115	0.49	0.076	1059	3.8
13 Challenge	PI as loaded	269.74	0.000	0.410	0.240	1.67	0.503	913	0.7
	PI suggested	94.28	0.000	0.629	0.459	0.41	0.197	1704	11.9
	rule-based	128.83	0.176	1.000	0.830	–	3.648	40919	25.4
	sliding mode	103.02	0.000	0.895	0.725	0.73	0.394	1642	4.1
	scheduled PI	103.01	0.000	0.905	0.735	0.80	0.398	1680	4.0

loop's, to unwind the 1337 N m stored for dry asphalt. The law releases the brake fully and never locks the wheel there, as shown in Figure 4.5. On the challenge it meets the target with a worst error of 0.006 over the scored window, where the PI loop as loaded leaves 0.086 and its suggested tuning 0.001. It beats the rule-based cycle on every case in distance, lock, actuator travel and chatter.

The law and the scheduled PI loop agree within 0.06 m on every case in Table 4.7. At $\rho = 1$ the switching term saturates for 0.00% of the regulated time in every configuration, and replacing $\mathrm{sat}(\sigma/\phi)$ by σ/ϕ gives a bit-identical command and slip on all eight comparison cases. The layer sized to the brake lag is wider than any slip error these roads produce, so at the shipped setting the law is a proportional-integral loop whose gain is scheduled on speed and whose integral is bounded by the friction bound. Its robustness argument, the reaching condition, is never called on.

Three counterfactual runs remove one part of the law at a time, as listed in Table 4.8. Without the boundary layer ($\rho = 0$) the switching term acts every 2 ms as a relay between 0 and 2200 N m. The full switching gain catches the dry peak at once and the dry stop shortens to 32.77 m, but the chatter rises from 0.2 to 892.6 N m and the actuator travel from 937 to 69778 N m. In configuration 12 run to $t = 3.5$ s the ice-patch peak falls from 0.849 to 0.655 while the chatter rises from 12.3 to 538.8 N m, and run to the stop the wheel locks for 0.27 s after the low-speed exit. The engine's limit cycle runs at

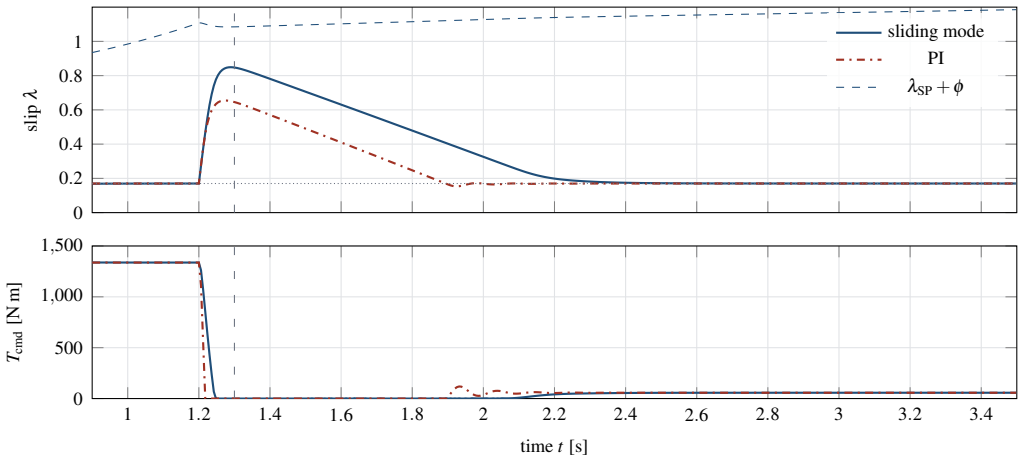


Figure 4.5 Configuration 12, dry asphalt turning to ice at $t = 1.2$ s, under the sliding-mode law at $\rho = 1$ (solid) and under the PI loop (dash-dotted), with the upper edge of the law's boundary layer dashed and the setpoint dotted. At $t = 1.3$ s (vertical line) the law's slip is 0.847 with $\sigma = 0.678$ inside a layer of $\phi = 0.915$ and its command is clamped at zero, while the PI loop's slip is 0.646. The upper edge of the layer stays above every slip reached.

Table 4.8 Counterfactual Variants of the Sliding-Mode Law, Run to the Stop

Case	Variant	Dist. [m]	Locked [s]	Peak λ	Chatter [N m]
Dry	$\rho = 0$	32.77	0.000	0.309	892.6
	no switching	38.15	0.108	1.000	1.2
	no integral	36.89	0.000	0.445	0.2
Wet	$\rho = 0$	41.05	0.000	0.535	792.0
	no switching	49.57	0.080	1.000	2.0
	no integral	40.91	0.282	1.000	0.9
Snow	$\rho = 0$	73.73	0.000	0.685	548.1
	no switching	84.30	0.924	1.000	2.0
	no integral	87.57	4.692	1.000	4.9
Ice	$\rho = 0$	65.23	0.000	0.713	308.5
	no switching	67.12	0.000	0.436	1.2
	no integral	65.25	14.128	1.000	5.6
Dry to ice	$\rho = 0$	212.15	0.272	1.000	372.5
	no switching	284.23	0.236	1.000	1.8
	no integral	249.66	14.068	1.000	4.8
Challenge	$\rho = 0$	95.99	0.000	0.503	403.4
	no switching	142.05	0.196	1.000	1.7
	no integral	94.91	14.180	1.000	6.1

16.7, 18.3 and 20.0 Hz at 24.6, 15.6 and 8.6 m/s against the 15.6 Hz the describing function predicts, with slip half-swings of 0.046, 0.072 and 0.101 against a predicted 0.087, 0.133 and 0.238. The relay is lopsided between the clamps, which halves the amplitude, and the tire's slope below the peak reduces it further, while the growth as $1/v$ survives. A thin layer sits between the two settings, with $\rho = 0.25$ giving a 0.610 peak and 31.4 N m of chatter in configuration 12 (saturated 9.5% of the time) and $\rho = 0.5$ giving 0.706 and 22.4 N m. Doubling the layer to $\rho = 2$ locks the wheel for 0.18 s on the ice patch. Without the switching term the equivalent control alone is an integral-only law, and most cases lock (snow 0.92 s, dry to ice 0.24 s). Without the integral, with $\hat{F}_x$ frozen at the middle of the bound, the wide layer leaves an offset that locks the wheel for 14.1 s on ice.

Safety is judged over every existing configuration as shipped and with each single extreme a learner can set (setpoint, slip filter, every vehicle and loop constant at its minimum and maximum,

Table 4.9 Worst Outcome per Structure Over Every Configuration and Single Extreme

Structure	Runs	With a lock	Longest lock [s]
PI	336	80	5.04 (6, noise 4 rad/s)
Rule-based	348	201	9.11 (6, $\tau_b = 0.2$ s)
Sliding mode	300	60	4.97 (11, $v_0 = 40$, manual)

the load at -20 and $+40\%$, and 4 rad/s of sensor noise), 276 cases shared by the three structures, plus each structure's own settings at their extremes. Table 4.9 gives the worst outcome per structure. The law locks the wheel in fewer runs, 60 of 300 against 80 of 336 for the PI loop and 201 of 348 for the rule-based cycle. Over the shared cases it locks where the PI loop does not, or for more than 0.05 s longer, in 19 cases, and the PI loop does so against it in 36. The longest of the 19 is 1.41 s on the challenge with a 0.2 s brake, where the PI loop locks for 0.50 s, and in 24 further cases the law's slip overshoots the PI loop's by more than 0.1. One of the 19 is a shipped configuration. Under the law configuration 8 locks the wheel for 0.126 s after the low-speed exit, because the law reaches the latch at $t = 2.27$ s with slip at 0.194, just past the dry peak and rising, and the lagged brake applying 1340 N m against the 1334 N m hold, so the falling branch carries the wheel to lock. The PI loop reaches its latch with slip at 0.090. The cause is the margin of Equation (4.29) at a 0.12 s brake, 0.6° against 41.8° at the declared 20 ms, since the law keeps its gain from the declared lag.

The friction bound shapes the load trend. Inside the bound the law's loss to the PI loop on dry asphalt grows with the load, from 1.61 m at 0% (34.58 against 32.97 m) to 2.57 m at $+40\%$ (27.20 against 24.63 m), and the scheduled PI loop loses the same (27.19 m), so the cause is the lower gain and not the projection. A load of $+60\%$, past the bound and past what the input allows, pins $\hat{F}_x$ at $\bar{F}_x$. The law then under-brakes at a mean slip of 0.091 and stops in 25.81 m against 22.18 m, without locking.

4.6.7 The Application Display

The application runs on the framework of Section 1.6, with Manual and Auto modes and a full export of the logged trajectory. The vehicle speed and the tire surface speed are displayed together because their divergence is the definition of the controlled variable, and Figure 4.4 shows configuration 6 paused 4.83 s into the stop.

4.7 Technology in Practice

What Measures It

Slip is not measured. It is computed from two speeds, and only one of them comes from an instrument at the wheel. Airplane antiskid systems first used the braked-wheel sensor only to detect a fast deceleration, and the later ones that regulate slip also need a reference for the ground speed [7]. Tests on a transport airplane's main wheel found the largest braking friction at a slip of about 0.1, while the side force fell so fast that above about 0.2 little cornering capability was left, so about 0.1 was suggested as the best target for an antiskid design [25]. The chapter's sensor model is the wheel speed with optional noise of 0.01 s correlation time, the vehicle speed reaches the estimator exactly, and its setpoint of 0.170 sits at the dry peak itself.

> **TECHNOLOGY** **Sensor: Wheel Speed and a Ground-Speed Reference**
>
> Each braked wheel of an airliner carries a sensor whose signal is proportional to the wheel's angular velocity. Most antiskid systems derive the ground speed from these sensors too, and some read it from an unbraked nose wheel or a small auxiliary wheel [7]. A slip-ratio system that holds the wheel about 10 % below the ground speed never lets it spin up, so it needs a reference from elsewhere [25]. When the brakes went on before the wheels had spun up on a wet runway, the logic took a ground speed well below the true one, and the traces of the front wheels showed deep skids for about 30 s [7].

What Moves It

The brake is moved by fluid pressure. In a car's antilock system the controlling device sends its signals to modulators, which adjust the brake actuating forces [26]. In the NASA tests the antiskid control valve sat between the pilot's metering valve and the brake, and a control box fed by the wheel-speed sensor drove it [25]. The chapter's modulator is a 20 ms first-order lag on torque with a clamp at 2200 N m, and its rule-based cycle applies and releases at fixed rates, as the early on-off valves did [7].

> **TECHNOLOGY** **Final Element: A Servo Valve and a Brake With Hysteresis**
>
> An airliner's antiskid logic modulates the hydraulic pressure at each wheel brake through a servo control valve. Early valves were on-off and applied or released pressure at a single rate, while later ones are constant-gain or variable-gain servo valves [7]. Behind the valve, the brake does not turn pressure into torque along a single curve. Tests during antiskid cycling showed large hysteresis loops, so one pressure gave a wide range of torque. Neither a linear spring with viscous damping nor an undamped nonlinear spring reproduced them, while a nonlinear spring with friction damping cut the torque error significantly [25]. The chapter's lag on torque carries no hysteresis.

Where the Controller Runs

In a car, control is executed by electronic control units, microprocessors with input and output interfaces, and a modern car carries more than a hundred of them. Once computers were introduced they took on anti-lock braking, electronic braking and stability control, braking the wheels one by one [2]. No open, vendor-neutral source read for this chapter gives the execution period of a production antilock unit. The chapter's laws run every 2 ms, and one function holds its structure: the velocity-form PI with back-calculation and the low-speed latch, or in its place the rule-based cycle or the sliding-mode law. Figure 4.6 draws the loop on vehicle hardware.

What Is Modernizing It

The first development replaces hydraulics at the brake. In aircraft braking, electromechanically actuated brakes have found wide acceptance as an alternative to hydraulic ones, and advanced antiskid controllers take advantage of them [27]. The second is electric braking under the antilock function. The United States light-vehicle brake standard covers electrically actuated service brakes and electric transmission of the brake signal, and an electric vehicle whose regenerative braking is part of the service brake must have that braking controlled by its antilock system [26]. The third is learning the peak on line. One design estimates the vehicle's states and the tire curve's parameters with a particle filter while it brakes, choosing each action for what it will both earn and reveal, and in simulation it stopped up to 8.5 % shorter than extremum seeking [28]. Another drives automated vehicles to high slip on purpose during empty trips, to estimate the peak friction of the road [29].

> **IN PRACTICE** **What a Brake Test Counts as a Locked Wheel**
>
> In the United States, Standard No. 135 (49 CFR 571.135) governs light-vehicle brakes. It defines an antilock system as the part of the service brake that controls wheel slip by sensing wheel rotation, passing those signals to controlling devices and sending their outputs to modulators. Wheel lockup means 100 % slip, and in many of its stops no wheel may lock for longer than 0.1 s above 15 km/h. Any electrical failure of the antilock system must light the brake warning indicator, and the test sequence includes stops with the antilock function failed [26]. The chapter's lock flag counts the time spent above a slip of 0.92.

4.8 Limitations

The controller is given the vehicle reference speed exactly. Estimating that speed from four wheel speeds when all four may be slipping is the hardest part of a production anti-lock system, and none of it is present here, so the application should not be read as a model of anti-lock braking end to end.

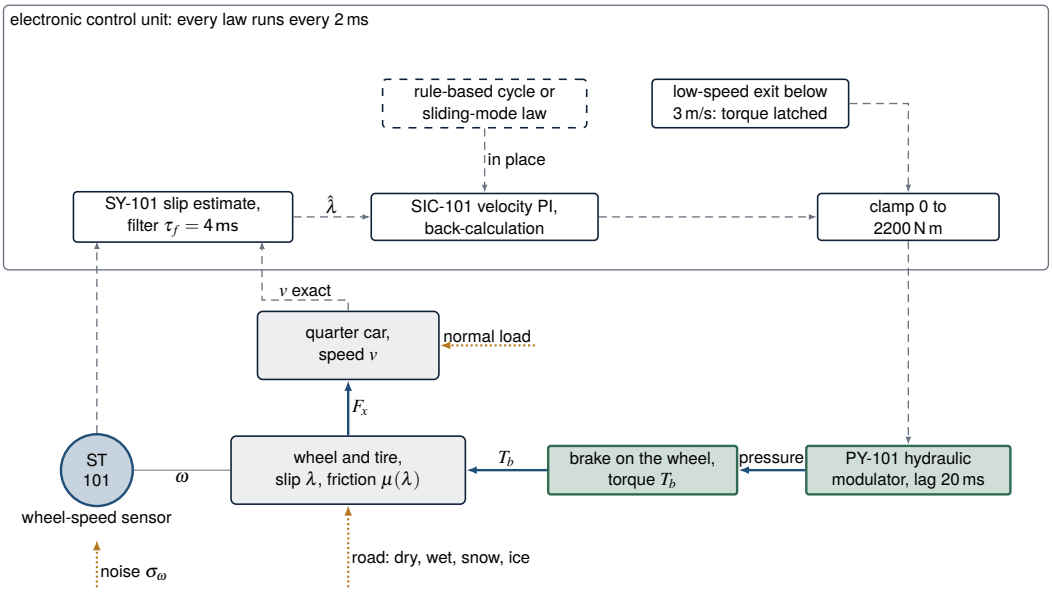


Figure 4.6 The slip loop on vehicle hardware. The wheel-speed sensor ST-101 reads the wheel, and the slip estimate SY-101 divides by the vehicle speed, which the chapter hands over exactly. The velocity-form PI SIC-101 (or, in its place, the rule-based cycle or the sliding-mode law, dashed) sends a torque command, clamped to the brake limit, to the hydraulic modulator PY-101, and below 3 m/s the exit latches a hold torque. Notice that the controller sees neither the road nor the load (dotted arrows), so a change in either reaches it only through the slip it causes.

The friction map is static and single valued. Relaxation length, temperature, wear and water film thickness are absent, and a dynamic friction model would change both the achievable bandwidth and the shape of the transition in configuration 6 [14]. On ice the curve has no peak at all, and the displayed peak slip of 0.0150 is a convention rather than a maximum.

⚠ OPEN CHALLENGE Finding a Peak Without Falling Off It

Estimates of the slip and of the road are often made apart from the antilock controller, which leads to sub-optimal braking. The peak is harder to identify on snow or ice, where it is less defined than on dry or wet roads [28]. Ordinary driving gives little help, because it stays at slips below about 0.05, too low to show the region near the peak [29]. For aircraft braking on a runway that can change fast, the estimation is called an extremely challenging and still open task [27]. Progress would be a loop that learns the peak while braking on it, without the lock it exists to prevent.

The scorecard was designed for regulation against a setpoint, and two of its figures need care here. Overshoot and settling are not reported, because the slip setpoint never steps, and the peak slip error reported instead includes the initial approach from $\lambda = 0$, so it reads 0.151 even for the reference stop. The theoretical best distance is frozen when the brake is applied and is therefore wrong by construction after a surface change, which is deliberate but has to be stated whenever configuration 6 is quoted.

The linearization in Section 4.4 holds v fixed, which is a good approximation because the slip dynamics are one to two orders of magnitude faster than the speed dynamics at every speed in Table 4.4. The phase margins quoted there are still properties of a frozen plant rather than guarantees over a whole stop.

The sliding-mode law's model leaves out the brake lag, the slip filter and the sample delay, which enter only through the pole β and the layer width, and it keeps β from the declared 20 ms brake, so a slower brake in the plant erodes its margin without the law knowing, to 0.6° at the 0.12 s of configuration 8. Its friction bound covers the shipped roads up to dry asphalt's peak at $+40\%$ load. A force beyond it leaves the estimate pinned and the wheel under-braked. Its reaching condition holds only while the brake can deliver the switching torque, and at the shipped layer it is never exercised, so the application demonstrates the boundary-layer trade rather than the robustness of an ideal sliding mode. The vehicle's mass, radius and inertia are treated as coded specifications for every controller, so a learner's change to them reaches the controllers as a recalibrated vehicle rather than as model error.

4.9 Conclusion

This chapter has used a browser simulation of quarter-car braking with a Burckhardt tire to study regulation to the maximum of a plant characteristic. Four states are integrated with fourth-order Runge-Kutta on a 2 ms step, the friction coefficient comes from three coefficients per surface, and the four shipped surfaces place the friction peak between $\lambda^* = 0.0150$ on ice and 0.1700 on dry asphalt with peak friction between 0.0500 and 1.1700. The constant-torque equilibrium is shown to lose stability at the maximum of the hold-torque curve rather than at the maximum of the friction curve, at $\lambda_c = 0.1661$ and 1337.0 N m on dry asphalt, and that boundary carries no speed dependence while the loop crossover does. A fixed-gain proportional-integral loop on slip stops the car in 33.0 m against a theoretical 31.8 m, where a locked wheel needs 47.9 m, and the same tuning falls from 40.6° of phase margin at 27 m/s to 15.2° at the 3 m/s handover. A mid-stop change from dry asphalt to ice steps the required torque by a factor of 23.4 and produces a slip excursion to 0.655 that recovers in 0.66 s, and the mirror-image change produces no excursion above the setpoint. A sliding-mode law with its boundary layer sized to the brake lag never saturates its switching term, acts as a speed-scheduled PI loop that stops 1.6 m later on dry asphalt and peaks at 0.849 on the ice patch, and locks the wheel in 60 of 300 stress runs against 80 of 336 for the fixed-gain loop. The natural extensions would estimate the surface on line, add a relaxation length to the tire, model the reference-speed estimator, and carry the lateral force that makes anti-lock braking worth fitting on ice. A moving horizon estimator or a Kalman filter could treat the Burckhardt coefficients as slowly varying parameters and reconstruct the friction peak from the measured wheel speed, which is the identification problem of Section 4.2 [20]. Nonlinear model predictive control could enforce the torque limit and a slip ceiling

as explicit constraints instead of relying on anti-windup and a low-speed exit [30]. The application gives a learner a control loop whose setpoint sits exactly where its process gain vanishes.

KEY POINTS

- A braked tire gives its largest force at a small slip, and past that slip more brake torque gives less force.
- The constant-torque equilibrium loses stability at the maximum of the hold-torque curve, 1337.0 N m on dry asphalt, a boundary with no speed in it.
- The same PI tuning falls from 40.6° of phase margin at 27 m/s to 15.2° at the 3 m/s handover.
- An ice patch is a windup problem with the output far from its limit: slip runs to 0.655 and recovers in 0.66 s.

Where else this applies. Volume 4, Chapter 2 holds a compressor off the peak of its characteristic, and Chapter 2 meets the same fold of the holding input.

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

[3] U. Kiencke and L. Nielsen, *Automotive Control Systems: For Engine, Driveline, and Vehicle*, 2nd ed. Berlin: Springer Berlin Heidelberg, 2005.

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

[5] S. M. Savaresi and M. Tanelli, *Active Braking Control Systems Design for Vehicles*. London: Springer London, 2010.

[6] Z. Jiao, X. Liu, F. Li, and Y. Shang, "Aircraft antiskid braking control method based on tire–runway friction model," *Journal of Aircraft*, vol. 54, no. 1, pp. 75–84, 2017.

[7] W. B. Horne, J. L. McCarty, and J. A. Tanner, "Some effects of adverse weather conditions on performance of airplane antiskid braking systems," NASA Langley Research Center, Hampton, VA, Technical Note NASA TN D-8202, 1976, nASA Technical Reports Server 19760019083.

[8] H. Leiber and A. Czinczel, "Antiskid system for passenger cars with a digital electronic control unit," SAE International, SAE Technical Paper 790458, 1979.

[9] C. M. Farmer, "New evidence concerning fatal crashes of passenger vehicles before and after adding antilock braking systems," *Accident Analysis & Prevention*, vol. 33, no. 3, pp. 361–369, 2001.

[10] A. Erke, "Effects of electronic stability control (ESC) on accidents: A review of empirical evidence," *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 167–173, 2008.

[11] National Highway Traffic Safety Administration, "Standard no. 121; air brake systems, 49 CFR 571.121," Code of Federal Regulations, Title 49, Chapter V, edition of October 1, 2024, U.S. Government Publishing Office, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/CFR-2024-title49-vol6/pdf/CFR-2024-title49-vol6-sec571-121.pdf

[12] H. B. Pacejka and E. Bakker, "The magic formula tyre model," *Vehicle System Dynamics*, vol. 21, no. sup001, pp. 1–18, 1992.

[13] C. Canudas de Wit, H. Olsson, K. J. Åström, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 419–425, 1995.

[14] C. Canudas-de Wit, P. Tsiotras, E. Velenis, M. Basset, and G. Gissinger, "Dynamic friction models for road/tire longitudinal interaction," *Vehicle System Dynamics*, vol. 39, no. 3, pp. 189–226, 2003.

[15] C. Unsal and P. Kachroo, "Sliding mode measurement feedback control for antilock braking systems," *IEEE Transactions on Control Systems Technology*, vol. 7, no. 2, pp. 271–281, 1999.

[16] A. Harifi, A. Aghagolzadeh, G. Alizadeh, and M. Sadeghi, "Designing a sliding mode controller for slip control of antilock brake systems," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 6, pp. 731–741, 2008.

[17] S. Drakunov, U. Ozguner, P. Dix, and B. Ashrafi, "ABS control using optimum search via sliding modes," *IEEE Transactions on Control Systems Technology*, vol. 3, no. 1, pp. 79–85, 1995.

[18] T. A. Johansen, I. Petersen, J. Kalkkuhl, and J. Ludemann, "Gain-scheduled wheel slip control in automotive brake systems," *IEEE Transactions on Control Systems Technology*, vol. 11, no. 6, pp. 799–811, 2003.

[19] E. Dincmen, B. A. Guvenc, and T. Acarman, "Extremum-seeking control of ABS braking in road vehicles with lateral force improvement," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 230–237, 2014.

[20] F. Gustafsson, "Slip-based tire-road friction estimation," *Automatica*, vol. 33, no. 6, pp. 1087–1099, 1997.

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

[22] V. Utkin, "Variable structure systems with sliding modes," *IEEE Transactions on Automatic Control*, vol. 22, no. 2, pp. 212–222, 1977.

[23] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Prentice Hall, 1991.

[24] K. D. Young, V. I. Utkin, and Ü. Özgüner, "A control engineer's guide to sliding mode control," *IEEE Transactions on Control Systems Technology*, vol. 7, no. 3, pp. 328–342, 1999.

[25] S. M. Stubbs and J. A. Tanner, "Review of antiskid and brake dynamics research," in *1980 Aircraft Safety and Operating Problems, Part 2*. Hampton, VA: NASA Langley Research Center, 1981, pp. 555–562, nASA Technical Reports Server 19810010540.

[26] National Highway Traffic Safety Administration, "Standard no. 135; light vehicle brake systems, 49 CFR 571.135," Code of Federal Regulations, Title 49, Chapter V, edition of October 1, 2024, U.S. Government Publishing Office, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/CFR-2024-title49-vol6/pdf/CFR-2024-title49-vol6-sec571-135.pdf

[27] F. Crocetti, G. Costante, M. L. Fravolini, and P. Valigi, "Tire-road friction estimation and uncertainty assessment to improve electric aircraft braking system," in *29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 330–335.

[28] B. Sullivan, J. Jiang, G. Mavros, and W.-H. Chen, "An exploration-exploitation approach to anti-lock brake systems," arXiv preprint 2306.14730 [eess.SY], 2023. [Online]. Available: https://arxiv.org/abs/2306.14730

[29] Z. Liang, H. Zhou, H. Huang, and X. Li, "High-slip-ratio control for peak tire–road friction estimation using automated vehicles," in *2026 IEEE Intelligent Vehicles Symposium (IV)*, 2026, pp. 1901–1906.

[30] 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.

4.10 Exercises

4.1 ⌕ Concept check · 1 min

Only one signal of the quarter-car model in Figure 4.1 reaches the slip estimator through an instrument, and the app can add noise to it. Which signal is it?

(a) The vehicle speed v, read by a ground-speed sensor on the body
(b) The friction coefficient μ, inferred from the brake pressure
(c) The wheel angular speed ω, read by the wheel-speed sensor
(d) The applied brake torque T_b, read by a load cell at the caliper

4.2 ⌕ Concept check · 2 min

In configuration 11 the pedal is held at the 2200 N m limit on dry asphalt, and the car stops in 47.9 m against 33.0 m under the PI loop. Why does braking harder lengthen the stop?

(a) The brake command sits on its limit, so the wheel receives less torque than the pedal asks
(b) The low-speed exit releases the brake below 3 m/s, and the car coasts to rest
(c) The 20 ms hydraulic lag withholds the full torque for most of the stop
(d) The wheel locks, and a sliding tire delivers $\mu(1) = 0.7601$ instead of the peak 1.1700

4.3 ⌕ Concept check · 2 min

A PI slip loop holds $\lambda = 0.170$ when the road changes without warning. Which change threatens the wheel, and why?

(a) A fall in friction, because the integral holds a torque the new road cannot carry and pushes slip onto the falling branch

(b) A rise in friction, because the wheel spins up toward free rolling and the integral then winds up against the upper limit

(c) Both equally, because the loop is linear about its setpoint and answers either sign of change in the same way

(d) Neither, because back-calculation holds the integral at the applied torque through any change of the road

4.4 Theory and calculation · 6 min

Wet asphalt has the Burckhardt coefficients $c_1 = 0.8570$, $c_2 = 33.822$ and $c_3 = 0.3470$. Use Equations (4.7) to (4.9) and $g = 9.81\,\mathrm{m/s^2}$.

(a) Find the peak slip λ^* and the peak friction μ^*. *Checkpoint:* 0.1308 and 0.8013.

(b) Find the friction of a locked wheel, $\mu(1)$, and the percentage of the peak force it loses. *Checkpoint:* 0.5100, a loss of 36.4 %.

(c) Find μ at the default setpoint 0.170 as a fraction of μ^*, and the ideal stopping distance from 25 m/s. *Checkpoint:* 0.7953, 99.2 % of the peak, and 39.8 m.

4.5 Theory and calculation · 8 min

The default vehicle has $m = 340\,\mathrm{kg}$, $r = 0.33\,\mathrm{m}$, $J = 1.7\,\mathrm{kg\,m^2}$ and $g = 9.81\,\mathrm{m/s^2}$, with no load change, on dry asphalt ($c_1 = 1.2801$, $c_2 = 23.990$, $c_3 = 0.5200$, $\mu(0.170) = 1.1700$).

(a) Use Equation (4.13) to find the brake torque that holds $\lambda = 0.170$, and the share of it that decelerates the wheel itself. *Checkpoint:* 1336.9 N m, of which 49.1 N m or 3.7 %.

(b) Find the tire slope μ' and the pole p of Equation (4.15) at $\lambda = 0.25$ and $v = 27\,\mathrm{m/s}$, and the time constant of the mode. *Checkpoint:* $\mu' = -0.444$, $p = +4.05/\mathrm{s}$, about 0.25 s.

(c) Find the same pole at 5 m/s without repeating the calculation, and say what the result means for the slip loop. *Checkpoint:* 21.9/s, a time constant of 46 ms.

4.6 Theory and calculation · 8 min

Take the loop of Equation (4.17) on dry asphalt at $\lambda_{\mathrm{SP}} = 0.170$ with the default tuning $K_c = 6000\,\mathrm{N\,m}$ and $\tau_I = 0.12\,\mathrm{s}$.

(a) Show that for an unstable plant pole $p > 0$ the phase margin is $\arctan(\tau_I \omega_c) - 90^\circ + \arctan(\omega_c/p) - \arctan(\tau_b \omega_c)$, and evaluate it at 27 m/s, where $p = 0.42/\mathrm{s}$, $\tau_b = 0.020\,\mathrm{s}$ and the crossover is 36.0 rad/s. *Checkpoint:* 40.6°.

(b) Evaluate it at 3 m/s, where $p = 3.82/\mathrm{s}$ and the crossover is 135.0 rad/s. *Checkpoint:* 15.2°.

(c) Evaluate it for the slow actuator of configuration 8, $\tau_b = 0.12\,\mathrm{s}$, at 27 m/s, where the crossover is 19.0 rad/s, and say what the run will show. *Checkpoint:* −1.3°.

4.7 App challenge · 8 min

Load scenario 1, Rolling wheel, brake released. It runs in Manual from 27 m/s on dry asphalt at the default seed. Make four stops, each with Brake torque set before pressing Run and with Reset between them (Reset keeps the pedal): 1250, 1337, 1338 and 1340 N m. Run each to the stop, which takes under 3 s, at a playback speed of 0.25, and read the Stopping distance and Wheel locked tiles. Export the file of the 1250 and 1337 N m runs.

(a) Read both tiles for 1337 and 1338 N m. *Checkpoint:* 32.7 m with 0.00 s, and 32.9 m with 0.27 s.

(b) Read both tiles for 1250 and 1340 N m. *Checkpoint:* 34.8 m with 0.00 s, and 34.1 m with 0.87 s.

(c) Read column `slip` at 1.0 s in the two exported files, and compare the values with the slip λ_c at which the hold torque is largest. *Checkpoint:* 0.092 and 0.155, within 0.002, against $\lambda_c = 0.1661$.

Accept distances within 0.1 m and lock times within 0.01 s.

4.8 App challenge · 8 min

Load scenario 6, Dry → ice, one second into the stop, which runs the PI loop in Auto with the road turning to ice at 1.2 s, at the default seed. Make three runs to 3.5 s at a playback speed of 0.25, and export the file of each. For each run read the Peak slip error tile at 3.5 s, the first logged time after 1.2 s from which column `slip` stays within ± 0.03 of the setpoint 0.170, and column `vehicle_speed_mps` at 1.2 s.

(a) Run the scenario as loaded. *Checkpoint:* 0.484, 1.86 s and 13.77 m/s.

(b) Press Reset, set Integral time to 1 s before pressing Run, and run again. *Checkpoint:* 0.328, 2.58 s and 16.16 m/s.

(c) Reload the scenario, set Brake torque limit to 900 N m in the Parameters of Show Model before pressing Run, and run again. Explain why this run recovers fastest. *Checkpoint:* 0.259, 1.68 s and 17.98 m/s.

Accept the tile within 0.002, the time within 0.02 s and the speed within 0.02 m/s.

4.9 App challenge · 7 min

Load the graded challenge, listed last as wet asphalt that turns to ice. It stops from 20 m/s with $K_c = 1200\,\mathrm{N\,m}$ and $\tau_I = 0.8\,\mathrm{s}$, the road turns to ice at 1.5 s, and the target is slip within ± 0.03 of the setpoint from 2.4 s to 3.6 s with no more than 0.02 s of lock. Use the default seed and a playback speed of 0.25, run to 3.6 s, and export the file of each run. The worst error is the largest $|\texttt{slip} - \texttt{slip_sp}|$ over the rows from 2.40 to 3.60 s, as the app's verdict reports it.

(a) Run the challenge as loaded. Find the worst error, the largest slip before 1.5 s and column `vehicle_speed_mps` at 1.5 s. *Checkpoint:* 0.086, 0.023 and 15.43 m/s, no lock.

(b) Press Reset, set Proportional gain to 6000 N m and Integral time to 0.12 s before pressing Run, and run again. Find the worst error and the speed at 1.5 s. *Checkpoint:* 0.0007, within 0.001, and 8.45 m/s, no lock.

(c) The retuned loop lets slip peak higher on the ice than the loaded one. Explain why it still passes.

4.10 Application area · 4 min

A wheel-speed sensor on a wheel of radius 0.33 m carries noise of standard deviation $\sigma_\omega = 2.0\,\mathrm{rad/s}$, as in configuration 10, while the vehicle speed is known exactly.

(a) Find the standard deviation of the slip noise at 27, 5 and 3 m/s. *Checkpoint:* 0.024, 0.13 and 0.22.

(b) Give two reasons, one from the measurement and one from the plant, why the controller stops regulating slip at 3 m/s.

4.11 Application area · 3 min

The model omits rolling resistance and aerodynamic drag. Take a rolling resistance coefficient of 0.01, which adds a deceleration of $0.01g$.

(a) Compare that deceleration with what the tire supplies at its peak on ice and on dry asphalt. *Checkpoint:* 0.0981 m/s^2, 20 % of 0.491 on ice and under 1 % of 11.5 on dry asphalt.

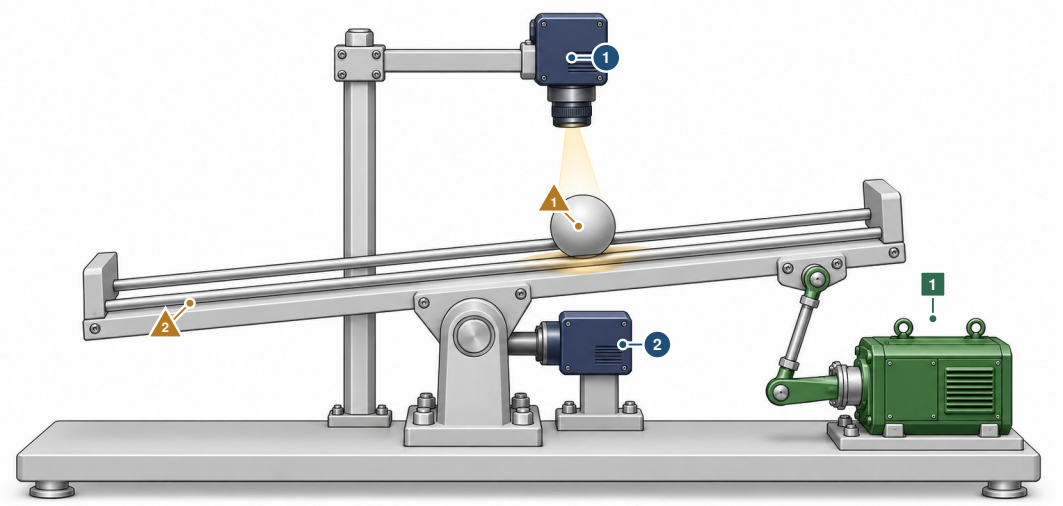
(b) Find the ideal stopping distance on ice from 8 m/s with and without it. *Checkpoint:* 54.4 m against 65.2 m.

4.12 Review the agent · 5 min

In scenario 8, Slow brake actuator ($\tau_b = 0.12\,\mathrm{s}$), slip oscillates between 0.084 and 0.264 at the default tuning. An AI assistant replies as follows. “The loop oscillates because its gain is too high for the slower actuator. Halve the proportional gain to 3000 N m and keep $\tau_I = 0.12\,\mathrm{s}$. A lower gain moves the crossover down to where the 0.12 s lag costs less phase, so the phase margin rises and the oscillation dies out.” Find the flaw in the proposal and say how it could be checked.

CHAPTER 5

A Desktop Ball and Beam Whose Only Damping Comes From the Controller



The controller moves 1 servo that tilts the beam through a push rod

The controller reads 1 camera above the beam, the ball's position 2 beam angle readback at the pivot

What disturbs it 1 a flick that sets the ball rolling 2 bent beam or unlevel table

The servo tilts the beam to bring the ball to its setpoint, and once the ball rolls almost nothing but the controller slows it. Illustration, not to scale.

IN PLAIN TERMS A Rolling Ball That Nothing Slows Down

The main idea. A ball rests on a long, thin beam that a small motor can tilt. The controller tilts the beam to bring the ball to a chosen spot, and once the ball rolls, almost nothing else can stop it.

The challenge. Suppose the controller only pushes harder the farther the ball is from its spot. Then the ball swings past again and again, wider each time. It also has to know how fast the ball is moving, and a noisy or late reading of the ball's place limits how well it can tell.

Why it matters. Spacecraft, disk-drive heads and machine slides move the same way, with little to slow them.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 5.2, the fact on event cameras in Section 5.3, the two try-its and the fact on the inner loop in Section 5.6, 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 5.6), Technology in Practice (Section 5.7), Limitations (Section 5.8) and the open challenge

The control question A ball on a tilting beam is two integrators with almost no friction, so a level beam leaves the ball rolling and a loop that knows only the position cannot hold it still. Where does the damping come from, and how much of it do the servo's lag, the sensor's noise, a camera's delay, a slow inner loop and an angle limit take away?

Learning objectives

- Derive the rolling factor of a ball on two rails and the acceleration a beam angle gives it.
- Predict the period and growth per swing of a proportional loop through a lag, and the damping a derivative adds.
- Explain why a derivative on a noisy reading needs a filter, and what the filter and a camera's delay cost.
- Decide from a stopping distance whether any controller can keep a flicked ball on the beam.
- Compare PD, PID and a four-state observer LQR on travel, integrated error and bend offset.

Prerequisites Sections 1.4.7 (lag limit, actuator headroom), 1.4.5 and 1.4.6 (LQR, Bryson's rule, observers) and 1.4.9 (growth through a lag), and Sections 1.6.1 and 1.6.2 (engine and velocity-form PID). Chapter 3 for a desktop kit with its own sample time and limits.

The app https://apmonitor.com/apps/ball-and-beam/

See also Volume 2, Chapter 4 for a sticking valve whose limit cycle no tuning removes, and Volume 2, Chapter 12 for profile feedforward on a machine-tool axis. Volume 3, Chapter 3 for the inner-loop speed of a cascade measured on a fermenter, and Volume 3, Chapter 10 for a position loop that commands an angle.

5.1 Introduction

Tilt a beam by 1° and a steel ball resting on its rails starts to roll at 0.112 m/s^2. Level the beam again and nothing slows the ball except a rolling resistance worth a twentieth of a degree of tilt. In the simulator of this chapter, a ball that has rolled for one second from the middle of a 1 m beam tilted 1° is moving at 10.6 cm/s when the beam is leveled, and it still rolls off the end 4.45 seconds later at 8.8 cm/s. Stopping it takes an opposite tilt held as long as the first one, and a level beam at the instant the speed reaches zero. A hand on a slider is too late for both, which is the first lesson the rig teaches.

The rig is built to correct a common belief, that a controller which pushes harder the farther the output is from its target will bring it there. On a plant with no damping of its own a proportional law is a spring. With 0.30 °/cm of tilt per centimeter of error the ball swings past its target with a period of 3.47 s, close to the 3.43 s of an undamped oscillator, and the servo's 60 ms lag feeds energy in at every swing, so each swing is 1.37 to 1.42 times the last until the ball leaves the beam at 16.41 s. Doubling the gain shortens the period to 2.48 s and loses the ball sooner, at 9.94 s. Only a term that knows the ball's speed adds damping, and the sensor that supplies the speed then limits how much of it the loop can carry.

The shape is general. Two integrators from a force or an angle to a position, with little natural damping, describe a spacecraft turning on its thrusters, a positioning stage on air bearings and a disk-drive head, and the double integrator is a standard test of how controllers behave under off-nominal conditions [1, 2]. A dead zone like the ball's breakaway is the stiction of a valve stem or the static friction of a machine slide, and integral action across one makes a loop hunt between two sides of its setpoint [3]. Volume 2, Chapter 4 models a sticking valve whose limit cycle no tuning removes. A position loop that asks an inner drive for an angle is a cascade, whose inner loop must be faster than the outer one it serves [4]. Each of these appears in the simulator as a case of its own.

The published work on the ball and beam is mostly about one design on one rig, a nonlinear stabilizing law [5–7] or a comparison of model-based and model-free controllers on a laboratory apparatus [8], and the virtual and remote laboratories built around it give access to a rig rather than a way to separate its effects [9, 10]. A student who sees the ball ring on hardware cannot tell how much of the ringing comes from the servo, the sensor, the sampling or the rail. This chapter separates those causes on a simulated desktop ball and beam, brings them in one by one, and measures what each one costs a loop. The rolling-ball model, the shared velocity-form PID, the trajectory feedforward [11],

the dither [12] and the linear quadratic regulator are all standard methods. The measurements are made on one model and one seed.

Section 5.2 follows the rig through the literature. Section 5.3 states the model, with every equation, constant and default, and with the ball's true state logged beside each sensor's reading. Section 5.4 analyzes the linear loop. It shows a P-only growth per swing that follows $e^{\pi\omega(\tau_s+T_s/2)}$ less a fixed loss to the rolling resistance, the damping that only the derivative supplies, and what the sensor, a delay and an inner drive cost. Section 5.5 describes the controllers as coded, together with a loop view whose original-law stability verdict comes from the eigenvalues of the sampled closed-loop matrix. Section 5.6 reports ten cases and a relay race as measured. Among the results are a camera that costs the loop 14.4° of phase at its 5.01 rad/s crossover, a cascade that goes unstable when the inner bandwidth falls below 1.53 times the outer crossover, and a ninth case that no controller can meet. Section 5.7 describes the sensors, servos and controllers of real rigs. Sections 5.8 and 5.9 state the limitations and summarize the results, and the exercises follow in Section 5.10.

5.2 The Ball and Beam in the Laboratory and the Literature

Rolling-ball rigs entered the control literature as laboratory apparatus that deserved analysis in their own right. A ball rolling inside a hoop that a motor turns, a close relative of the beam, was described in 1983 as a laboratory system of that kind [13]. The ball on a beam became a benchmark for nonlinear design a few years later, as the history segment of this section tells. Through the 1990s the plant stayed a proving ground for semi-global stabilization by output feedback [6] and for global asymptotic stabilization [7]. The simulator keeps the centripetal term in its model and leaves it out of its linear analysis, and case 4 below shows the term at work when a noisy derivative shakes the beam.

> ⧖ HISTORY **A Bead on a Wire Defeats Exact Linearization, 1989**
>
> A research memorandum of 1989 took the ball and beam, a common undergraduate control laboratory experiment, as the motivating example for approximate input-output linearization [14], the method that Section 1.4.3 introduces. Its ball slid without friction along the beam, like a bead on a wire, so the beam could turn freely without losing it. Differentiating the position brought in the input with a coefficient that is zero whenever the beam's angular velocity or the ball's position is zero, so the relative degree is not defined and exact input-output linearization fails. Full state linearization fails too. Neglecting the centrifugal term, the $r\dot{\alpha}^2$ of Equation (5.1), gave approximate designs that tracked a large sinusoid where a design on the Jacobian linearization went unstable. The work appeared as a journal paper in 1992 [5].

Laboratory practice followed a simpler path. A proportional-derivative law on the measured position is what most rigs run, and its stability on the ball and beam has been analyzed directly [15]. The difficulty in practice is the speed. A laboratory study that modeled the rig with its nonlinear and coupling effects tuned a linear quadratic regulator by a genetic algorithm, and because the position sensor was noisy it estimated the ball's speed with a state observer before comparing the result with PID control [8]. The original laws offer the same two routes to the speed, a filtered difference of the reading and an α-β estimator. Access then widened through virtual and remote laboratories that put a ball and beam behind a web page [9], one of them inside the Moodle course management system [10], and a desktop heat-transfer kit built around a microcontroller showed what a small, cheap, sampled rig offers as a benchmark for modeling and control [16].

Three remedies the cases teach come from other machines. Dead zones and the hunting that integral action makes across them were catalogued for machines with friction, along with dither as a compensation [3]. The theory of dither treats a fast periodic signal as a way to smooth a nonlinearity seen by a slower loop [12], and process practice adds short pulses to the command of a sticking control valve for the same purpose [17]. Motion systems plan their setpoint as a trajectory and feed forward the effort that trajectory needs, which leaves the feedback loop only the error [11]. Volume 2,

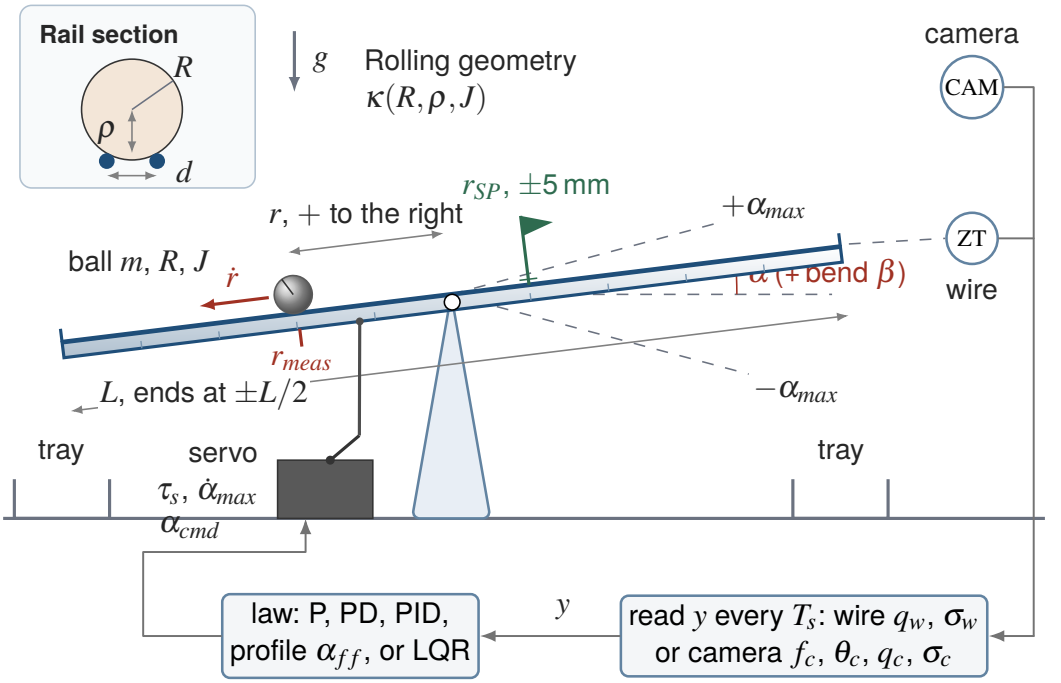


Figure 5.1 The simulated ball and beam, as the Rig view draws it, and the kit's chain from the reading to the servo, not to scale. The beam of length L is tilted by α (with an unknown bend β) within $\pm\alpha_{max}$ by a servo with lag τ_s and rate limit $\dot{\alpha}_{max}$ that receives the command α_{cmd}. The ball of mass m, radius R and inertia J sits at r with speed $\dot{r}$, and the section shows it touching rails d apart at the rolling radius ρ, which sets the rolling factor κ. The target flag marks r_{SP} with its ± 5 mm band, the short tick below the rail is the wire's quantized reading r_{meas}, and the trays catch a ball past $\pm L/2$. Every T_s the kit reads y from the wire (ZT) or the camera (CAM) and runs its law.

Chapter 12 applies a jerk-limited profile with its feedforward to a machine-tool axis. Case 6 brings the dead zone and the dither to the beam, and cases 5 and 11 the profile and its feedforward.

5.3 The Kit as Simulated

The simulated kit is a 1.0 m aluminum beam on a center pivot, tilted by a hobby servo on the desk through a horn and a push rod 12 cm left of the pivot, as shown in Figure 5.1. Two rails 12.7 mm apart carry a solid steel ball 25.4 mm across (66 g), or a hollow ping-pong ball 40 mm across (2.7 g), and a resistance wire between the rails reads the ball's position over ±45 cm. A camera above the beam can replace the wire. The position r is measured along the beam from the pivot, positive to the right, and the beam angle α is positive when it raises the right end, so a positive angle rolls the ball toward $-r$. Past either end the ball drops into a tray. Every dimension, limit and threshold of the rig is a teaching value chosen for a desktop kit rather than taken from any product, which the model dialog and the Instructions page both state. The signals are sorted by role in Table 5.1.

5.3.1 The Ball on Two Rails

A sphere of mass m, radius R and moment of inertia J about its center rolls without slip on two rails a distance d apart. It touches them at the rolling radius ρ, so it turns through r/ρ while it travels r, and

Table 5.1 Signals of the Kit by Role

Symbol	Role	Nominal	Range
r	CV and state, ball position	per case	SP ±45 cm
$\dot{r}$	state, ball speed	0	—
α	state, beam angle	0	±15°
ω	state, beam rate (torque beam)	0	—
α_{cmd}	MV, beam angle command	per law	$\pm\alpha_{max}$
y	measured, wire or camera reading	r	±45 cm wire
$\hat{\alpha}$	measured, servo readback	α	0.01° steps
β	DV, bent beam or unlevel table	0	±2°
$\Delta\dot{r}$	DV, finger flick	20 cm/s	±50 cm/s, event
$\Delta\alpha$	DV, bump to the beam	2°	event
—	DV, sensor dropout	1 s	event
$m/\hat{m}$	DV, heavier ball (torque beam)	1	1 to 3
$\hat{\dot{r}}$	estimate, α-β speed	—	—
PM	output, phase margin of the tuning	—	—

Lagrange's equations for the ball on a beam turning at $\dot{\alpha}$ give Equation (5.1).

$$\left(1+\frac{J}{m\rho^2}\right)\ddot{r} = r\dot{\alpha}^2 - g\sin(\alpha+\beta) - g\sin\alpha_k\, s$$
$$\rho = \sqrt{R^2 - d^2/4}, \qquad \kappa = \frac{1}{1+J/(m\rho^2)} \tag{5.1}$$

where β is a tilt the controller does not know about, α_k is the rolling resistance written as an equivalent tilt, and $s = \pm 1$ is the rolling direction, frozen over each integration step, which keeps the derivative function smooth inside the step. Multiplying through by κ gives the acceleration the code integrates. For a solid ball $J = \frac{2}{5}mR^2$, so $\kappa = 5/7$ on a flat track. On rails one radius apart $\rho = 11.00$ mm for the steel ball and $\kappa = 0.652$, and the hollow ping-pong ball with $J = \frac{2}{3}mR^2$ has $\rho = 18.97$ mm and $\kappa = 0.574$ on the same rails. The rails make the ball spin faster for the same travel, so more of gravity's work goes into the spin. Near rest $\ddot{r} \approx -\kappa g\sin\alpha$, which is 0.558 m/s^2 at 5° and 1.656 m/s^2 at the 15° limit. The term $r\dot{\alpha}^2$ flings the ball outward when the beam turns fast. The rolling resistance, $\alpha_k = 0.05°$ or a coefficient of 0.00087, decelerates a rolling steel ball at 5.6 mm/s^2.

5.3.2 The Breakaway and BALL OFF

Between integration steps a latch decides whether the ball rolls. With the drive $a_d = \kappa\big(r\dot{\alpha}^2 - g\sin(\alpha+\beta)\big)$, a rolling ball stops when it is slower than the creep speed $v_c = 2$ mm/s, or its speed changes sign inside a step, while $|a_d| \le \kappa g\sin\alpha_0$, and a ball at rest starts only when $|a_d| > \kappa g\sin\alpha_0$, as shown in Equation (5.2).

$$\text{stuck} \;\Leftrightarrow\; |\dot{r}| < v_c \text{ and } |a_d| \le \kappa g\sin\alpha_0, \qquad \alpha_0 = 0.4° \tag{5.2}$$

A stuck ball has $\dot{r} = 0$ and $\ddot{r} = 0$. The breakaway α_0 is the dent, seam or speck of dirt on a real rail. It decides where a PD law leaves the ball, within about α_0/K_c of its setpoint, which is 1.33 cm at 0.30 °/cm. When $|r|$ passes $L/2 = 50$ cm the ball has left the beam. The instant is found by interpolating back inside the step, $t_{off} = t_k - \Delta t\,(|r_k| - L/2)/(|r_k| - |r_{k-1}|)$, the event is marked on the trends, the ball falls into the tray in 0.3 s on the drawing, and the run stops when it lands.

5.3.3 The Servo-Driven and the Torque-Driven Beam

The hobby servo is a stiff position loop of its own, seen through the lever as a first-order lag with a rate limit, as shown in Equation (5.3).

$$\dot{\alpha} = \text{clamp}\left(\frac{\alpha_{cmd} - \alpha}{\tau_s},\ \pm\dot{\alpha}_{max}\right), \qquad |\alpha_{cmd}| \le \alpha_{max} \tag{5.3}$$

where $\tau_s = 60\,\text{ms}$, $\dot{\alpha}_{max} = 150^\circ/\text{s}$ and $\alpha_{max} = 15^\circ$. The lever only rescales the servo's angle, so it is drawn and not modeled, and the ball's weight does not move the beam. The advanced torque-driven beam replaces the servo by a motor at the pivot, and the beam then feels the ball, as shown in Equation (5.4).

$$\begin{aligned} (J_b + mr^2)\,\ddot{\alpha} &= \tau - mgr\cos\alpha - 2mr\dot{r}\dot{\alpha} - b_b\dot{\alpha} \\ \tau &= \text{clamp}\big(K_p(\alpha_{cmd} - \alpha) - K_d\dot{\alpha} + \hat{m}gy\cos\alpha,\ \pm\tau_{max}\big) \end{aligned} \tag{5.4}$$

where $J_b = 0.042\,\text{kg}\,\text{m}^2$, $b_b = 0.01\,\text{N}\,\text{m}\,\text{s/rad}$, $\tau_{max} = 3\,\text{N}\,\text{m}$, and the inner proportional-derivative loop runs every millisecond with $K_p = J_b\omega_{in}^2$ and $K_d = \max(0,\ 2\zeta_{in}\omega_{in}J_b - b_b)$ from the bandwidth ω_{in} and damping ζ_{in} the reader sets. Its last term feeds forward the declared ball's weight on the measured position. The inner loop reads the beam angle and rate exactly, as an encoder and a tachometer fine enough to treat as exact would, and the heavier-ball disturbance is the mass $m - \hat{m}$ that the feedforward does not know.

5.3.4 What the Kit Reads

The wire reads the ball once every control interval $T_s = 10\,\text{ms}$ with no delay, and the camera catches a frame every $1/f_c$ on its own clock and reports it θ_c later, as shown in Equation (5.5).

$$\begin{aligned} y_k &= \text{clip}\Big(q_w\,\text{rnd}\Big(\frac{r(t_k) + \sigma_w n_k}{q_w}\Big),\ \pm x_w\Big) \\ y_{cam}(t) &= q_c\,\text{rnd}\Big(\frac{r(t_f) + \sigma_c n_f}{q_c}\Big), \quad t_f + \theta_c \le t \end{aligned} \tag{5.5}$$

where rnd rounds to the nearest integer, $q_w = 0.5\,\text{mm}$, $\sigma_w = 1\,\text{mm}$, $x_w = 45\,\text{cm}$, t_f is the newest frame on the grid k/f_c that has arrived, $f_c = 30\,\text{Hz}$, $\theta_c = 33\,\text{ms}$, $q_c = 2\,\text{mm}$ and $\sigma_c = 0.5\,\text{mm}$. A camera reading is therefore 33 to 66 ms old. The wire draws its unit normal deviates n_k from the engine's generator, seeded at 20261019, and the camera draws n_f from a stream of its own seeded from the same number, so switching sensors or changing one noise level leaves the other's noise as it was. The servo's angle readback $\hat{\alpha}$ is rounded to 0.01°. A sensor dropout freezes the reading for one second.

> ✦ DID YOU KNOW? **A Pencil Balanced by Event Cameras**
>
> Event cameras report each pixel's brightness change as it happens, not whole frames. In an early demonstration two of them watched a pencil standing on a small movable cart, its state was estimated with a latency below 1 ms, and a simple hand-tuned PID controller kept it upright [18]. The kit's camera reports a frame 33 to 66 ms old.

5.3.5 Integration, Starting Point and Replay

Table 5.2 lists the constants. A fourth-order Runge–Kutta step of 1 ms advances the four states, the controller is invoked every millisecond but executes its law only when a full T_s has passed, and the trends record at 50 Hz. Each step follows the order of Section 1.6.1, with the inner loop of the torque beam run beside the controller and the breakaway latch and the BALL OFF test applied after the integration. A case starts with the ball at rest at its starting position, on a level beam in automatic mode or at the manual angle in manual mode, and a run plays at 0.1 to 5 times real time. Reset and the reload button act as Section 1.6.1 describes. The application passes its 51 verification checks, among them $\kappa = 5/7$ flat and 0.652 on rails one radius apart, conservation of $\frac{1}{2}\dot{r}^2/\kappa + gr\sin\alpha$ to 10^{-9} with the beam still and no rolling resistance, a ball that never starts below the breakaway, the P-only period on an ideal servo within 2 % of $2\pi/\sqrt{K_c\kappa g}$, and, with the noise off, halving the step moving the BALL OFF instant of case 2 and the settling time of case 3 by less than 10 ms.

Table 5.2 Model Constants. H Is Hidden, A Adjustable in the Model Dialog

Symbol	Meaning	Value	Range	
L, x_w	beam length, wire range	1.0 m, ±0.45 m	fixed	H
g	gravity	9.81 m/s^2	fixed	H
m, R	steel ball, solid	66 g, 12.7 mm	fixed	H
m, R	ping-pong ball, hollow	2.7 g, 20 mm	fixed	H
d	rail spacing	12.7 mm	0 to 24	A
α_k	rolling resistance	0.05°	0 to 0.3	A
α_0	breakaway angle	0.4°	0 to 1.5	A
v_c	creep speed	2 mm/s	fixed	H
τ_s	servo lag	60 ms	5 to 300	A
$\dot{\alpha}_{max}$	servo rate limit	150 °/s	20 to 600	A
α_{max}	beam angle limit	±15°	1 to 20	A
J_b	beam inertia (torque beam)	0.042 kg m^2	0.01 to 0.2	A
b_b	pivot damping	0.01 N m s/rad	0 to 0.2	A
τ_{max}	motor torque limit	3 N m	0.5 to 10	A
σ_w, q_w	wire noise, quantization	1 mm, 0.5 mm	0–5, 0.1–5	A
f_c, θ_c	camera rate, latency	30 Hz, 33 ms	10–120, 0–200	A
q_c, σ_c	camera pixel, noise	2 mm, 0.5 mm	0.2–10, 0–5	A
κ	rolling factor, steel, ping-pong	0.652, 0.574	derived	
ρ	rolling radius, steel, ping-pong	11.00 mm, 18.97 mm	derived	

5.3.6 Assumptions

One ball rolling without slip on a rigid, straight beam. Slip, bouncing, the ball's wobble between the rails and the beam's flexing are left out. A real ball skids when the beam is jerked faster than friction can spin it up, which would weaken the $r\dot{\alpha}^2$ fling of case 4.

A stiff servo. The servo is a first-order lag with a rate limit and is not moved by the ball. A real hobby servo has a dead band, gear backlash and a speed that depends on its load, and its lever makes the beam angle a nonlinear function of the horn angle.

Rolling resistance and breakaway as a latch. The values of α_k and α_0 are teaching values for a ball that presses on a resistance wire, and a clean steel ball on steel rails rolls with less. They decide where a PD law stops the ball and the size of the hunting in case 6.

Ideal instruments apart from noise, rounding and delay. The wire's noise is white at the control interval, without contact bounce or drift, and the camera's frames are exact snapshots with a fixed latency, without blur, dropped frames or jitter.

Teaching thresholds. The ±5 mm band, the 1 s that makes an arrival, the challenge's 15 legs in 60 s and the rig's limits are teaching values. A real kit's angle limit, servo speed and lag come from its servo's data sheet and its lever geometry. The torque-driven beam has no mechanical stop beyond its command limit.

5.4 Two Integrators and a Lag

5.4.1 The Linear Plant

About rest on a level beam, with the rate limit, the angle limit, the breakaway, the rolling resistance, the quantization and the centripetal term left out, the servo-driven kit from the angle command to the position is Equation (5.6).

$$P(s) = \frac{R(s)}{A_{cmd}(s)} = \frac{-\kappa g}{s^2(\tau_s s + 1)} \tag{5.6}$$

The double integrator contributes −180° of phase at every frequency, the servo takes more, and the sign is the reverse action of the loop. A proportional law $\alpha_{cmd} = K_c(y - r_{SP})$ on an ideal servo

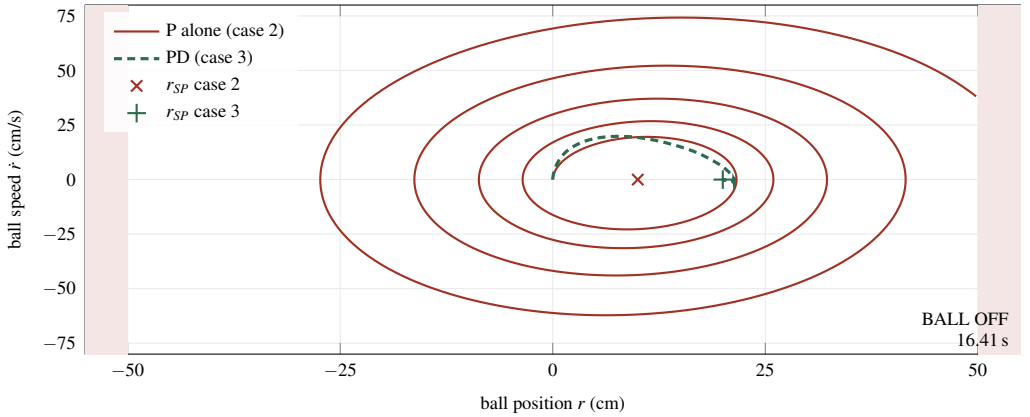


Figure 5.2 The ball's position against its speed from the engine, true values every 20 ms. Case 2 as shipped (solid line, P alone, 0.30 °/cm, setpoint 10 cm from 1 s, marked ×) spirals out at 1.37 to 1.42 a swing until the ball leaves the right end at 16.41 s. Case 3 as shipped (dashed line, PD, $\tau_D = 0.70$ s, setpoint 20 cm from 1 s, marked +) winds in from 1 s to 8 s and stops 2.5 mm past its setpoint. The shaded bands are past the ends of the beam.

makes the ball an undamped oscillator at $\omega = \sqrt{K_c \kappa g}$, a mass on a spring, with a period of 3.43 s at 0.30 °/cm (0.524 rad/m). The servo's lag and the hold of the sampled command act at that frequency as a delay of about $\tau_e = \tau_s + T_s/2 = 65$ ms, and a delay τ_e in $\ddot{r} = -\omega^2 r(t - \tau_e)$ moves the roots to $s \approx \omega^2 \tau_e/2 \pm j\omega$. Each swing therefore grows by $e^{\pi\omega\tau_e}$, which is 1.45 at 0.30 °/cm. The rolling resistance takes a fixed $4 \sin \alpha_k / K_c$ off each full swing, 6.7 mm at that gain, so the measured growth of 1.37 to 1.42 sits a little below the formula, and the sampled loop's least damping ratio is −0.058. The planning value for the kit was a gain of 0.1 rad/m. There the formula gives 1.18 a swing, but the rolling resistance takes 3.5 cm off every 7.9 s swing, and a 10 cm swing spirals in. The kit therefore ships with 0.30 °/cm, where the growth is visible inside its 20 s trend window.

Pause. Try Exercise 5.1 before reading on.

5.4.2 Derivative Is the Only Damping

A derivative term adds $K_c \tau_D \dot{r}$ to the angle, and on an ideal servo the loop becomes $s^2 + \kappa g K_c \tau_D s + \kappa g K_c$, with a damping ratio $\zeta = \omega\tau_D/2$ that the derivative time alone sets at a given gain. At 0.30 °/cm and $\tau_D = 0.70$ s that is 0.64. With the servo's lag the characteristic polynomial is $\tau_s s^3 + s^2 + \kappa g K_c \tau_D s + \kappa g K_c$, and the Routh test requires $\tau_D > \tau_s$ at every gain. This is the lag limit of Equation (1.9): the derivative time must be longer than the lag, and a proportional law alone fails it for any lag. The derivative is the phase lead the loop has, and the Loop view reads the result as a phase margin of 48.5° at 2.70 rad/s. The phase plane makes the two laws visible, as shown in Figure 5.2. Under P alone the orbit spirals out until it reaches the end of the beam, and under PD it winds into the target and stops there on the breakaway.

5.4.3 What the Sensor, the Delay and the Drive Cost

The derivative differences a reading that moves in 0.5 mm steps with 1 mm of noise. Over one 10 ms interval that noise is about 0.14 m/s of false speed, which a derivative time of 1 s at 0.30 °/cm turns into several degrees of command, so a filter is needed, and its lag is paid at the crossover. The camera is a delay. A frame is on average half a frame old when it is read, so the Loop view treats it as $\theta_c + 1/(2f_c) = 49.7$ ms, rounded to five control intervals, and a delay θ costs $\omega\theta$ of phase at the crossover and nothing in gain. A torque-driven beam with its inner loop closed is a second-order lag

Table 5.3 Controller Settings, Defaults and Ranges

Symbol	Setting	Default	Range
—	structure	PD	P, PD, PID, LQR4
K_c	proportional gain	0.30 °/cm	0 to 2
τ_I	integral time, PID only	1.0 s	0 to 30
τ_D	derivative time	0.70 s	0 to 2
τ_f	derivative filter	30 ms	0 to 300
T_s	control interval	10 ms	2 to 50
I_{max}	integral limit	±3°	0.2 to 15
a_p, v_p	profile acceleration, speed	0.8 m/s², 0.6 m/s	0.05–3, 0.05–1.5
A, P	relay marks, period of a leg	±30 cm, 6 s	5–45, 1–20
A_d, f_d	dither amplitude, frequency	±1.5°, 10 Hz	0–3, 1–50
$\omega_{\alpha\beta}$	α-β bandwidth	15 rad/s	2 to 60
ω_{in}, ζ_{in}	inner loop, torque beam	20 rad/s, 0.7	1–60, 0.2–1.5
r_s, α_s	LQR position, angle scales	5 cm, 3°	0.5–30, 0.5–15

where the servo was a first-order one, and its phase at the outer crossover comes out of the outer loop's margin.

5.4.4 A Limit No Tuning Overcomes

A ball moving at v toward an end stops in no less than $v^2/(2\kappa g \sin\alpha_{max})$, whatever law decides when to tilt. For the ping-pong ball on a beam limited to ±3° the largest deceleration is 0.295 m/s², so a ball flicked at 30 cm/s needs 15.3 cm to stop (15.0 cm with the rolling resistance helping), and case 9 gives it 10 cm. This is the actuator-headroom limit of Section 1.4.7 in the form it takes on a double integrator. It belongs to the plant and the actuator, and no controller removes it.

5.5 The Kit's Controller as Coded

The controller reads only the wire's or the camera's reading y, its age and a dropout flag, and the angle readback $\hat{\alpha}$, with the tachometer rate also supplied to the four-state law. It never reads the ball's true position or speed, which the CSV export logs as columns prefixed `truth_`. It executes once a whole T_s has passed on the 1 ms clock, writes α_{cmd} clamped to $\pm\alpha_{max}$, and holds it until the next execution. All laws are reverse acting, so a ball to the right of its reference asks for a positive angle. The settings are listed in Table 5.3.

5.5.1 The Shared PID

The PID is the shared velocity-form algorithm of Section 1.6.2, written in Equation (5.7) in the symbols of this loop.

$$\begin{aligned} e_k &= y_k - r_{ref,k}, \qquad P_k = K_c e_k \\ D_k &= K_c \tau_D \,(z_{f,k} - z_{f,k-1})/T_s \\ S_k &= S_{k-1} + (P_k - P_{k-1}) + (D_k - D_{k-1}) \\ \hat{u}_k &= b_k + S_k + I_{k-1} \\ \alpha_{cmd} &= \mathrm{clamp}(\hat{u}_k + \delta_k,\ \pm\alpha_{max}) \\ I_k &= \mathrm{clamp}\big(I_{k-1} + \big[K_i e_k + (u_k - \hat{u}_k)/\tau_I\big]T_s,\ \pm I_{max}\big) \end{aligned} \tag{5.7}$$

where z_f is the measured variable through the exact lag $z_{f,k} = a\,z_{f,k-1} + (1-a)z_k$ with $a = \mathrm{e}^{-T_s/\tau_f}$ ($a = 0$ when $\tau_f = 0$), $z = y - r_{ref}$ with the profile on and $z = y$ without it, $K_i = K_c/\tau_I$ for the PID, and the integral is updated only while $K_i \neq 0$, $u_k = \mathrm{clamp}(\hat{u}_k, \pm\alpha_{max})$ drives the back-calculation, δ_k is the dither, and b_k is the bias that carries the feedforward and the hand-over. Here $\hat{u}_k$ and α_{cmd} play

the parts of v_k and u_k in Equations (1.13) to (1.15), and the tracking time is τ_I. Without the profile r_{ref} is the setpoint, held constant between moves, so a setpoint step gives no derivative kick. Beyond the shared algorithm, the integral alone is clamped to $\pm I_{max}$. With the α-β option the PID's own derivative is off and $K_c\tau_D(\hat{\dot{r}} - v_{ref})$ enters through b_k. The saturation flag compares the unclamped law with $\pm\alpha_{max}$ and is cleared every execution.

In manual mode every law tracks the manual angle, so automatic mode starts without a bump, as in Section 1.6.2. Tracking leaves the manual angle less P and D in the integral, and a P or PD law would then hold that bias for good, which on a plant that rests only on a level beam keeps the ball off its setpoint. On the first automatic execution the part of the integral the law may not keep (all of it for P and PD, whatever lies beyond $\pm I_{max}$ for the PID) moves into b_k and decays as e^{-t/τ_h} with $\tau_h = 1$ s. When the ball of case 1 is caught by hand at −11.35 cm and the loop is switched to automatic at 6 s, it is 0.32 cm from its setpoint at 30 s under PD and 0.41 cm under PID.

5.5.2 Profile, Feedforward and Dither

With the profile on, a new setpoint is reached on a trapezoid, accelerating at a_p to the speed v_p and braking at a_p to stop on the target, and the feedforward is the angle that gives the profile's acceleration, as shown in Equation (5.8).

$$\alpha_{ff} = -\arcsin\frac{a_{ref}}{\hat{\kappa} g} \tag{5.8}$$

where $\hat{\kappa}$ is the rolling factor of the ball the controller is told is loaded. The generator runs $\mathrm{round}(\tau_s/T_s) = 6$ intervals ahead of the reference the PID tracks, so the beam is at the feedforward's angle when the ball needs it, and the PID acts on the tracking error $y - r_{ref}$. At 0.8 m/s^2 the feedforward is 7.18°. The breakaway dither adds a square wave of $\pm A_d$ to the command, its sign taken from the half-period index $\lfloor 2 f_d t + 10^{-6} \rfloor$ so that a 10 Hz wave on the 10 ms interval is exactly five intervals high and five low. A first-order servo at 60 ms passes $1/\sqrt{1+(2\pi f_d \tau_s)^2} = 0.256$ of the fundamental at 10 Hz to the beam.

5.5.3 Speed Estimate and State Feedback

The α-β estimator updates on every fresh reading, as shown in Equation (5.9).

$$\begin{aligned} \bar{r} &= \hat{r}_{k-1} + T_s\hat{\dot{r}}_{k-1}, \qquad \hat{r}_k = \bar{r} + a(y_k - \bar{r}) \\ \hat{\dot{r}}_k &= \hat{\dot{r}}_{k-1} + \frac{b}{T_s}(y_k - \bar{r}) \end{aligned} \tag{5.9}$$

with $a = 2\omega_{\alpha\beta}T_s$ and $b = (\omega_{\alpha\beta}T_s)^2$, the small-$\omega T$ form of a double pole, so 0.30 and 0.0225 at 15 rad/s and 10 ms, which place the estimate's error poles at 10.8 and 24.9 rad/s. The comparison law is a discrete linear quadratic regulator on the declared servo-beam model $[r, \dot{r}, \alpha]$, discretized for a zero-order hold at T_s and designed with Bryson's weights $Q = \mathrm{diag}(1/r_s^2, 0, 0)$ and $R = 1/\alpha_s^2$ by the shared Riccati module, as Section 1.4.5 describes. Chapter 7 designs a discrete regulator with the same rule on a cart and pole. It acts as $\alpha_{cmd} = \alpha_{ff} - K[y - r_{ref},\ \hat{\dot{r}} - v_{ref},\ \hat{\alpha} - \alpha_{ff}]$, and at the defaults $K = [-1.0283\ \mathrm{rad/m},\ -0.6254\ \mathrm{rad\,s/m},\ 0.2166]$. The beam rate is not a state of a first-order servo, so the law has three gains, and it reads the readback and the estimate, never the truth.

For the original laws, the Loop view writes each as a discrete state-space exactly as it runs, closes it around Equation (5.6) sampled at T_s, decides stability from the eigenvalues of that sampled closed loop, and takes the phase margin where the open-loop magnitude last drops through unity.

5.5.4 Four-State Feedback on the Torque Rig

The additional LQR4 uses $x = [r, v, \alpha, \omega]^T$ and commands the angle $u = \alpha_{cmd}$ through the existing inner motor loop. It is distinct from the three-state servo comparison above. Its local model assumes

Table 5.4 Four-State Design Used Throughout the Comparison

Setting	Value
Decision interval T_s	10 ms
Q diagonal inverse-square scales	3 cm, 0.3 m/s, 10°, 2 rad/s
R inverse-square command scale	5°
Observer bandwidth ω_o	12 rad/s
Inner bandwidth, damping	20 rad/s, 0.7
Adjustable position scale	0.5 to 30 cm
Adjustable command scale	0.5 to 15°
Adjustable observer bandwidth	2 to 40 rad/s

a level beam, small angles, nominal weight cancellation and negligible friction:

$$\begin{aligned} \dot{r} &= v, \qquad \dot{v} = -\kappa_n g_n \alpha, \\ \dot{\alpha} &= \omega, \quad \dot{\omega} = \omega_{in}^2 (u - \alpha) - 2\zeta_{in}\omega_{in}\omega. \end{aligned} \tag{5.10}$$

Declared steel geometry gives $\kappa_n = 0.652173913$ and $g_n = 9.81\,\mathrm{m/s^2}$. The selected ball and controller settings ω_{in}, ζ_{in} are known, while plant-parameter edits remain mismatch. The nonlinear plant and unchanged inner loop remain Equations (5.1) and (5.4).

For $\dot{x} = Ax + Bu$, the shared Riccati routines compute $\Phi = \exp(AT_s)$ and $\Gamma = \int_0^{T_s} \exp(A\tau)B\,d\tau$. For regulation at a fixed equilibrium, the discrete regulator minimizes $\sum_{k=0}^{\infty}(\tilde{x}_k^T Q \tilde{x}_k + \tilde{u}_k^2 R)$ on this unclipped model [19], where $\tilde{x} = x - [r_{ref}, v_{ref}, \alpha_{ff}, 0]^T$ and $\tilde{u} = u - \alpha_{ff}$. The initial design is retained across every case: The gain is $K = [-2.73608860, -1.20491006, 0.60986450, 0.02793657]$, with units rad/m, rad s/m, 1 and s, respectively. The applied command is

$$u_k = \mathrm{clamp}\big(\alpha_{ff,k} - K(\hat{x}_k - x_{ref,k}) + b_{h,k},\ \pm\alpha_{max}\big). \tag{5.11}$$

The model and gain are cached separately for each engine and recomputed on explicit design-setting changes. Profile feedforward uses nominal κ_n, g_n in Equation (5.8), with the arcsine argument clipped to $[-1, 1]$, and a lead of $\mathrm{round}(2\zeta_{in}/(\omega_{in}T_s))$ intervals. There is no future disturbance information or error integral.

The Luenberger observer [20] of Section 1.4.6 first predicts $\bar{x}_k = \Phi\hat{x}_{k-1} + \Gamma u_{k-1}$, then replaces its angle and rate by the current encoder and tachometer readings. For a fresh position captured at $t_c = t_k - d_k$, its residual is $\varepsilon_k = y_k - \bar{r}(t_c)$, where d_k is the reported age and $\bar{r}(t_c)$ is interpolated from estimated history. With h the larger of T_s and the time since the last corrected capture,

$$\begin{aligned} p &= \exp(-\omega_o h), \quad a = 1 - p^2, \quad b = (1-p)^2, \\ \hat{r}_k &= \bar{r}_k + (a + d_k b/h)\varepsilon_k, \qquad \hat{v}_k = \bar{v}_k + (b/h)\varepsilon_k. \end{aligned} \tag{5.12}$$

History after t_c receives the correction transported to each stored time. Held/dropout frames do not correct. A 1 s interpolation window is capped at 1024 entries. Entry resets the calendar and initializes $[y, 0, \hat{\alpha}, \hat{\omega}]^T$ from instruments. Frames predating activation wait for history warm-up, while dropout entry holds until a real reading. True ball velocity is never an input.

Nominal wire feedback poles are $0.97016824 \pm 0.02700964i$ and $0.83613899 \pm 0.10361177i$. Current-update observer poles are $\exp(-\omega_o T_s) = 0.88692044$ twice and zero twice. These results describe the frictionless, unsaturated wire model, not delayed-camera or nonlinear stability. The original frequency-margin plots are withheld under LQR4.

LQR4 never changes the physical rig. On the servo it visibly holds the previous command with zero activity. It takes precedence over the stored original LQR switch, which is restored on return to P, PD or PID. Matched PD and PID runs therefore explicitly turn that switch off. Manual tracking, entry, recovery and departure seed the first sampled applied command, with transient bias b_h washing

Table 5.5 The Ten Cases and the Challenge

#	Configuration and events	Teaching point
1	Manual, beam at 1°, ball at rest at 0	it always overshoots
2	P 0.30 °/cm; setpoint 0 to 10 cm at 1 s	a spring, and the lag
3	PD, τ_D 0.70 s, τ_f 30 ms; to 20 cm at 1 s	derivative damps
4	PD, τ_D 1.00 s, $\tau_f = 0$; to 20 cm at 1 s	noise against phase
5	quick PD 1.00 °/cm, 0.40 s; relay ±30 cm every 6 s	profile and feedforward
6	PD, beam bent $\beta = 0.5°$	offset, hunting, dither
7	quick PD on the camera; to 20 cm at 1 s	phase the delay eats
8	quick PD, torque beam ω_{in} 20 rad/s; to 20 cm at 1 s	inner loop faster
9	ping-pong ball, camera, ±3°, at 40 cm; flick 30 cm/s at 1 s	infeasible by arithmetic
10	LQR4, torque beam; to 20 cm at 1 s, flick 20 cm/s at 8 s, bend 0.5° at 16 s	estimation, travel and offset
11	P 0.30 °/cm; relay ±35 cm every 4 s for 60 s; flick 20 cm/s at 20 s; $\beta = 0.5°$ from 35 s	the relay race

out over 1 s. Startup in AUTO instead computes feedback at $t = 0$. Invalid instruments, design or calendar reject an update, hold a finite bounded applied command, mark the hold and visibly rearm on valid data. This adds no error integral. Existing angle-command and applied-torque caps remain, without a mechanical angle stop.

5.6 Eleven Cases, Measured

The cases appear in Table 5.5. The numbers in this section come from headless runs of the engine the browser uses, on the default seed, with each case either unchanged or carrying the single change a sentence names entered before Run. Overshoot is measured on the true position past the new setpoint, and settling time from the setpoint change until the ball stays within ±5 mm. Each case produces the same run when it is the first one opened, when it follows another case from the menu, and after Reset.

5.6.1 By Hand and Proportional Alone

As shipped, case 1 starts the beam at 1° with the ball at rest in the middle, and the ball leaves the left end at 3.07 s at 32.6 cm/s. Leveling the beam at 1 s, when the ball is 5.3 cm left of center at 10.6 cm/s, only slows the escape to 5.45 s. A tilt of −1° from 1 to 2 s and then a level beam stops the ball 11.35 cm left of center. Case 2 moves the setpoint of the P-only loop to 10 cm at 1 s. The ball crosses it and comes back with periods of 3.46 to 3.48 s between swings on the same side, against 3.43 s for the ideal servo, swings 11.6, 16.0 and 22.3 cm past the target on the right, and leaves the beam at 16.41 s at 37.5 cm/s. The Loop view gives a phase margin of −6.8°. With a 5 ms servo the swings shrink from 10.0 to 9.4 cm over 20 s, slowed only by the rolling resistance. Halving the gain to 0.15 °/cm lengthens the period to 4.88 s and slows the growth to 1.16 a swing, rising to 1.25 as the swings widen and the fixed loss to the rolling resistance becomes a smaller share, so the ball lasts until 37.56 s. Doubling it to 0.60 °/cm gives 2.48 s and BALL OFF at 9.94 s. A higher gain makes the same lag a larger share of a shorter period.

> **TRY IT IN THE APP** **2 · Proportional Alone: Spring and Lag**
>
> Load the scenario and press Run, and watch the Phase view. After the setpoint moves to 10 cm, the orbit of position against speed spirals outward: each swing past the target is 1.37 to 1.42 times the last, about 3.47 s apart, until the ball leaves the beam at 16.41 s.

5.6.2 Derivative, and the Sensor That Limits It

With $\tau_D = 0.70$ s the 20 cm step of case 3 overshoots by 1.31 cm, and the ball rests there for a while, as listed in Table 5.6, and Figure 5.3 draws three of its columns. A 1.31 cm error asks for 0.39° at this gain, just inside the 0.4° breakaway, so the ball waits until the wire's noise tips the beam over it, and it is inside the band to stay 4.09 s after the step. Longer derivative times raise the margin

Table 5.6 Loop Margins and Step Responses on the 20 cm Steps. PM and ω_c From the Loop View, Overshoot and Settling From the Run

Case	Setting	PM (°)	ω_c (rad/s)	Overshoot (cm)	Settling (s)
3	τ_D 0.30 s	22.5	1.99	9.66	—[a]
	τ_D 0.50 s	38.0	2.28	4.47	—[a]
	τ_D 0.70 s (shipped)	48.5	2.70	1.31	4.09
	τ_D 1.00 s	55.6	3.49	0.03	2.62
4	$\tau_f = 0$ (shipped)	60.2	3.43	9.43	—[b]
	τ_f 30 ms	55.6	3.49	0.03	2.62
	τ_f 100 ms	43.2	3.51	0.00	3.42
	τ_f 200 ms	29.7	3.37	0.00	—[a]
7	wire	38.3	5.01	1.28	1.44
	camera (shipped)	24.0	5.01	5.75	1.69
	camera, 0.60 °/cm, 0.50 s	32.1	3.83	2.90	1.72
8	ω_{in} 20 rad/s (shipped)	34.2	5.20	1.23	1.31
	ω_{in} 15 rad/s	26.7	5.18	2.51	1.25
	ω_{in} 10 rad/s	11.4	5.08	7.79	3.97
	ω_{in} 6.7 rad/s	−9.1	4.76	17.1	—[c]

[a] Stops on the breakaway outside the band: 6.3 mm, 7.2 mm and 7.6 mm short. [b] Flung 66 mm to 94 mm beyond the setpoint from 8 s on. [c] Swings between 10.6 cm short and 14.7 cm past for good.

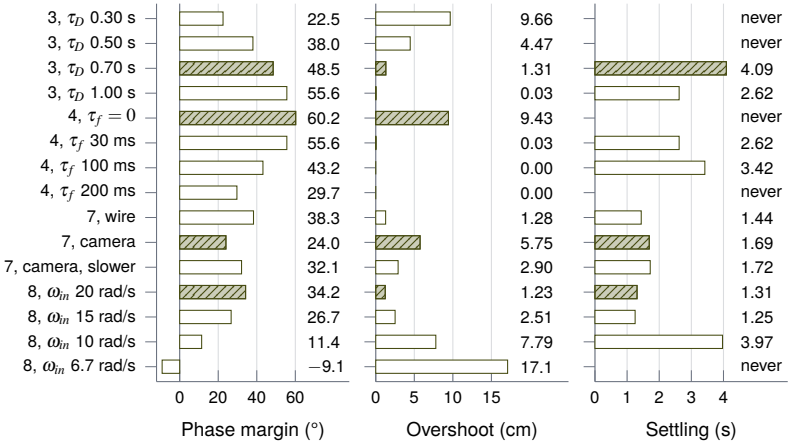


Figure 5.3 The 20 cm steps of Table 5.6: the phase margin from the Loop view (left), the overshoot (middle) and the settling time (right) of each run. Hatched bars are the setting each case ships with and white bars its variants, and "never" marks the five runs that never stay in the ±5 mm band. Notice that the unfiltered derivative of case 4 has the largest margin of all and still never settles, and that the inner loop of case 8 loses its margin while its crossover hardly moves.

and the crossover together, and at 1.00 s the overshoot is 0.3 mm. Shorter ones leave the ball on the breakaway outside the band.

Case 4 runs $\tau_D = 1.00$ s with the filter off. The Loop view rates it best, at 60.2°, because the margin describes small, slow motions. In the run the command jumps 7.43° RMS from one interval to the next, the servo spends 0.56 s of the first 15 s at its rate limit, and the beam shakes at 67 °/s RMS. That shaking flings the ball outward through $r\dot{\alpha}^2$ until a mean tilt of 2.22° holds it, against $r\langle\dot{\alpha}^2\rangle/g = 2.22°$, so the ball sits 66 to 94 mm beyond its setpoint. A 10 ms filter cuts the chatter to 3.79° and still leaves the ball 18 to 24 mm out. At 30 ms the chatter is 1.45° and the ball settles in 2.62 s for a margin of 55.6°. At 200 ms the chatter falls to 0.30° but the margin has fallen to 29.7° and the ball stops 7.6 mm short. The filter buys quiet with phase.

🕮 **Pause.** Try Exercise 5.2 before reading on.

Table 5.7 Matched Torque-Rig Runs, Seed 20261019. Each Group Lists PD, PID and LQR4 in That Order

	IAE (m s)			Logged travel (deg)			Last 5 s error (mm)			Motor cap time (s)		
Case	PD	PID	LQR4	PD	PID	LQR4	PD	PID	LQR4	PD	PID	LQR4
Step	0.09956	0.13432	0.12204	2386.6	2370.7	243.4	0.30	-0.33	0.84	0.017	0.017	0.016
Flick only	0.02671	0.04749	0.05169	2389.1	2377.4	221.4	0.16	-0.18	-0.59	0.000	0.000	0.000
New lesson	0.17922	0.17201	0.25993	2387.2	2388.9	255.1	-4.34	-0.14	-8.63	0.017	0.017	0.016
Constant bend	0.45983	0.14587	0.26433	1198.6	1204.5	210.0	-15.70	-0.50	-8.96	0.000	0.000	0.000
Mass ×3	0.35549	0.14704	0.28936	2373.1	2363.2	244.9	9.59	-0.23	6.86	0.017	0.017	0.016
Camera step	0.11724	0.15234	0.11383	816.0	875.4	229.3	0.44	0.21	0.37	0.023	0.029	0.016
Camera + dropout	0.42825	0.46312	0.42073	1050.8	1285.5	235.1	0.18	0.21	-0.02	0.076	0.091	0.040
Motor cap 0.5 N m	0.23071	0.22183	0.30677	2373.2	2364.4	275.6	-3.29	0.10	-8.67	7.738	7.813	0.780
Limited angle	0.04797	0.04797	0.04797	3.0	3.0	3.1	—	—	—	0.000	0.000	0.000
Torque race, base tuning	10.27726	0.67968	8.79947	2492.8	63.6	1118.3	-135.41	—	-120.35	0.338	0.022	0.339
Torque race, suggested tuning	10.79869	10.75737	10.99966	4011.0	4023.0	1440.9	-145.59	-141.30	-151.35	0.000	0.000	0.324

Requested windows are 30 s, or 60 s for race variants. All limited-angle runs exit the rail at 1.353 s and stop at 1.654 s. Base-tuning race PID exits at 1.969 s and stops at 2.269 s. Their IAE and travel use these actual short windows, and their final errors are undefined. The camera/dropout variant has the step at 1 s, flick at 8 s and a one-second hold from 8.05 s, without the bend. Mass mismatch triples physical ball mass. The motor-cap case keeps the new lesson events. Limited-angle runs retain the ping-pong ball, camera, 3 deg command cap and 30 cm/s flick. Torque races explicitly change the original physical rig and are not the graded servo challenge.

5.6.3 The Relay, the Bent Beam and the Camera

Case 5 sends the setpoint between −30 cm and +30 cm every 6 s under a quicker PD. A 60 cm step asks the proportional term for 60°, four times the limit, so every leg pins the beam at ±15°, for 3.93 s of the first 24 s in all, and the legs overshoot by 4.9 to 5.2 cm and settle in 2.08 to 2.10 s. With the profile and its feedforward the overshoot is 0.6 to 0.8 cm, settling takes 2.16 to 2.38 s, and the beam never reaches its limit. All four legs arrive within ±5 mm both ways, but only the profiled loop has angle to spare.

> **TRY IT IN THE APP** **5 · The Relay: Back and Forth on Purpose**
>
> Load the scenario and press Run. Every 60 cm leg pins the beam at ±15°, and the legs overshoot by 4.9 to 5.2 cm. Press Reset, turn Profile and feedforward on and run again: the overshoot falls to 0.6 to 0.8 cm, and the beam never reaches its limit.

In case 6 the beam is bent 0.5° and the PD stops the ball 15.8 mm short of its setpoint, where the bend and the proportional tilt leave less than the breakaway. Switched to PID with $\tau_I = 1$ s, the integral tilts the beam until the ball breaks loose, but by then it has tilted past what holding the ball needs, so the ball rolls through the setpoint and sticks on the other side. The stops alternate sides, 4.7 to 7.9 mm from the setpoint, every 2.2 to 4.0 s, and the HUNTING status, three alternating stops within 15 s, comes on at 7.18 s. The 10 Hz dither keeps the ball from sticking for long, and it then drifts slowly within ±4.0 mm without another hunting alarm. A dither of ±0.5° is too small, and HUNTING returns at 13.28 s. At 20 Hz the servo passes about half as much of the wave, and the ball again holds on alternate sides 5.1 to 9.8 mm from the setpoint, although its edges unstick the ball often enough that the detector stays quiet.

Case 7 puts the quick PD on the camera. At the 5.01 rad/s crossover the 50 ms delay is 0.25 rad, and the margin falls from 38.3° on the wire to 24.0° on the camera, a loss of 14.4°. The 20 cm step then overshoots by 5.75 cm instead of 1.28 cm. More derivative is no cure, because the camera's margin peaks at 24.7° at $\tau_D = 0.45$ s and falls again as the crossover moves up into the delay. A slower loop, 0.60 °/cm and 0.50 s, recovers 32.1°.

5.6.4 The Cascade and the Limit

On the torque-driven beam of case 8 the quick PD crosses over at 5.20 rad/s with 34.2° of margin when the inner loop has a bandwidth of 20 rad/s, and the step settles in 1.31 s. The outer loop is unstable for inner bandwidths below 7.93 rad/s, 1.53 times its crossover. At a third of the shipped

Table 5.8 The Relay Race Over 60 s, Default Seed

Configuration	Arrivals	Note
P 0.30 °/cm (shipped)	0	off 1.19 s
PID 1.30 °/cm, τ_I 3 s, τ_D 0.30 s, τ_f 40 ms, profile (suggested)	15	passes
suggested without the integral (PD)	12	legs 10, 12, 14
suggested without the profile	0	off 1.26 s
suggested with the filter off	0	7.90 s at limit
case 3's PD on the profile	0	—
LQR on the profile, r_s 2 cm, α_s 5°	15	—

bandwidth, 6.7 rad/s, the margin is −9.1° and the ball swings between 10.6 cm short of its setpoint and 14.7 cm past it for as long as the run lasts, with the command at its limit for 11.0 s of the 20 s. The rule of thumb that an inner loop should be three to five times faster than the outer one is the margin kept above that 1.53. Volume 3, Chapter 3 measures the same rule on a fermenter, and Volume 3, Chapter 10 closes a position loop that, like this one, commands an angle.

> **✦ DID YOU KNOW? Faster Than the Crossover, and Still Unstable**
>
> From the simulator: on the torque-driven beam of case 8, an inner bandwidth of 6.7 rad/s is still faster than the outer loop's crossover of 4.76 rad/s. Yet the phase margin is −9.1°, the ball overshoots by 17.1 cm and it never settles (Table 5.6).

Case 9 flicks the ping-pong ball toward the near end at 30 cm/s with the beam held to ±3°, and the ball leaves at 1.36 s. It does so in all 145 configurations of a verification sweep, P, PD and PID from 0.3 to 10 °/cm with derivative times of 0, 0.3 and 1 s, other integral times, the α-β speed, faster sampling, the state feedback at twenty weightings on both sensors and a bang-bang law on the wire, with BALL OFF between 1.27 and 1.41 s. The 15.3 cm stopping distance of Section 5.4 decides it before any controller is chosen. Given ±10°, the state feedback on the wire with $r_s = 1$ cm and $\alpha_s = 10°$ stops the ball at 47.9 cm.

🕮 **Pause.** Try Exercise 5.3 before reading on.

5.6.5 The Relay Race

The challenge carries the ball between marks at ±35 cm, the command alternating every 4 s for 60 s, through a 20 cm/s flick at 20 s and a 0.5° bend from 35 s. A leg counts as an arrival when the ball has been within ±5 mm of its mark for the last second before the next command, judged on the true position every millisecond, and BALL OFF scores zero. The configurations are compared in Table 5.8. The shipped P-only loop pins the beam at 15° on the first 70 cm step and loses the ball at 1.19 s. The suggested loop arrives on all 15 legs, each inside the band 2.48 to 2.61 s after its command with a worst error of 1.5 to 3.4 mm over its last second, and it scores 15 at the default seed and seeds 1 to 50, and at 21 seeds with the step halved, while the shipped loop scores 0 at all of them. Every part of the suggested setting is necessary except the step from the shipped 30 ms filter to 40 ms, which also scores 15 at eleven seeds. The race is judged as the case ships it, so changing the relay, the rig, a model constant or a disturbance voids an attempt. At the fastest speed, 5, the 60 s race plays in 12 s.

Without the integral the loop misses legs 10, 12 and 14, three of the legs to the left mark after the bend. Back at 0.30 °/cm, at $\tau_D = 0.70$ s or at $\tau_I = 1$ s, no leg arrives at eleven seeds, apart from one leg at one seed for the short integral time. Case 3's gentle PD fails on the profile because its tilt is inside the breakaway over most of the last centimeter. The P-only swing of case 2 just before the ball leaves the beam is shown in Figure 5.4.



Figure 5.4 Proportional control alone, case 2 on the Phase view, captured at $t = 16.30$ s, 0.11 s before the ball leaves the beam (the header that prints the time is cut). The picture is a composite of two parts of the screen: the Phase view above and the trends below. The badge row (1) reads the ball, the setpoint, the beam and the speed estimate, and calls the loop's phase margin unstable. The orbit (2) of position against speed has spiraled out from the 10 cm setpoint, and the swing panel (3) reads a last swing of −37.3 cm after one of −26.3 cm, a growth of $\times 1.42$ and a period of 3.48 s, with the loop at rest unstable. The ball is at 45.4 cm (the wire reads 44.9 cm) moving right at 49.4 cm/s, which the α-β estimate puts at 60.9 cm/s, and the beam is at 9.55° on its way to the commanded 10.46°. The 10 s trends show the swings growing in position about the dashed setpoint (4), in angle and in speed, where the estimate (5) runs beside the true speed.

5.6.6 Observer LQR Against PD and PID

Case 10, Four states, one gain vector, combines a 20 cm step at 1 s, a 20 cm/s flick at 8 s and a 0.5° bend from 16 s. Table 5.7 matches torque rig, inner settings, instruments, noise, seed, initial state and events. PD/PID keep the source case's tuning, or the named suggested tuning, with the old LQR switch off. LQR4 retains Table 5.4. Scorecard IAE uses raw meters. Logged command travel sums 50 Hz samples, not all 100 Hz moves or physical wear. Final error averages true $r - r_{SP}$ over the last 5 s, and motor-cap time counts requested torque beyond its limit.

In the new lesson LQR4 reduces logged travel from PD's 2387.247° to 255.059°, but increases IAE from 0.1792172 to 0.259 932 2 m s and leaves −8.629 mm mean error instead of −4.344 mm. PID leaves −0.139 mm with IAE 0.172 012 2 m s. At 24 s both balls are stationary, at 19.561 cm under PD and 19.137 cm under LQR4, while the latter's observer estimates 0.931 cm/s. An unknown input and friction can bias the nominal velocity estimate. Static feedback, estimation, chosen weights and the breakaway all affect the outcome, so this offset is not attributed to one cause. The paired trajectories

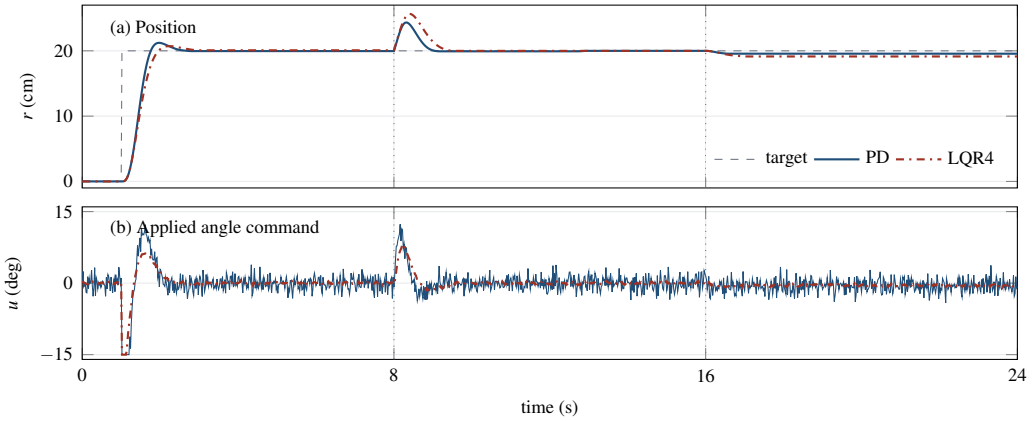


Figure 5.5 Common torque-rig lesson under PD (solid) and LQR4 (dash-dotted), seed 20261019, with the target as a thin dashed line. The step at 1 s, flick at 8 s and bend at 16 s give less logged command motion under LQR4 but a larger final position error. Every plotted point is a 50 Hz historian sample through 24 s. These traces have a shorter window than the 30 s table, and position is simulation truth.

are shown in Figure 5.5.

A fixed analytical diagnostic eliminates static angle feedback to give $K_c = -K_1/(1 + K_3) = 1.699\,576\,95\,\text{rad/m}$ and $\tau_D = K_2/K_1 = 0.440\,376\,84\,\text{s}$. With the original α-β estimator at 12 rad/s, it gives lesson IAE 0.346 513 0 m s, travel 495.657° and mean error −4.495 mm. This algebraic diagnostic still has a different estimator and is not an optimized or stronger PD comparator.

The step and push favor PD's IAE, and PID leaves smaller bend/mass errors. With a 0.5 N m motor cap LQR4 saturates less but has larger IAE and a true angle peak of 18.841° despite a 15° command cap. All three laws lose the limited-angle ball. Failed metrics end at the actual stop, including the 0.3 s fall, without final-window error or settling. Their shorter IAE window is not a win.

All 44 torque comparisons, including the analytical diagnostic, repeat at half the integration step. The lesson's travel/error tradeoff persists, but the camera/dropout ranking reverses. At 1 ms, PD/LQR4 IAE is 0.428 246 1 m s and 0.420 727 6 m s. At 0.5 ms, it is 0.421 755 8 m s and 0.441 870 6 m s. A robust camera advantage is not established. Torque races are distinct from the original graded servo race, which retains scores 0 shipped and 15 suggested. LQR4 on that unchanged servo holds, has zero activity and scores 0.

The eleven torque LQR4 runs have zero numerical failures. A browser measurement recorded when the law was added gives 7.792 μs per update at 1× and 3.002 μs at 5×, or 0.0779% and 0.1498% of one core, and the initial design costs 2.2 and 1.5 ms. These timings depend on the host and include timer quantization and JIT warm-up.

5.7 Technology in Practice

What Measures It

The one controlled variable is the ball's position along the beam, and no instrument on the kit measures its speed. The resistance wire makes the ball part of a voltage divider, read once every control interval. The camera above the beam is the other route. A frame camera delivers pictures at a fixed rate and must wait for a global exposure time, and motion blur is typical of it on fast motion [18]. The chapter's camera is a clean 33 ms delay on a 30 Hz frame clock, without blur, dropped frames or jitter, so the 14.4° it costs case 7 is a lower bound for a real one, as Section 5.8 states. The beam's own angle comes back from the servo's readback, rounded to 0.01°.

TECHNOLOGY Sensor: A Resistance Wire Between the Rails

On one laboratory ball and beam a high-resistance metal strip runs along the beam with a voltage across it, and the ball connects it to the other rod. The voltage sensed on that rod, from −5 V to +5 V over an effective length of 42 cm, gives the ball's position [8]. The same study found its signals noisy, most of all the resistive displacement sensor's, because of how they are read and wired. A numerical derivative of the noisy signals caused problems, so a filter or a state observer was required for the ball's speed [8]. The chapter's wire carries that noise as 1 mm of white noise with 0.5 mm rounding, and its dropout freezes the reading for a second.

What Moves It

The final element is a hobby servo that tilts the beam through a horn and a push rod. A research rig drives its beam instead with a direct-current motor through a gearbox of three simple gears and a lever arm, reads the gear's angle with a potentiometer over $\pm 180^\circ$, and relates the beam's angle to it through the lever's geometry, a sine relation that small angles make linear [8]. The chapter's lever only rescales the servo's angle, and its torque-driven beam, a motor at the pivot under an inner angle loop, is the kit's version of a motor-driven beam.

TECHNOLOGY Final Element: A Hobby Servo

A hobby servo is a position loop in a small case. A potentiometer inside senses the output shaft, the shaft position is compared with the desired one, and the amplified difference drives a small direct-current motor through gears that turn the shaft to remove the error [21]. Its command is a pulse 0.7 to 2.0 ms wide repeated every 20 ms, and the width sets the position of the output arm. One teaching laboratory removed the internal circuit so a microcontroller could close the loop itself, and on one small servo without end stops the potentiometer turns only about 210° before it slips on the shaft [22]. The chapter's servo is a 60 ms lag with a rate limit, and Section 5.3.6 names the dead band and backlash it leaves out.

Where the Controller Runs

One laboratory rig runs its laws on a desktop computer with analog-to-digital and digital-to-analog boards, and an interface board reads the sensors and applies the motor voltage [8]. A low-cost teaching rig puts the loop on a microcontroller board that drives a servo's motor through two outputs used as an H-bridge, with pulse-width modulation at 490 Hz, while an unmodified servo takes a new command pulse every 20 ms [22]. The chapter's kit runs its law every 10 ms on a 1 ms clock, and the inner loop of the torque-driven beam runs every millisecond. The velocity-form PID, with its profile, feedforward bias and dither, or the regulator that replaces it, is the one function that holds the chapter's structure. Figure 5.6 draws the loop on kit hardware.

What Is Modernizing It

First comes the event camera. Each pixel works on its own and reports a change of brightness as soon as it is detected, with a latency of about 10 µs on the laboratory bench and below a millisecond in the real world, and with less motion blur than a frame camera. New methods are needed to process that output [18]. It attacks the delay of case 7. The second is event-based control, in which the control signal changes asynchronously: an event-triggered controller recomputes when, for example, the tracking error exceeds a threshold, and a self-triggered one decides itself when to run next. The same survey states the principle behind both, that it makes no sense to change the control much faster than new information arrives or the actuators can move [18], the kit's servo and camera seen from the controller's side.

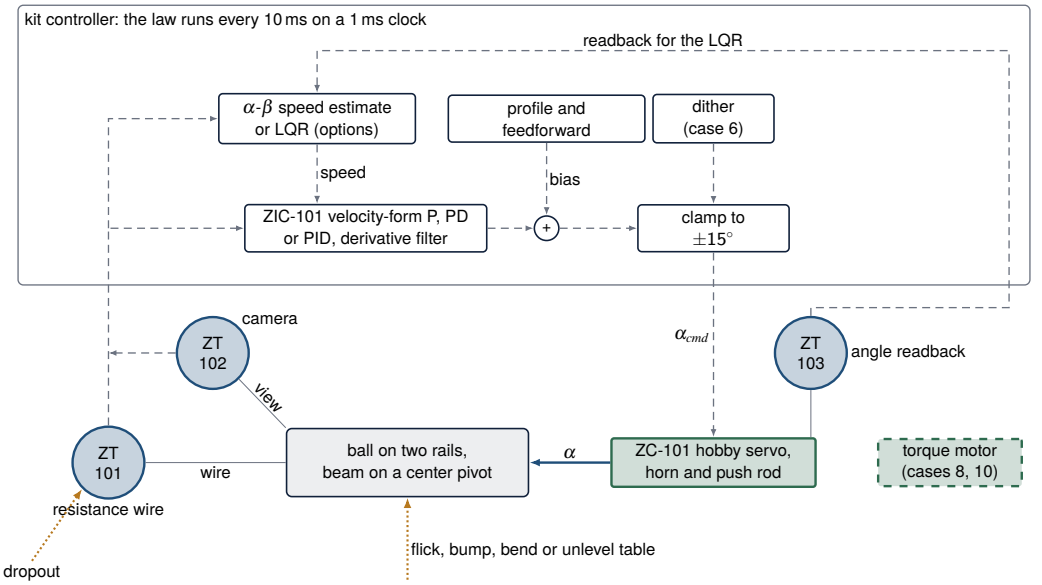


Figure 5.6 The position loop on kit hardware. The resistance wire ZT-101, or the camera ZT-102 that can replace it, reads the ball's position for the velocity-form controller ZIC-101, whose speed term comes from a filtered difference or from the α-β estimate. The profile's feedforward enters as a bias, the dither (case 6) rides on the command, and the clamped angle command goes to the hobby servo ZC-101, whose own position loop tilts the beam through a horn and a push rod; ZT-103 is its angle readback. The dashed torque motor replaces the servo on the torque-driven beam. Notice that nothing in the loop measures the ball's speed: every damping term has to be computed from the position reading.

> ⚠ **IN PRACTICE** **Filter the Derivative Before Tuning It**
>
> An ideal derivative has a high gain for high-frequency signals, so measurement noise makes large variations in the control signal. Practice replaces it with the derivative of a low-pass filtered signal, and the filter time constant is typically chosen as the derivative time divided by a factor in the range 5 to 20. Filtering the measured signal with a second-order filter instead also gives good high-frequency roll-off, which improves robustness [2]. Either way the filter is part of the tuning: in case 4 no filter shakes the beam, and a long one costs phase at the crossover.

5.8 Limitations

The kit is a teaching design whose sizes and limits make a consistent desktop rig without being fitted to any product. The ball rolls without slip on a rigid beam, so skidding, bouncing and beam flex are absent, and the servo is a lag with a rate limit that the ball's weight never moves. The wire's noise is white and Gaussian, and the camera is a clean delay without blur, dropped frames or jitter, so the costs measured in cases 4 and 7 are lower bounds for real sensors.

The linear analysis of the Loop view leaves out the rate limit, the angle limit, the breakaway and the centripetal term, which is why case 4's unfiltered loop reads as the best-damped one while its run is the worst. The camera enters that analysis as a whole number of control intervals. The torque-driven beam has no mechanical stop, and in case 8 detuned the beam reaches 16.44° while its command is held at 15°. The original servo state feedback is a comparison law, and no nonlinear law of the kind the literature built on this plant is implemented.

LQR4 has no integral or disturbance-bias state. Clipping, breakaway, unknown forces and delayed instruments are outside its nominal pole calculation. The observer can be biased, and delayed

initialization is approximate. Numerical hold bounds the applied command without ensuring ball survival. The camera/dropout refinement reversal precludes a robust IAE advantage.

OPEN CHALLENGE **Stabilizing With an Actuator That Runs Out**

A study of the ball and beam under a limited motor voltage judged the control of underactuated systems, with fewer controls than configuration variables, far from closed: design methods do not exist for many that matter in applications. When the linearized plant has unstable modes and the control is bounded, only initial states inside a controllability domain can be brought back, and the basin of attraction of any saturated feedback lies inside it. A law for the straight beam made the two coincide, one for a circular beam only approached it, both in simulation [23]. Progress would be such laws for more machines, tested on hardware. In case 9 the stopping distance decides before a controller is chosen.

5.9 Conclusion

A ball on a beam is two integrators with nothing to slow them, and this simulator models it as a desktop kit with rolling on rails, a breakaway, a lagging servo, a quantized wire or a delayed camera, and an optional torque-driven beam. Proportional control alone swings the ball past its target with a period the gain sets, 3.47 s at 0.30 °/cm, and the servo's lag makes each swing 1.37 to 1.42 times the last, close to $e^{\pi\omega\tau_e}$, until the ball leaves the beam at 16.41 s. Derivative action supplies the damping, 48.5° of margin at the shipped setting, and the wire's noise limits it, since the unfiltered derivative shakes the beam hard enough to fling the ball 66 mm to 94 mm off its setpoint. The camera's delay costs 14.4° of margin, the cascade fails below an inner bandwidth of 1.53 times the outer crossover, integral action hunts across the breakaway until a dither calms it, and no tuning keeps a flicked ping-pong ball on a beam limited to $\pm 3^\circ$. A PID on a profile with its feedforward turns the relay race from 0 arrivals into 15. The four-state torque observer LQR lowers logged command travel in its lesson while losing IAE and bend-offset performance to PID. The friction model is a latch and the instruments are idealized. A fuller model would add a servo with backlash and a dead band, a camera with blur and jitter, and a ball that can skid. The kit shows that a loop holding a ball still has to know how fast the ball is moving.

KEY POINTS

- A ball on a beam is two integrators with nothing to slow them, so a level beam leaves it rolling.
- Proportional control alone swings the ball past its target, and the servo's lag makes each swing 1.37 to 1.42 times the last until the ball leaves the beam.
- Derivative action supplies the damping, and the wire's noise limits how much of it the loop can carry.
- The camera's delay costs 14.4° of margin, the cascade fails below an inner bandwidth of 1.53 times the outer crossover, and no tuning keeps a flicked ping-pong ball on a beam limited to $\pm 3^\circ$.

Where else this applies. Volume 2, Chapter 4 meets the same dead zone as a sticking valve whose limit cycle no tuning removes, and Volume 3, Chapter 3 measures the inner-loop speed of a cascade on a fermenter.

References

[1] V. G. Rao and D. S. Bernstein, "Naive control of the double integrator," *IEEE Control Systems*, vol. 21, no. 5, pp. 86–97, 2001.

[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] B. Armstrong-Hélouvry, P. Dupont, and C. Canudas de Wit, "A survey of models, analysis tools and compensation methods for the control of machines with friction," *Automatica*, vol. 30, no. 7, pp. 1083–1138, 1994.

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

[5] J. Hauser, S. Sastry, and P. Kokotović, "Nonlinear control via approximate input-output linearization: the ball and beam example," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 392–398, 1992.

[6] A. R. Teel, "Semi-global stabilization of the 'ball and beam' using 'output' feedback," in *1993 American Control Conference*, 1993, pp. 2577–2581.

[7] C. Barbu, R. Sepulchre, W. Lin, and P. V. Kokotović, "Global asymptotic stabilization of the ball-and-beam system," in *Proceedings of the 36th IEEE Conference on Decision and Control*, vol. 3, 1997, pp. 2351–2355.

[8] M. Keshmiri, A. F. Jahromi, A. Mohebbi, M. H. Amoozgar, and W.-F. Xie, "Modeling and control of ball and beam system using model based and non-model based control approaches," *International Journal on Smart Sensing and Intelligent Systems*, vol. 5, no. 1, pp. 14–35, 2012.

[9] J. L. Diez, M. Valles, and A. Valera, "Virtual and remote laboratory of the ball and beam system," in *2009 EAEEIE Annual Conference*, 2009, pp. 1–6.

[10] M. Guinaldo, L. de la Torre, R. Heradio, and S. Dormido, "A virtual and remote control laboratory in Moodle: The ball and beam system," *IFAC Proceedings Volumes*, vol. 46, no. 17, pp. 72–77, 2013.

[11] 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.

[12] G. Zames and N. Shneydor, "Dither in nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 5, pp. 660–667, 1976.

[13] P. E. Wellstead, "The ball and hoop system," *Automatica*, vol. 19, no. 4, pp. 401–406, 1983.

[14] J. Hauser, S. Sastry, and P. Kokotović, "Nonlinear control via approximate input-output linearization: The ball and beam example," Electronics Research Laboratory, University of California, Berkeley, Memorandum UCB/ERL M89/95, 1989. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/1989/1300.html

[15] W. Yu and F. Ortiz, "Stability analysis of PD regulation for ball and beam system," in *Proceedings of 2005 IEEE Conference on Control Applications (CCA 2005)*, 2005, pp. 517–522.

[16] J. Park, R. A. Martin, J. D. Kelly, and J. D. Hedengren, "Benchmark temperature microcontroller for process dynamics and control," *Computers & Chemical Engineering*, vol. 135, p. 106736, 2020.

[17] T. Hägglund, "A friction compensator for pneumatic control valves," *Journal of Process Control*, vol. 12, no. 8, pp. 897–904, 2002.

[18] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.

[19] R. E. Kalman, "When is a linear control system optimal?" *Journal of Basic Engineering*, vol. 86, no. 1, pp. 51–60, 1964.

[20] D. G. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.

[21] Electrical & Computer Engineering Department, University of Utah, "ECE 1250 lab 5: Servo (control system)," Laboratory handout. [Online]. Available: https://ctools.ece.utah.edu/ECE1250/ECE1250_Lab05_ServoNew.pdf

[22] C. Cohenour, "A low-cost control system experiment for engineering technology students," in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio, 2017.

[23] Y. Aoustin and A. Formal'skii, "Ball on a beam: stabilization under saturated input control with large basin of attraction," *Multibody System Dynamics*, vol. 21, no. 1, pp. 71–89, 2009.

5.10 Exercises

5.1 ○ Concept check · 2 min

Case 2 runs a proportional law alone at 0.30 °/cm. After the setpoint step each swing past the target is 1.37 to 1.42 times the one before it on the same side, until the ball leaves the beam. What feeds the growth?

(a) The rolling resistance, which pushes the ball the way it is already rolling

(b) The wire's noise, which the proportional term passes straight to the beam

(c) The servo's lag and the held command, which apply each tilt about 65 ms late

(d) The centripetal term $r\dot{\alpha}^2$, which flings the ball outward on every swing

5.2 Concept check · 2 min

In case 4 the derivative filter is off. The Loop view gives this tuning a phase margin of 60.2°, more than any filtered setting, yet the ball comes to sit 66 to 94 mm beyond its setpoint. Why does the margin mislead?

(a) The margin is computed for the camera, while the case runs on the wire
(b) A setpoint step kicks the unfiltered derivative, and the integral winds up
(c) The margin leaves out the servo's lag, which the raw derivative excites
(d) The analysis leaves out the rate limit and the $r\dot{\alpha}^2$ term that the shaking beam excites

5.3 Concept check · 1 min

In case 9 a ping-pong ball resting 10 cm from the end of a beam limited to $\pm 3°$ is flicked toward the end at 30 cm/s. Every one of 145 configurations, from P to state feedback and a bang-bang law, loses it. What decides the outcome?

(a) The stopping distance at the angle limit is longer than the 10 cm the ball has
(b) The camera's 50 ms delay leaves too little phase margin for any stabilizing gain
(c) The light ball's small rolling factor makes the loop gain too low to react in time
(d) The integral term winds up against the angle limit while the ball rolls toward the end

5.4 Theory and calculation · 5 min

The steel ball has radius $R = 12.7\,\text{mm}$ and $J = \frac{2}{5}mR^2$, and it rolls on rails $d = 12.7\,\text{mm}$ apart. Use $g = 9.81\,\text{m/s}^2$.

(a) Find the rolling radius ρ and the rolling factor κ, and compare κ with its value on a flat track. *Checkpoint:* 11.00 mm and 0.652, against $5/7 = 0.714$ flat.

(b) Find the acceleration near rest at a beam angle of 1° and at the 15° limit. *Checkpoint:* $0.112\,\text{m/s}^2$ and $1.656\,\text{m/s}^2$.

(c) The hollow ping-pong ball has $R = 20\,\text{mm}$ and $J = \frac{2}{3}mR^2$. Find ρ, κ and the acceleration at 3° on the same rails. *Checkpoint:* 18.97 mm, 0.574 and $0.295\,\text{m/s}^2$.

5.5 Theory and calculation · 9 min

Use the steel ball's $\kappa = 0.652$, $g = 9.81\,\text{m/s}^2$, the servo lag $\tau_s = 60\,\text{ms}$, the control interval $T_s = 10\,\text{ms}$ and the rolling resistance $\alpha_k = 0.05°$.

(a) A proportional law with $K_c = 0.30\,°/\text{cm}$ (0.524 rad/m) runs on an ideal servo. Find the natural frequency and the period. *Checkpoint:* 1.830 rad/s and 3.43 s.

(b) Treat the lag and the hold as a delay $\tau_e = \tau_s + T_s/2$. Find the growth per swing $e^{\pi\omega\tau_e}$ and the fixed loss $4\sin\alpha_k/K_c$ per full swing. Repeat both at 0.15 °/cm, and estimate the ratio of a swing to the one before it when that one was 11.4 cm. *Checkpoint:* 1.45 and 6.7 mm. At 0.15 °/cm, 1.30 and 13.3 mm, a ratio of about 1.19.

(c) Add a derivative time. Find the damping ratio on an ideal servo at $\tau_D = 0.70\,\text{s}$ and the derivative time that gives $\zeta = 1$. Then include the servo's lag and state the condition on τ_D for stability. *Checkpoint:* 0.64 and 1.09 s, and $\tau_D > \tau_s = 60\,\text{ms}$ at any gain.

5.6 Theory and calculation · 6 min

In case 9 the ping-pong ball ($\kappa = 0.574$) rests 10 cm from the end of a beam limited to $\pm 3°$, and a flick sends it toward the end at 30 cm/s. Ignore every delay and assume the beam is already at its limit.

(a) Find the shortest stopping distance, without and with the rolling resistance $\alpha_k = 0.05°$ helping. *Checkpoint:* 15.3 cm and 15.0 cm.

(b) Find the smallest angle limit that stops the ball within 10 cm, and the largest flick that 3° can absorb in 10 cm. *Checkpoint:* 4.58° and 24.3 cm/s.

(c) Would the steel ball ($\kappa = 0.652$) survive the same flick at 3°? *Checkpoint:* No: it needs 13.4 cm.

5.7 App challenge · 8 min

Load scenario 2, Proportional alone: spring and lag, and keep the Phase view. Use the default seed.

(a) Run as shipped until BALL OFF stops the run. Read the On the beam tile and the swing panel: the last swing, the swing before it on the same side, the growth per swing and the period. *Checkpoint:* Off at 16.41 s, −37.3 cm after −26.3 cm, ×1.42 and 3.48 s.

(b) Press Reset, set Proportional gain to 0.15 °/cm, run, and pause between 12 and 13 s. Read the growth and the period. *Checkpoint:* ×1.17 and 4.88 s.

(c) Run on at speed 5 until BALL OFF. Read the On the beam tile and the growth per swing. *Checkpoint:* Off at 37.56 s, growth ×1.25.

Explain why halving the gain lengthens the period by about $\sqrt{2}$, and why the growth in (c) is larger than in (b).

5.8 App challenge · 7 min

Load scenario 4, The sensor: a derivative of noise. Use the default seed and read the phase margin on the Loop view in each part.

(a) Run as shipped and pause at about 15 s. Read the Beam chatter, Servo at its rate limit and Last overshoot tiles. *Checkpoint:* 7.43° within 0.05°, 0.56 s within 0.03 s, 9.4 cm, PM 60°.

(b) Press Reset, set Derivative filter to 30 ms, and run to about 15 s. Read the same tiles and Since the last move. *Checkpoint:* 1.45°, 0.00 s, 0.0 cm, settled in 2.62 s, PM 56°.

(c) Press Reset, set Derivative filter to 200 ms, and repeat (b). *Checkpoint:* 0.30°, still settling, PM 30°.

State what the filter buys and what it costs.

5.9 App challenge · 7 min

Load the challenge scenario, Challenge, the relay race. Use the default seed and speed 5, at which each 60 s run takes 12 s.

(a) Run as shipped. Read the Arrivals tile. *Checkpoint:* 0, ball off within the first 1.5 s.

(b) Press Reveal suggested tuning, then Reset, and run the whole race. Read Arrivals and Challenge target. *Checkpoint:* 15 of 15, met.

(c) Press Reset, set Controller structure to PD, keep everything else, and run again. Read Arrivals, and find on the ball trend which legs miss. *Checkpoint:* 12 of 15. The misses are legs to the left mark after the bend at 35 s.

5.10 Application area · 4 min

A kit builder considers replacing the 30 Hz camera with 33 ms of latency by a 120 Hz camera with 10 ms of latency, under the quick PD of case 7, which crosses over at 5.01 rad/s.

(a) Find the equivalent delay $\theta_c + 1/(2f_c)$ of each camera and the phase it costs at the crossover. *Checkpoint:* 49.7 ms and 14.3°, then 14.2 ms and 4.1°.

(b) The wire gives this loop 38.3°. Estimate the margin on the faster camera, and name two costs of a real camera that the model leaves out. *Checkpoint:* About 34°. Blur, dropped frames, jitter, lighting and price.

5.11 Application area · 4 min

Case 6 calms the hunting across the 0.4° breakaway with a square wave of $\pm 1.5°$ at 10 Hz added to the command. The servo is a first-order lag of 60 ms, and a square wave of amplitude A has a fundamental of amplitude $4A/\pi$.

(a) Find the amplitude of the fundamental at the beam for $\pm 1.5°$ at 10 Hz, $\pm 1.5°$ at 20 Hz and $\pm 0.5°$ at 10 Hz. *Checkpoint:* 0.49°, 0.25° and 0.16°.

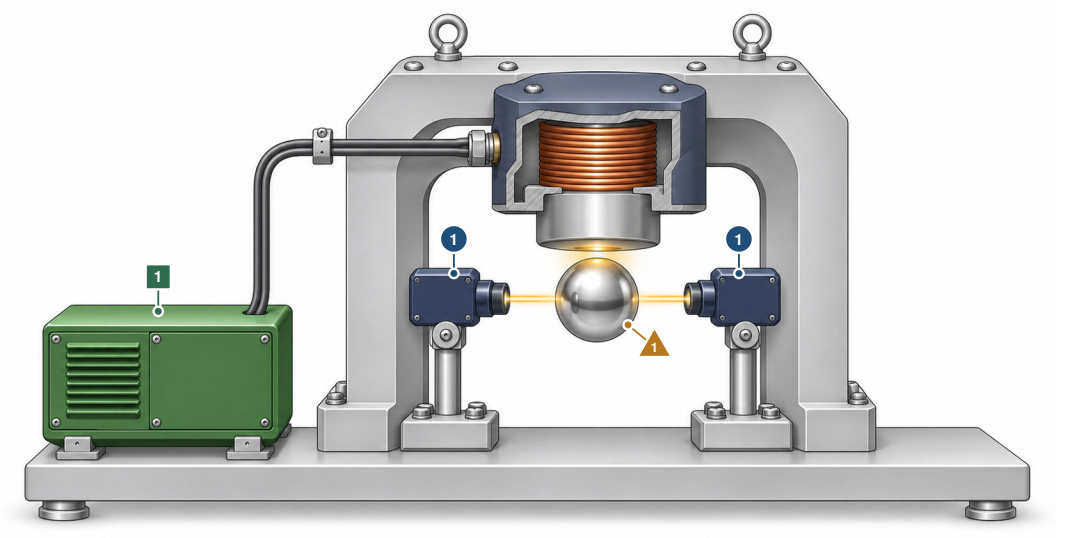
(b) Compare each with the breakaway, and name one cost of dither on a real valve or machine slide. *Checkpoint:* Only the first exceeds 0.4°. Wear and vibration.

5.12 Review the agent · 5 min

An AI assistant reviews case 8, where the quick PD on the ball commands the angle loop of the torque-driven beam, and writes: “The outer loop crosses over at 5.20 rad/s. Any inner angle loop faster than that is fast enough, so the inner bandwidth can be cut from 20 to 6.7 rad/s to save motor effort, with the outer tuning left as it is.” Find the flaw, and say how you would check the claim in the app.

CHAPTER 6

Magnetic Suspension Virtual Laboratory to Teach Control of Unstable Systems



The controller moves 1 current amplifier wired to the coil

The controller reads 1 infrared source and detector, a beam across the gap

What disturbs it 1 a heavier ball or a push away from the magnet

The controller resets the coil current a thousand times a second to hold the ball where the magnet's pull just balances its weight, a balance that does not last without feedback. Illustration, not to scale.

IN PLAIN TERMS A Steel Ball That Will Not Stay Put

The main idea. An electromagnet pulls a small steel ball upward against its weight. At one distance the pull and the weight balance, but if the ball drops a little the pull weakens and the ball falls faster.

The challenge. So a controller watches the ball with a beam of light and changes the current in the magnet a thousand times a second. It must react to how fast the ball moves, and it cannot ask for more current than the power supply can give.

Why it matters. Fast spinning machines and some trains float on the same kind of loop, with nothing touching and no oil.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 6.2, the fact on gaps in service in Section 6.3, the try-its in Sections 6.4 and 6.7, the fact on the derivative filter in Section 6.8, and the key points
Study	55 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	15 min	the first page, the gain limits and results (Sections 6.7 and 6.8), Technology in Practice (Section 6.10), Limitations (Section 6.11) and the open challenge

The control question A steel ball below an electromagnet has an equilibrium at every gap and none is stable, so even the exact balancing current loses it. Which terms of a PID hold the ball, how large must each be, and what bounds the gain from above?

Learning objectives

- Linearize the inverse-square force law about a gap and compute the unstable pole and the control effectiveness.
- Derive the three Routh conditions on K_c, τ_D and τ_I, and explain why proportional or integral action alone cannot hold the ball.
- Predict from the balancing current which gaps an amplifier ceiling can hold, and how much headroom is left for a disturbance.
- Explain how the derivative filter trades current chatter against the phase lead that holds the ball.
- Compare a PID and an explicit MPC on a step into the current ceiling and say what decides the outcome.

Prerequisites Sections 1.4.3 (linearization), 1.4.7 (lag limit, actuator headroom), 1.4.5 and 1.4.6 (LQR, Kalman filter), and Sections 1.6.1, 1.6.2 and 1.6.3 (engine, velocity-form PID, MPC option). Chapter 5 for the same lag limit on a servo.

The app https://apmonitor.com/apps/magnetic-levitation/

See also Volume 2, Chapter 2 for the velocity-form PID and its anti-windup, Volume 2, Chapter 3 for a setpoint outside the actuator range, Volume 2, Chapter 6 for one tuning across a nonlinear gain, and Volume 2, Chapter 12 for an explicit MPC on a machine axis. Volume 9, Chapter 4 for offset-free predictive control with an estimated disturbance.

6.1 Introduction

Almost every simulation used to teach feedback describes a plant that would come to rest somewhere definite if the controller were switched off. Tanks drain to a level, reactors settle at a temperature, exchangers reach a duty, and the loop then improves a condition the process already reaches on its own, which makes a first course in control a course about performance. A magnetic suspension has no such resting state. The upward force on a ferromagnetic ball falls as the square of the gap, so at every distance there is a current that balances the weight exactly, and at that current a small displacement away from the magnet weakens the force that was holding the ball while a small displacement toward the magnet strengthens it. The equilibrium is a maximum of the potential rather than a minimum. Setting the current to eleven correct digits and releasing the ball puts it either on the pole face or on the floor.

Hardware suspension rigs make the argument well and have been documented as undergraduate projects for four decades [1, 2]. A rig is also a power amplifier, a photodetector and a machined pole face, and the time available in a single laboratory period is usually spent assembling it rather than measuring it. However, the published rigs and the published control designs both begin after the instructive failure has been eliminated, so there is no inexpensive place to observe an unstable plant diverge. This chapter presents a simulated magnetic suspension in which manual control is attempted and fails first, and derives from the implemented loop a closed-form condition for when feedback succeeds.

The model rests on standard mechanics. Newton's second law on a single translational degree of freedom supplies one state equation, the reluctance force of a gapped magnetic circuit supplies the driving force as $\alpha i^2/h^2$, and the coil inductance divided by its resistance supplies a first-order actuator lag, so the whole plant is three states and one input [3]. Linearizing an attractive force that depends on the reciprocal square of a displacement produces a positive position stiffness rather than a negative one, which is the origin of the right-half-plane pole and the reason a controller has to supply the stiffness the plant does not have [4, 5]. The scenarios built on that model teach the distinction between an equilibrium and a stable equilibrium, linearization about an operating point and the meaning of an open-loop pole, the separate jobs of the proportional, integral and derivative terms, Routh stability

testing of a fourth-order characteristic polynomial, operating-point dependence of process gain, actuator saturation and anti-windup, and the amplification of sensor noise by derivative action [6, 7]. Industry runs the same loop wherever contact is removed on purpose, in the active magnetic bearings that hold turbomolecular pumps and gas compressors and in the suspension of a guideway vehicle, and in every one of those machines the feedback loop is a load-bearing member [8, 9].

The analysis gives a stability certificate for the loop as implemented. Three Routh inequalities fix the minimum proportional gain, the minimum derivative time and the maximum integral action in terms of the physical constants, and they agree with the nonlinear simulator to within one setting increment. The model is a three-state suspension covering forces, amplifier, sensor, hard stops and numerics, specified closely enough that a second implementation reproduces the same trajectories. Its scenarios make the two mechanisms that bound the gain from above switchable one at a time, so each can be met on its own. An explicit model predictive controller beside the PID adds a measured comparison on a step into the current ceiling, in which runs that remove the ceiling, the proportional kick or the look-ahead one at a time show what each of them decides. Suspension modeling, PID tuning and explicit predictive control are all mature, and the chapter uses each of them as it stands.

Section 6.2 traces the problem from the impossibility of a static suspension to magnetic bearings and guideway vehicles. Section 6.3 states every relation and constant required to rebuild the suspension from this text alone. Section 6.4 linearizes it about a gap and shows why manual operation at the exact balancing current fails. Section 6.5 describes the PID and the explicit MPC, Section 6.6 derives the three stability conditions and tests them against the simulator, and Section 6.7 meets the amplifier ceiling and the sensor noise that bound the gain from above. Section 6.8 sets out the scenarios, the graded challenge and the comparison of the MPC with the PID, and Section 6.9 gives the numerics, and Section 6.10 describes the sensors, amplifiers and controllers of real suspensions. Sections 6.11 to 6.13 state the limitations, the wider lesson and the conclusions, and the exercises of Section 6.14 repeat the measurements on the app.

6.2 Background

The impossibility result predates the engineering. A static arrangement of fixed magnets and ferromagnetic bodies cannot produce a stable suspension, and Braunbek sharpened that statement in 1939 by separating the materials that evade it from the materials that cannot, so diamagnets and superconductors float passively while ordinary ferromagnetics do not [10]. The practical consequence is direct. A steel ball can only be held if something varies the field, and that something is a feedback controller.

> ⧖ HISTORY **A Servo Holds a Rotor in the Air, From the Late 1930s**
>
> A paper of 1937 showed that a vertical ferromagnetic needle can be held in equilibrium by magnetic forces alone if the field is varied, and a 6 g rotor held that way spun at about 1200 revolutions per second [11]. The electromagnetic support developed in the late 1930s was a one-dimensional system: a servo controlled one direction, and only the shape of the field held the others, the single axis this chapter keeps. By 1954 that support had been developed for more than a decade and a half, and an electronically controlled solenoid held a high-speed rotor in vacuum as a frictionless bearing. In 1957 a French aeronautical research office first showed that a wind tunnel model can be held by magnetic suspension, with no support to disturb the flow [11].

Industry adopted the result. Active magnetic bearings replaced oil films in turbomolecular pumps, machine-tool spindles, flywheels and gas compressors, trading contact and lubrication for a control system that has to compute continuously for the machine to stay assembled. The design literature treats the levitation loop, its power stage and the rotordynamics it supports as a single artifact [8]. Transport adopted it in parallel. Electromagnetic-suspension vehicles regulate the same attractive force across a gap of a few millimeters beneath a guideway rail, while electrodynamic suspension becomes passively stable above a threshold speed. Both families are surveyed together [9], and an

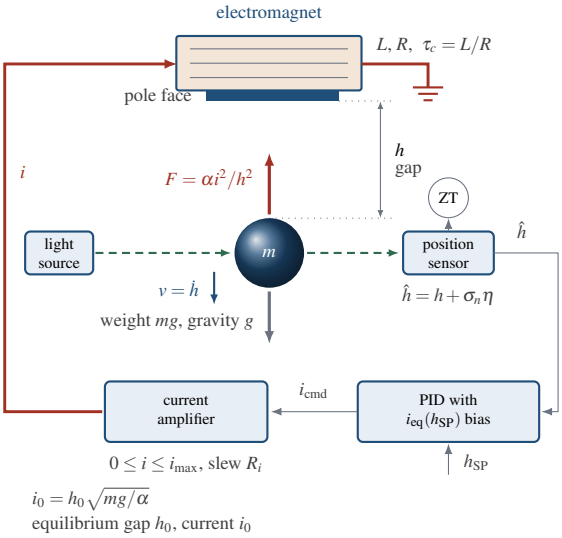


Figure 6.1 Magnetic suspension rig with the electromagnet, the steel ball, the light-beam position sensor and the current path. The gap h is measured downward from the pole face, so the gap rate $v = \dot{h}$ is positive while the ball escapes. Coil current i produces the attractive force $\alpha i^2/h^2$ against the weight mg. The amplifier holds the current between zero and $i_{\max}$ at a slew rate no greater than R_i, and the coil contributes the lag $\tau_c = L/R$. Transmitter ZT reports $\hat{h}$ to a PID controller biased by the equilibrium current i_0 at the setpoint gap h_{SP}.

assessment of maglev economics places the control problem well below the guideway in the cost of a deployment [12].

The single-degree-of-freedom ball and coil became the standard object of study because it isolates the difficulty from everything else. The electromagnet design and the force law for this exact geometry were derived for teaching use [3]. Linear state feedback was compared against exact feedback linearization on one apparatus [13], squaring the current command was shown to cancel most of the nonlinearity [14], a nonlinear controller was built from the same starting point [15], and adaptation was added for an uncertain payload [16]. Underneath the specific designs sits the general accounting for what a right-half-plane pole costs, in the sensitivity integrals of Freudenberg and Looze [17] and in the argument that unstable plants must be respected rather than merely tuned [18]. The simulator of this chapter makes that accounting visible on a plant a student can operate.

6.3 Suspension Model

The simulated rig is a 20 g steel sphere below a single electromagnet driven by a current amplifier, with a light beam across the gap serving as the position measurement. Distance is measured downward from the pole face, so the gap h grows as the ball falls and a positive gap rate $v = \dot{h}$ means the ball is escaping. The controlled variable is the gap reported by the position sensor and the manipulated variable is the current command. There is one loop, one actuator and no inner cascade, so every mechanism in this chapter acts inside a single feedback path. Figure 6.1 draws the rig with each symbol placed where the equations use it.

Table 6.1 sorts the signals into states, the controlled and manipulated variables, the two disturbances and the computed outputs. Two latched events bound the excursion. Below $h_{\min} = 2\,\mathrm{mm}$ the ball is against the pole face and the simulator raises a magnet-contact fault, and above $h_{\max} = 60\,\mathrm{mm}$ it raises a loss-of-capture fault, because beyond that radius no available current produces enough force. Both events annotate the trend, and both clear if the loop recovers the ball, which it can do

Table 6.1 Signals and Their Roles. Nominal Values Are Those of the Default 20 mm Scenario at $t = 0$.

Symbol	Role	Nominal	Range
h	state, true gap	20 mm	2 to 60 mm
$v = \dot{h}$	state, gap rate	0	—
i	state, coil current	1.778 A	0 to $i_{\max}$
$\hat{h}$	CV, measured gap	20 mm	—
h_{SP}	setpoint	20 mm	8 to 34 mm
i_{cmd}	MV, current command	1.778 A	0 to $i_{\max}$ applied
Δm	DV, mass error	0 %	−20 to +20 %
σ_n	DV, sensor noise 1σ	0 mm	0 to 0.5 mm
F	output, magnetic pull	0.196 N	—
mg	output, weight	0.196 N	—
i_{eq}	output, balancing current	1.778 A	—
p	output, unstable pole	31.3 rad/s	—

from contact and cannot do from loss of capture. Three states are integrated in SI units with time in seconds.

🕮 **Pause.** Try Exercise 6.1 before reading on.

6.3.1 Force Balance

The magnet attracts the ball upward with a force proportional to the square of the current and inversely proportional to the square of the gap, while gravity acts downward in the direction of increasing h, as shown in Equation (6.1).

$$m_{\mathrm{eff}}\ddot{h} = m_{\mathrm{eff}}\,g - \frac{\alpha\, i^2}{h^2}, \qquad m_{\mathrm{eff}} = m\left(1 + \tfrac{\Delta m}{100}\right) \tag{6.1}$$

where m is the nominal ball mass in kg, Δm is the mass error in percent set as a disturbance, g is gravitational acceleration in m/s^2 and α is the force coefficient in N m^2/A^2. Both sides of Equation (6.1) carry units of N. The coefficient α lumps coil turns, pole geometry and permeability into one constant, which is the standard reduced form for this configuration [3]. Inside the derivative evaluation the gap is floored at 1×10^{-4} m and the squared gap at 1×10^{-9} m^2, so an integrator excursion cannot divide by zero. Setting $\ddot{h} = 0$ gives the current that balances the weight at any gap, as shown in Equation (6.2).

$$i_{\mathrm{eq}}(h) = h\sqrt{\frac{mg}{\alpha}} = K_{\min}\,h, \qquad K_{\min} \equiv \sqrt{\frac{mg}{\alpha}} \tag{6.2}$$

The constant $K_{\min} = 88.9\,\mathrm{A/m}$ is not a curiosity of the arithmetic. Section 6.6 shows that it is exactly the proportional gain below which no controller of the implemented form stabilizes the ball, at any gap whatsoever. A mass disturbance scales the balancing current by $\sqrt{1 + \Delta m/100}$, and the scaled value is reported alongside the measured gap.

6.3.2 Amplifier, Coil and Sensor

The current cannot step. A first-order current source with a rate limit and a hard ceiling stands between the controller and the coil, as shown in Equations (6.3) and (6.4).

$$\tau_c\,\dot{i} = i_{\mathrm{app}} - i, \qquad 0 \le i \le i_{\max} \tag{6.3}$$

$$i_{\mathrm{app}} = \mathrm{slew}\left(i_{\mathrm{cmd}},\, i_{\mathrm{app}}^{-},\, R_i,\, \Delta t_c\right) \tag{6.4}$$

where τ_c is the coil time constant in s, i_{app} is the current actually applied to the coil in A, i_{app}^{-} is its value at the previous control execution, R_i is the slew limit in A/s and Δt_c is the control period. The

function slew$(\cdot)$ moves the previous applied value toward the command by at most $R_i \Delta t_c$. Physically $\tau_c = L/R$ for a coil of inductance L and resistance R, and the ratio is the adjustable constant rather than the two separately. Setting $\tau_c = 0$ replaces Equation (6.3) by $i = i_{\text{app}}$, and that option exists because τ_c turns out to decide how much derivative action the loop needs. The controller never reads a state. It reads the sensor output of Equation (6.5).

$$\hat{h} = h + \sigma_n \eta(t) \tag{6.5}$$

where σ_n is the noise standard deviation in m and η is a first-order-filtered Gaussian sequence of correlation time $\tau_n = 4\,\text{ms}$, scaled so that its stationary standard deviation is unity. The sequence is drawn from a seeded generator, so a noisy scenario replays identically on any machine, as Section 1.6.1 describes.

6.3.3 Hard Events and Initial Condition

Contact and escape are discontinuous and are applied after the integration step rather than inside the derivative function. If $h \le h_{\min}$ the gap is clipped there and any velocity toward the magnet is zeroed, which is a perfectly inelastic stop against the pole face. If $h \ge h_{\max}$ the gap is clipped and any velocity away from the magnet is zeroed. A negative current is clipped to zero. The contact flag clears above $1.05\,h_{\min}$ and the loss-of-capture flag clears below $0.98\,h_{\max}$, and that hysteresis is what stops the fault banner from flickering at the boundary.

Every scenario starts at the algebraic equilibrium of the setpoint in force at the start of the run, with $h(0) = h_{\text{SP}}$, $v(0) = 0$ and $i(0) = \min\left(i_{\text{eq}}(h_{\text{SP}}),\, i_{\max}\right)$, so a case whose balancing current exceeds the ceiling starts with the coil already at its limit. Two scenarios override $h(0)$ by $\pm 1\,\text{mm}$. No iteration is required because Equation (6.2) is closed form, and that is itself instructive, because the steady state of this plant is trivial to compute and useless on its own.

> **+ DID YOU KNOW? Gaps in Service Are Far Smaller**
>
> An attractive-suspension vehicle surveyed in 1992 kept a clearance of typically 7 to 10 mm to its guideway and levitated even at rest [19]. A compressor on active magnetic bearings had a typical radial clearance of 0.3 mm, and its auxiliary ball bearing 0.15 mm [20]. The chapter's ball hangs 20 mm below its pole face.

Table 6.2 carries every constant a second implementation would need. The proportional gain is entered in A/mm and used internally in A/m, a factor of 1000 that has to be respected when the stability conditions of Section 6.6 are compared with an entered value. Three disturbances are available in any scenario, a mass error spanning $\pm 20\,\%$, a sensor noise standard deviation reaching 0.5 mm, and three one-shot events that displace the ball 1 mm away from the magnet, add 10 % mass, or apply a 0.02 m/s velocity impulse away from the magnet. The impulse is the most selective of the three, because it changes no position and no force, so proportional and integral action see nothing at the instant it lands and only the derivative term reacts.

6.3.4 Assumptions

Only the mechanisms that alter the tuning survive in the model. Every assumption is given below with the operating range in which it is sound and the situation that breaks it.

1. *One translational degree of freedom.* The ball moves only along the coil axis. This is accurate for a centered ball in a rig with a pole face wide relative to the sphere, and it is exactly what a single-axis teaching rig is built to approximate. It fails for any question about lateral guidance, rotation about the axis, or the coupled roll and heave modes that dominate a suspended vehicle [9], and none of those can be studied here.

2. *Inverse-square force law with a lumped coefficient.* The relation $F = \alpha i^2 / h^2$ holds while the core is unsaturated and the gap is small compared with the pole-face diameter, which covers

Table 6.2 Parameters, Controller Settings and Numerics. Every Plant Parameter Is Operator-Adjustable Over the Range Shown Except the Contact Gap $h_{\min}$.

Symbol	Meaning	Value	Unit	Range
Plant				
m	ball mass	0.020	kg	0.010-0.040
α	force coefficient	2.4832×10^{-5}	$\mathrm{N\,m^2/A^2}$	$1\text{-}5\times10^{-5}$
g	gravity	9.81	$\mathrm{m/s^2}$	1.6-24.8
τ_c	coil time constant	0.010	s	0-0.060
$i_{\max}$	current ceiling	3.0	A	1.0-6.0
R_i	current slew limit	400	A/s	20-2000
$h_{\min}$	contact gap	0.002	m	fixed
$h_{\max}$	capture radius	0.060	m	0.040-0.120
τ_n	noise correlation time	0.004	s	fixed
Controller				
K_c	proportional gain	0.400	A/mm	0-1.5
τ_I	integral time	0.50	s	0.05-20
τ_D	derivative time	0.020	s	0-0.08
τ_f	derivative filter	0.001	s	0-0.02
χ_{FF}	equilibrium feedforward	on	—	on/off
Derived at the 20 mm operating point				
$K_{\min}$	$\sqrt{mg/\alpha}$	88.9	A/m	—
i_0	balancing current	1.7778	A	—
p	unstable pole $\sqrt{2g/h_0}$	31.32	rad/s	—
$1/p$	divergence time constant	31.9	ms	—
B	control effectiveness $2g/i_0$	11.04	$\mathrm{m/(s^2\,A)}$	—
Numerics				
Δt	integration step (RK4)	5×10^{-4}	s	—
Δt_c	controller sample period	1×10^{-3}	s	—
—	historian rate	50	Hz	—

the 2 to 60 mm range the simulator uses with α fixed. It fails at large gap, where flux fringing makes the true exponent smaller than two, and at high current, where core saturation makes the force grow more slowly than i^2. A design study would need the flux-path model rather than the lumped constant [3].

3. *Point mass and rigid ball.* The sphere is a point mass with no induced-dipole correction, no eddy currents and no remanent magnetization, so the force depends on the present current alone. This is adequate at the millimeter-per-second speeds the loop produces in normal operation. It fails during a fast escape, where eddy-current drag in a solid steel sphere would add a velocity-dependent term that the model does not carry, and it fails for a hysteretic material, where the force would depend on the current history.

4. *First-order current amplifier.* The power stage is a current source with time constant τ_c, a slew limit R_i and a ceiling $i_{\max}$. This represents a current-mode amplifier working well inside its voltage headroom, which is the normal operating region. It fails when the supply voltage limits first, because back electromotive force from a moving ball and the gap-dependence of the coil inductance both change the achievable $\mathrm{d}i/\mathrm{d}t$, and the inductance dependence matters most over a wide gap range.

5. *Ideal gap sensor apart from added noise.* The measurement is the true gap plus a colored noise of correlation time 4 ms. The noise model is realistic for a photodetector chain, and the correlation time is short relative to the 31.9 ms divergence constant, so the noise is genuinely

broadband as far as the loop is concerned. The sensor has no bandwidth limit, no quantization, no calibration drift and no dependence of gain on gap, so any question about a sensor whose scale factor changes with distance is outside the model.

6. *Perfectly inelastic pole face.* Contact clips the gap and zeros the approach velocity. The result is the correct qualitative outcome, a ball that stays on the magnet until the current is reduced. It fails as a description of the impact itself, because a real sphere rebounds, deforms locally and can chip a coating, and no restitution coefficient is available to change that.

7. *Loss of capture declared at a fixed radius.* Escape is declared at $h_{\max}$ rather than derived from the field. At the default 3 A ceiling the true limit is $i_{\max}/K_{\min} = 33.8\,\text{mm}$, so the 60 mm radius is conservative in the sense that the ball is already unrecoverable well before it is reached. The fixed radius is wrong whenever the ceiling is changed, and the ceiling is adjustable, so the declaration point does not track the physics.

8. *Constant m, g and α during a run.* Mass changes only through the mass-error disturbance and the two mass events. The coefficient α does not drift with coil temperature, although coil resistance in a real rig rises with dissipation and pulls the current down at a fixed drive voltage. Over the tens of seconds a scenario runs, thermal drift is negligible, and over a laboratory session on hardware it is not.

9. *Fixed-step integration and a fixed control period.* Classical fourth-order Runge-Kutta on a 0.5 ms step resolves an instability whose time constant is 31.9 ms with a margin of more than sixty steps per *e*-folding, and the 1 ms control period is comfortable against the same figure. Both become marginal near the 8 mm setpoint limit, where the unstable pole reaches 49.5 rad/s and the coil contributes another at 100 rad/s. Reducing the gap below about 5 mm would require a smaller step.

10. *Continuous-time stability analysis.* The three conditions derived in Section 6.6 come from a linearized continuous-time loop with no derivative filter and no sampling. They are accurate to within one setting increment against the nonlinear simulator, as Section 6.6 reports. They are not a substitute for a sampled-data treatment of a real kilohertz loop, and they say nothing about robustness margins.

11. *Omitted entirely.* Air drag, acoustic and structural resonance of the mounting, magnetic hysteresis, stray fields, amplifier deadband and current measurement noise are all absent. Leaving them out does not alter the feedback structure this application is built to expose.

6.4 Linearization and the Unstable Pole

Writing $h = h_0 + \delta h$ and $i = i_0 + \delta i$ with $i_0 = i_{\text{eq}}(h_0)$ and expanding Equation (6.1) to first order collapses both partial derivatives onto the equilibrium itself, because $\alpha i_0^2/h_0^2 = mg$, as shown in Equation (6.6).

$$\delta\ddot{h} = \underbrace{\frac{2g}{h_0}}_{p^2}\,\delta h \;-\; \underbrace{\frac{2g}{i_0}}_{B}\,\delta i \tag{6.6}$$

where $p = \sqrt{2g/h_0}$ in rad/s is the unstable pole and $B = 2g/i_0$ in m/(s^2 A) is the control effectiveness. This is the linearization of Section 1.4.3 taken at one gap. The open-loop poles are $\pm p$, and by Equation (6.2) the ratio of the two coefficients is a constant of the hardware rather than of the operating point, as shown in Equation (6.7).

$$\frac{p^2}{B} = \frac{2g/h_0}{2g/i_0} = \frac{i_0}{h_0} = K_{\min} \tag{6.7}$$

Table 6.3 Operating-Point Dependence. Closed-Loop Poles Are Roots of (6.16) at the Default Tuning $K_c = 400\,\mathrm{A/m}$, $\tau_I = 0.5\,\mathrm{s}$, $\tau_D = 20\,\mathrm{ms}$.

	11 mm	20 mm	28 mm
i_0 (A)	0.978	1.778	2.489
p (rad/s)	42.2	31.3	26.5
$1/p$ (ms)	23.7	31.9	37.8
B (m/(s^2 A))	20.1	11.0	7.88
dominant pair	-23.5 ± 105.3j	-18.5 ± 70.8j	-14.4 ± 56.2j
damping ratio ζ	0.22	0.25	0.25

Including the amplifier lag, the linearized transfer function from current command to gap is given by Equation (6.8).

$$\frac{\delta h(s)}{\delta i_{\mathrm{cmd}}(s)} = \frac{-B}{(s-p)(s+p)(\tau_c s+1)} \tag{6.8}$$

The negative sign is the physical statement that more current means a smaller gap, and it is the reason the controller runs a reversed error convention.

Pause. Try Exercise 6.2 before reading on.

Table 6.3 lists both quantities at the three gaps the application ships as scenarios. They move in opposite directions. Further from the magnet the instability is slower, which is an advantage, but B falls as well, because a larger equilibrium current means each additional ampere buys proportionally less acceleration. Their ratio does not move at all, by Equation (6.7). One tuning therefore produces three different closed-loop responses across the setpoint range, with the damping ratio nearly unchanged while the natural frequency falls by almost a factor of two.

6.4.1 Manual Operation at the Equilibrium Current

Manual operation sets the coil current directly, in 0.01 A increments, and a helper, the Set current to equilibrium button, evaluates Equation (6.2) at the present gap and applies it. The applied current is not approximately right. It is right to the last digit double precision carries.

Running that configuration untouched reports loss of capture at $t = 3.06\,\mathrm{s}$, and what grew over that interval was floating-point round-off. The computed net acceleration at the exact current is $1.8 \times 10^{-15}\,\mathrm{m/s^2}$, one unit in the last place of g, and the gap does not change in its last bit until $t = 1.95\,\mathrm{s}$. The scenario's stored manual current, 1.777 762 116 9 A, carries eleven significant digits and falls 1.3×10^{-11} A short of the exact value, which shifts the equilibrium by $\delta i / K_{\mathrm{min}} = 1.4 \times 10^{-13}\,\mathrm{m}$, and the same untouched run loses capture at $t = 0.90\,\mathrm{s}$. Covering 20 mm to the 60 mm capture radius from that offset at $p = 31.3\,\mathrm{rad/s}$ takes 0.86 s in the linearization. A deliberate displacement shortens the interval. A $+1$ mm offset loses capture at $t = 0.17\,\mathrm{s}$ and a -1 mm offset reaches the pole face at $t = 0.10\,\mathrm{s}$, faster because the instability is faster on the near side. A human visual-motor loop closes in roughly 250 ms, which is eight e-foldings, or a factor of 2500 by the time a correction is applied. Manual levitation is not difficult. It is arithmetically excluded. Feedback has more time than a hand but not much more. Switched to automatic operation during the fall from the $+1$ mm offset, the loop still catches the ball if the switch comes by $t = 0.07\,\mathrm{s}$, at a 24 mm gap.

TRY IT IN THE APP **2 · +1 mm Displacement (Dropped)**

Load the scenario, which starts in Manual with the ball 1 mm farther from the magnet than its balance point, press Run and watch the gap trend: capture is lost at $t = 0.17\,\mathrm{s}$. Reload it, run, and switch to Auto by $t = 0.07\,\mathrm{s}$, at a 24 mm gap: the loop still catches the ball.

6.5 Controller Design

One digital PID executes every $\Delta t_c = 1\,\text{ms}$, twice as slowly as the physics step. It is the velocity-form PID of Section 1.6.2 [7], with the derivative taken on the measurement so that moving the setpoint produces no derivative kick. The error is written with the measurement first, because the process gain in Equation (6.8) is negative and a larger gap must be answered with a larger current, as shown in Equations (6.9) to (6.12).

$$e_k = \hat{h}_k - h_{\text{SP}} \tag{6.9}$$

$$\Delta u_k^P = K_c\,(e_k - e_{k-1}) \tag{6.10}$$

$$\Delta u_k^I = \frac{K_c}{\tau_I}\,e_k\,\Delta t_c \tag{6.11}$$

$$\Delta u_k^D = K_c\,\tau_D\,\frac{\hat{h}_k^f - 2\hat{h}_{k-1}^f + \hat{h}_{k-2}^f}{\Delta t_c} \tag{6.12}$$

where e_k is the error in m and $\hat{h}_k^f$ is the filtered measurement in m. That filter is an exact zero-order-hold first-order lag of time constant τ_f, $\hat{h}_k^f = a\,\hat{h}_{k-1}^f + (1-a)\hat{h}_k$ with $a = \exp(-\Delta t_c/\tau_f)$, so it is stable for any $\tau_f \geq 0$ because a is never negative. The three increments are combined with the equilibrium bias and the demand is clipped and rate limited, as shown in Equation (6.13).

$$\begin{aligned} z_k &= z_{k-1} + \Delta u_k^P + \Delta u_k^D \\ c_k &= i_{\text{eq}}(h_{\text{SP}})\,\chi_{\text{FF}} + z_k + I_{k-1} \\ u_k &= \text{slew}(\text{clamp}(c_k, 0, i_{\max}), u_{k-1}, R_i, \Delta t_c) \\ I_k &= I_{k-1} + \Delta u_k^I + \frac{\Delta t_c}{T_t}\,(u_k - c_k) \end{aligned} \tag{6.13}$$

where u_k is the applied current command in A, $\chi_{\text{FF}} \in \{0, 1\}$ selects the equilibrium feedforward, z_k and I_k are in A, and $T_t = \tau_I$ is the tracking time of the back-calculation in s. These are Equations (1.13) to (1.15) of Section 1.6.2, with the running sum z_k in place of S_k, the demanded current c_k in place of v_k and the equilibrium current as the bias u_b. While the command is clipped or rate limited, the term $(\Delta t_c/T_t)(u_k - c_k)$ drives I_k toward the value that reproduces the applied current, removing $1/500$ of the excess per scan at the default $\tau_I = 0.5\,\text{s}$, so only the integral is corrected and the running sum z_k is never truncated [7]. The tracking time matters on this plant. A displacement step reaches the derivative path as a step in the measurement, and the 1 mm drop of Section 6.7 asks for 6.4 to 7.9 A on the scan at which it lands. A recursion that absorbed the whole clipped excess in one scan would load I_k with about $-5\,\text{A}$ and lose the ball after the drop at 28 mm. The feedforward term is unity times $i_{\text{eq}}(h_{\text{SP}})$ unless it is switched off, in which case the loop has to build the whole equilibrium current out of proportional and integral action.

In manual operation I_k is set so that $i_{\text{eq}}(h_{\text{SP}})\chi_{\text{FF}} + z_k + I_k$ equals the operator's current exactly, and both transfers are bumpless, as Section 1.6.2 describes. Output limits are 0 to $i_{\max}$ with the slew limit R_i applied afterwards, and the same slew acts on the manual command because a real amplifier cannot step either. The loop is direct acting in the ISA sense, since (6.9) writes the error as $\hat{h}_k - h_{\text{SP}}$ and a positive K_c therefore raises the current as the gap grows, which is the action $a = -1$ of Equation (1.12). The sample period is fixed at 1 ms regardless of the speed setting, and there is no output bias other than the feedforward. The model contains no transport delay, and the application implements no cascade, no gain scheduling and no feedforward beyond the algebraic bias in Equation (6.13), because those structures do not apply to a single-actuator suspension.

Three model layers are kept apart throughout. The simulation model is Equations (6.1) to (6.5), nonlinear and integrated at 0.5 ms. The analysis model is the linearization (6.6) with the amplifier lag, used to derive stability boundaries and, once discretized, as the prediction model of the explicit

controller of Section 6.5.1, and nowhere else. The PID's internal model is the single algebraic relation (6.2) evaluated at the setpoint, so the PID knows nothing of p, B or τ_c, and its feedforward uses the nominal mass, which makes it wrong by design whenever a mass disturbance is applied. The feedforward and the starting command are evaluated with the declared m, α and g of Table 6.2, fixed when the application loads, so a ball, magnet or gravity changed as a parameter also reaches the PID only through the measured gap.

6.5.1 Explicit Model Predictive Control

A second controller can take the coil in automatic operation in place of the PID. It is a model predictive controller (MPC) whose quadratic program is solved offline for every state it can meet. The predictive controllers of Section 1.6.3 solve their program at every decision, which a kilohertz loop cannot afford, and Volume 2, Chapter 12 uses a table of the same kind on a machine-tool axis. When the linear term and the right-hand sides of a quadratic program depend affinely on a parameter, its minimizer is piecewise affine in that parameter over polyhedral critical regions, one for each set of active constraints that is optimal somewhere, so the constrained linear quadratic regulator (LQR) can be stored as a table of regions and laws and evaluated by lookup [21], at a memory and search cost that grows quickly with the horizon and the number of constraints [22]. The PID stays the default, and only scenario 9 selects the MPC.

The prediction model is Equation (6.6) with the coil lag of Equation (6.3), linearized at $h_0 = 20\,\mathrm{mm}$ with the declared constants of Table 6.2 ($p^2 = 981.0/\mathrm{s}^2$, $B = 11.04\,\mathrm{m}/(\mathrm{s}^2\,\mathrm{A})$, $1/\tau_c = 100/\mathrm{s}$). Its state is $x = (h - h_{\mathrm{SP}},\, v,\, i - i_{\mathrm{ss}})$ and its input the move $\delta u = i_{\mathrm{cmd}} - i_{\mathrm{ss}}$, both measured from a target $(h_{\mathrm{SP}}, 0, i_{\mathrm{ss}})$, and an exact zero-order hold over $T_s = \Delta t_c = 1\,\mathrm{ms}$ gives $x_{k+1} = \Phi x_k + \Gamma\,\delta u_k$. The model is written from the declared physics rather than identified, because no open-loop step test exists for a plant that leaves its equilibrium in 32 ms, and central differences of the simulator's own derivative function reproduce $2g/h_0$ and $2g/i_0$ to relative errors of 5×10^{-11} and 7×10^{-11}. Its slopes remain those of 20 mm at every setpoint.

Every millisecond the controller plans three moves of the commanded current, one per period, and applies the first. The objective weighs the gap over the first two predicted steps, prices the state after the third with the LQR's cost-to-go and charges each move's distance from the target current, subject to the ceiling and the slew on all three moves, as shown in Equation (6.14).

$$\begin{aligned} \min_{\delta u_0, \delta u_1, \delta u_2} \; & \sum_{k=1}^{2} Q_h x_{1,k}^2 + x_3^T P x_3 + R \sum_{k=0}^{2} \delta u_k^2 \\ \text{subject to } & x_{k+1} = \Phi x_k + \Gamma\,\delta u_k, \quad x_0 = \hat{x} \\ & 0 \le i_{\mathrm{ss}} + \delta u_k \le i_{\max} \\ & |\delta u_k - \delta u_{k-1}| \le R_i T_s, \quad \delta u_{-1} = i_{\mathrm{app}}^{-} - i_{\mathrm{ss}} \end{aligned} \tag{6.14}$$

where $x_{1,k}$ is the predicted gap deviation, $\hat{x}$ is the estimated state, and P solves the discrete algebraic Riccati equation for Φ, Γ, $\mathrm{diag}(Q_h, 0, 0)$ and R, the terminal weight of the standard stabilizing construction [23], used here without a terminal constraint. Bryson's rule (Section 1.4.5) on the 1 mm band and on the 1.222 A between i_0 and the declared 3 A ceiling gives $Q_h = 1 \times 10^6/\mathrm{m}^2$ and $R = 0.669/\mathrm{A}^2$. With no constraint active the first move is the LQR's, $\delta u = 1.356\,\mathrm{A/mm}\,x_1 + 23.88\,\mathrm{A\,s/m}\,x_2 - 1.497\,x_3$, with closed-loop poles at $-62.5 \pm 80.3\mathrm{j}$ and −130.6/s, a rate gain 2.98 times the PID's $K_c\tau_D = 8\,\mathrm{A\,s/m}$, and feedback on the estimated coil current. Every constraint is hard, the ceiling and the slew are read as the PID reads them, and only inputs are constrained, so the program is feasible at every state.

In the parameter $\theta = (\hat{x},\, i_{\mathrm{app}}^{-} - i_{\mathrm{ss}},\, -i_{\mathrm{ss}},\, i_{\max} - i_{\mathrm{ss}},\, R_i T_s,\, 1)$ the target, the ceiling and the slew enter only the right-hand sides of the twelve inequality rows, so one table serves every setting a learner selects. Of the 173 sets of at most three independent active rows, each with an affine law from its optimality conditions, a set is kept when its critical region holds a ball of positive radius in the

stated domain. That domain is the application's own range of setpoints, ceilings and slews, gaps from contact to the capture radius, rates within ±1 m/s, coil currents of 0 to 6 A and disturbance estimates within ±3 m/s^2, about 30 % of g. It holds 91 regions, of which 63 occur at the declared 3 A and 400 A/s. Gaps to 120 mm and rates to ±5 m/s add one sliver whose center lies in a stored region, and boxes 10^4 times wider give 101 to 105 regions. One to five moves give 5, 23, 91, 283 and 710 regions, and the design admitted fewer than 100, which fixed the horizon at three.

A decision tests the regions in order, the unconstrained one first, each by the sign of its multipliers and by the twelve rows, which define the region, so a region found returns the exact optimum. A parameter that no region accepts is a miss. The PID, which tracks the applied current at every execution, then takes the coil without a bump, the record is marked, and after 20 found decisions in a row the table returns over a 20 ms blend. The lookup agrees with an online solve of the same program by the dual active-set solver of the series to 4.6×10^{-14} A on 100,000 uniform parameters (98,719 with a row active) and to 1.0×10^{-10} A within 10^{-6} of 3000 region boundaries, and 203,983 closed-loop decisions on every scenario and seven hard cases never missed. In Node.js a lookup has a median of 0.72 µs (0.93 µs in the slowest region) against 7.6 µs for the online solve. In the Chrome browser, on the application's frame loop at one and two times real time, the lookup adds 0.16 to 0.31 % of one core and the online solve 1.1 to 2.7 %, both measured on a host shared with other computations.

The MPC reads the measured gap, the setpoint, its applied current and the ceiling and slew. A constant-gain extended Kalman filter (Section 1.4.6) estimates the gap, the rate, the coil current and a constant downward acceleration $\hat{d}$ that stands for every unmodeled force. It predicts with Equations (6.1) and (6.3), the declared constants and $g + \hat{d}$, in two fourth-order Runge-Kutta substeps per period, and it corrects with the steady Kalman gain $(0.187,\ 19.25/\mathrm{s},\ 0,\ 901.9/\mathrm{s}^2)$ of the 20 mm linearization augmented with d, designed for 0.05 mm of measurement noise and 0.05 m/s^2 per step on d, which leaves error poles at $-51.6 \pm 83.8\mathrm{j}$, -103.3 and -100/s. The estimate shifts the target to the declared equilibrium of Equation (6.15), the target-shift form of offset-free predictive control [24].

$$i_{\mathrm{ss}} = h_{\mathrm{SP}} \sqrt{\frac{m\,(g + \hat{d})}{\alpha}} \tag{6.15}$$

On scenario 4 the estimate settles at $\hat{d} = 0.981\,\mathrm{m/s}^2$, the declared model's value for the 10 % heavier ball, with the gap 5.6×10^{-14} mm from its setpoint, while the same controller without the shift ends 0.193 mm off. Volume 9, Chapter 4 removes the offset of a poorer fuel with a Kalman filter that likewise estimates a constant disturbance, there on the fuel. The filter has no pole face, so a measured gap more than 5 mm from the prediction, twice the largest innovation of 38 runs without contact or loss of capture (2.44 mm), restarts it at rest at the measured gap with $\hat{d} = 0$.

The plant integrates Equations (6.1) to (6.5) with the run's parameters, while the MPC holds only the declared ones, so a ball, magnet, gravity or coil lag changed by a learner is model mismatch for the MPC, as it is feedforward error for the PID. The estimator runs in every mode and structure and writes nothing the PID reads, so every scenario and suggested tuning under the PID records the same run, to the last bit, as before the MPC was added.

6.6 Stability Conditions for the Closed Loop

Substituting the ideal PID transfer function $K_c(1 + 1/\tau_I s + \tau_D s)$ into Equation (6.8) and clearing denominators gives the closed-loop characteristic polynomial of Equation (6.16).

$$\tau_c \tau_I s^4 + \tau_I s^3 + \tau_I \left(B K_c \tau_D - p^2 \tau_c\right) s^2 + \tau_I \left(B K_c - p^2\right) s + B K_c = 0 \tag{6.16}$$

Applying the Routh-Hurwitz test and simplifying with $p^2/B = K_{\min}$ produces three conditions, each of which reads as a sentence about the hardware.

The first condition is that the proportional gain must exceed the magnetic stiffness, as shown in Equation (6.17).

$$K_c > \frac{p^2}{B} = K_{\min} = \sqrt{\frac{mg}{\alpha}} = 88.9\,\mathrm{A/m} \tag{6.17}$$

Below this value the loop cannot reverse the sign of the position feedback, because the magnet pulls harder with displacement than the controller pushes back. By Equation (6.7) the bound is identical at every gap, which is why the application can print it as one fixed number, 0.089 A/mm, labeled necessary but not sufficient.

The second condition is that the derivative time must exceed the coil time constant, as shown in Equation (6.18).

$$\tau_D > \tau_c \tag{6.18}$$

This is the sharpest result available from the analysis and the least obvious one. The Routh element that decides it evaluates to $\tau_I B K_c(\tau_D - \tau_c)$, so the proportional gain cancels completely and no amount of gain buys stability if the derivative lead is slower than the lag it has to overcome. With $\tau_c = 10\,\mathrm{ms}$ a derivative time of 6 ms fails at $K_c = 150\,\mathrm{A/m}$ and fails equally at 1000 A/m. It is the lag limit of Equation (1.9) in Section 1.4.7, written for the coil, and Chapter 5 finds the same condition with a servo's lag in place of the coil's.

The third condition is that integral action must be slow enough to be paid for out of the margin the first two conditions earn, as shown in Equation (6.19).

$$\left(BK_c - p^2\right)\left(\tau_D - \tau_c\right) > \frac{1}{\tau_I} \tag{6.19}$$

The two margins on the left form a budget that reset action spends. The default tuning has $BK_c - p^2 = 3434/\mathrm{s}^2$ and $\tau_D - \tau_c = 10\,\mathrm{ms}$, a product of 34.3/s against a requirement of 2/s, so the default sits well inside the region rather than on its edge.

6.6.1 Proportional-Only and Integral-Only Limits

Setting $\tau_c = 0$ and $\tau_D = 0$ in Equation (6.16) leaves $s^2 + (BK_c - p^2) = 0$, a pair of poles on the imaginary axis. Proportional action alone, however large, converts a runaway into an undamped oscillation and goes no further. It never levitates anything.

Keeping reset action but still without derivative action or coil lag gives $\tau_I s^3 + 0 \cdot s^2 + \tau_I(BK_c - p^2)s + BK_c = 0$. The s^2 coefficient is identically zero, so the Routh-Hurwitz necessary condition fails for every K_c and every τ_I. Integral action cannot stabilize this plant at any setting whatsoever. The ranking of the three terms is therefore fixed by the characteristic polynomial rather than by tuning preference, and it is the reverse of the ranking a first course usually establishes.

6.6.2 Checking the Boundaries Against the Simulator

Conditions (6.17) to (6.19) come from a linearization with no derivative filter and no sampling, so they are worth testing against the nonlinear model. Loading the graded scenario, which applies a $+5\,\%$ mass step at $t = 3\,\mathrm{s}$, and asking whether the ball is still captured at $t = 9\,\mathrm{s}$ brackets both testable boundaries within one setting increment. With $\tau_I = 0.5\,\mathrm{s}$ and $\tau_D = 20\,\mathrm{ms}$, Equation (6.19) predicts a critical gain of 107.0 A/m. The ball is lost at 95 and 105 A/m, survives at 110 A/m while spending 13.3 A of current travel against 1.1 A at 120 A/m, and is held cleanly at the higher gain. With $K_c = 150\,\mathrm{A/m}$ and $\tau_I = 5\,\mathrm{s}$, the same condition predicts a critical derivative time of 10.30 ms, and the ball is lost at 9 and 10 ms and held at 11 and 12 ms. Both sets of outcomes are plotted against the predicted boundaries in Figure 6.2.

The residual discrepancy is itself instructive. The capture test cannot place a boundary more finely than the slider increments, and at 110 A/m the ball survives in an oscillation whose swing is still growing, from 4.4 mm peak to peak at $t = 4\,\mathrm{s}$ to 5.9 mm at $t = 38\,\mathrm{s}$. Runs off the slider grid, which a learner cannot make, place the simulator's own boundaries slightly above the predicted ones, between

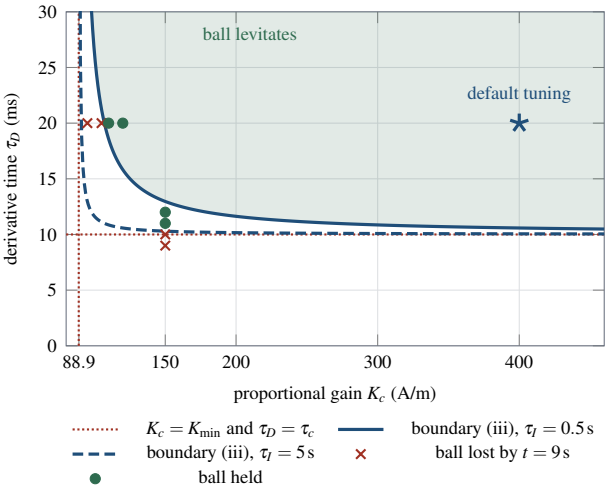


Figure 6.2 Predicted stability region and measured outcomes at the 20 mm gap. The shaded region satisfies all three conditions at $\tau_I = 0.5$ s. The four points at $\tau_D = 20$ ms were run at $\tau_I = 0.5$ s and bracket the predicted critical gain of 107 A/m. The four points at $K_c = 150$ A/m were run at $\tau_I = 5$ s and bracket the predicted critical derivative time of 10.30 ms, so they are judged against the dashed boundary. Crosses mark runs that lost the ball and filled circles runs that held it. The dotted lines are the asymptotes $K_c = K_{\min}$ and $\tau_D = \tau_c$, and the solid curve is the boundary at $\tau_I = 0.5$ s.

110 and 110.5 A/m and between 10.9 and 11.0 ms, because the 1 ms derivative filter and the 1 ms sample period each add a phase lag that Equation (6.18) does not know about, and both lags act in the one path that is holding the ball up.

6.7 Upper Bounds on the Proportional Gain

Conditions (6.17) to (6.19) are all lower bounds. If they were the whole story every gain would be set to its highest available value. Two mechanisms stop that, and the application separates them so each can be met alone.

6.7.1 Amplifier Ceiling

The static current requirement grows linearly with gap by Equation (6.2), so the default 3 A ceiling makes gaps beyond $i_{\max}/K_{\min} = 33.8$ mm unreachable in principle, and the highest setpoint offered, 34 mm, cannot be held at all. The ceiling consumes the dynamic range long before it consumes the static one. Table 6.4 reports the response to a 1 mm displacement away from the magnet, applied as a one-shot event, at three gaps with the tuning unchanged. Peak coil current is 1.27 A at 11 mm and 2.78 A at 28 mm for the identical disturbance, which is 42 % and 93 % of the ceiling, and at 28 mm the command itself sits on the 3 A ceiling for three scans. The increment the loop adds is almost the same at all three gaps, 0.29 A at 11 mm, 0.30 A at 20 mm and 0.29 A at 28 mm, so what consumes the headroom is the equilibrium current itself, which rises linearly with the gap. The gap at which the instability is slowest, 28 mm with $p = 26.5$ rad/s, is also the gap at which the amplifier has the least margin.

One scenario removes the ambiguity by lowering the ceiling to 2.4 A while asking for 28 mm, where $i_{\text{eq}} = 2.49$ A. No equilibrium exists at all, since the highest gap the reduced ceiling can hold is 27.0 mm. The command pins at the limit from the first scan, the ball leaves, and loss of capture is declared at $t = 0.18$ s with a gap error of 32 mm. Lowering the setpoint to 26 mm before the run starts holds the ball on 2.31 A, because the ball then starts at the new gap. Restoring the 3 A ceiling at 28 mm lets the coil start on the 2.49 A equilibrium current, and the ball is held. An actuator that

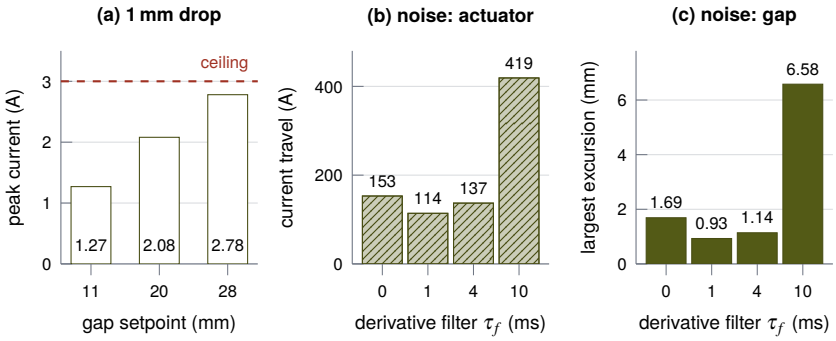


Figure 6.3 The two blocks of Table 6.4. (a) Peak coil current after the same 1 mm displacement at three gaps (white bars), against the 3 A ceiling (dashed line). (b) Current travel (hatched bars) and (c) largest gap excursion (solid bars) over 9.3 s of the noisy-sensor scenario for four derivative filters. Notice that the drop costs almost the same increment at every gap, so the headroom shrinks with the balancing current, and that both the actuator and the gap are best at the 1 ms filter, with 10 ms the worst on both counts.

physically cannot reach the demand is not a tuning problem, and that lesson transfers directly to valves, thrusters and inverters. It is the actuator headroom of Section 1.4.7 in its static form, and Volume 2, Chapter 3 meets it as a dilution that puts the required feed concentration outside the actuator range.

> **TRY IT IN THE APP** **5 · Current Saturation**
>
> Load the scenario and compare the `i_eq at SP` badge, 2.49 A, with the 2.4 A ceiling. Press Run: the command pins at the limit from the first scan and capture is lost at $t = 0.18$ s. Lower Gap setpoint to 26 mm, press Reset and run again: the ball is held on 2.31 A.

6.7.2 Sensor Noise

Derivative action is compulsory by Equation (6.18), and derivative action differentiates noise. The derivative path carries a gain of $K_c\tau_D = 8\,\mathrm{A\,s/m}$ at the default tuning, so a measurement wobbling by 0.15 mm with a 4 ms correlation time injects tenths of an ampere of pure chatter, a standard deviation of 0.44 A in the command at the default filter. The ball filters most of that out of the gap record, which wanders by 0.33 mm (one standard deviation), and the accumulated actuator movement is the reported quantity in which the chatter shows most.

The lower block of Table 6.4 sweeps the derivative filter at fixed noise. Total current travel over 9.3 s falls from 153 A unfiltered to 114 A at the default $\tau_f = 1$ ms, rises again to 137 A at 4 ms, and reaches 419 A at 10 ms, where the largest gap excursion grows from 0.93 mm to 6.58 mm as well. The same ordering holds for eight noise seeds, with travel of 107 to 124 A at 1 ms and 127 to 149 A at 4 ms, and Figure 6.3 draws both blocks of the table. That curve carries the design argument. A filter long enough to matter is a lag in exactly the path that is holding the ball up, and Equation (6.18) says that path has no margin to spare. The optimum is bounded below by the noise and above by the coil, and the usual rule of thumb $\tau_f \approx \tau_D/10$ puts it at 2 ms on this plant, within a factor of two of the 1 ms sample period, which is why physical suspensions run their loops at 10 kHz rather than at 1 kHz.

6.8 Results

Ten set-ups each change one quantity and name what they changed. The first three concern the existence of an equilibrium and are run in manual operation (the first opens in automatic operation and is switched by the learner). The next five run in automatic operation and concern the bounds on

Table 6.4 Measured Performance. Upper Block: a 1 mm Displacement Away From the Magnet at $t = 1$ s With the Default Tuning, Read at $t = 4.7$ s, by Which Time Every Entry Has Converged. Peak Coil Current Is Tracked at the 1 kHz Controller Rate. Lower Block: the Noisy-Sensor Scenario at $\sigma_n = 0.15$ mm, Read at $t = 9.3$ s.

1 mm displacement at three gaps			
h_{SP} (mm)	11	20	28
IAE (mm s)	0.030	0.033	0.049
Current travel (A)	1.46	1.25	1.44
Peak current (A)	1.27	2.08	2.78
Peak over ceiling (%)	42	69	93
Derivative filter under 0.15 mm sensor noise			
τ_f (ms)	0	1	4	10
IAE (mm s)	3.7	2.3	2.8	25.3
Current travel (A)	153	114	137	419
Largest gap excursion (mm)	1.69	0.93	1.14	6.58

Table 6.5 Scenario Set. Modes Are Initial Values and Either Mode May Be Selected at Any Time.

#	Set-up	Mode	Mechanism isolated
1	exact equilibrium, 20 mm	auto	equilibrium is not stability
2	+1 mm displacement	manual	runaway away from the magnet
3	−1 mm displacement	manual	faster runaway on the near side
4	+10 % ball mass	auto	feedforward error, reset removes it
5	$i_{max} = 2.4$ A, SP 28 mm	auto	demand outside actuator capability
6	$\sigma_n = 0.15$ mm	auto	derivative noise amplification
7	SP 28 mm	auto	slower pole, weaker effectiveness
8	SP 11 mm	auto	faster pole, stronger effectiveness
9	explicit MPC, SP 20 to 26 mm at $t = 1$ s	auto	step into the current ceiling, MPC against PID
10	+5 % mass at $t = 3$ s, $K_c = 150$ A/m, $\tau_I = 5$ s, $\tau_D = 6$ ms	auto	graded, violates (6.18)

the gain, the ninth sets the explicit MPC against the PID, and the last is graded. Table 6.5 lists them with the mechanism each one isolates. Every scenario leaves the remaining settings at their defaults, so the effect of the single change is unambiguous.

> ✦ DID YOU KNOW? **A Longer Filter Moves the Coil More**
>
> From the simulator: under 0.15 mm of sensor noise, the coil's current travel over 9.3 s is 153 A with no derivative filter, 114 A at 1 ms, 137 A at 4 ms and 419 A at 10 ms, where the largest gap excursion grows to 6.58 mm (Table 6.4).

The graded challenge is shown in Figure 6.4. It ships $K_c = 150$ A/m, $\tau_I = 5$ s and $\tau_D = 6$ ms, so Equation (6.17) holds while Equation (6.18) is violated by 4 ms, and the loop nonetheless holds a perfectly steady gap for three seconds because nothing has disturbed it. The mass step then excites the $+3.18 \pm 25.0$ rad/s pair predicted by Equation (6.16), the oscillation grows by a factor of e every 314 ms (320 ms between the second and third peaks of the record), the current command first reaches the 3 A ceiling on the third swing at the scan at $t = 3.64$ s, and the ball is gone at $t = 3.94$ s. Raising K_c is the response the first eight scenarios encourage, and it changes nothing here, because Equation (6.18) does not contain K_c. Every gain from 300 to 1500 A/m loses the ball between 3.56 and 3.73 s. The graded target is a gap held within ± 1 mm of setpoint from the mass step to $t = 12$ s with no time at a current limit and a peak below 3 A, and a suggested tuning can be

Table 6.6 Setpoint Step From 20 to 26 mm at $t = 1$ s (Scenario 9), Measured Over $t = 1$ to 3 s

Controller	IAE (mm s)	$n_{3\,\mathrm{A}}/n_{0\,\mathrm{A}}$	i_{pk} (A)	Outcome
PID, default tuning	62.84	1964/3	3.000	lost, 1.225 s
PID, LQR gains	64.80	1968/20	3.000	lost, 1.144 s
I-PD	2.98	0/0	2.308	held
MPC, online solve	0.319	69/12	2.998	held
MPC, lookup	0.319	69/12	2.998	held
Counterfactual runs				
PID, $i_{\max} = 6$ A	0.405	0/3	3.542	held
PID, 50 ms setpoint ramp	0.337	39/0	2.989	held
MPC law clipped, no look-ahead	0.325	70/13	2.999	held
MPC online, ten moves	0.257	55/12	2.993	held

IAE of the true gap at every 0.5 ms step from the event, finer than the 50 Hz scorecard of Table 6.4. $n_{3\,\mathrm{A}}$ and $n_{0\,\mathrm{A}}$ count the 1 ms decisions with the command at the ceiling and at zero, and i_{pk} is the peak coil current. The clipped law is the unconstrained law clipped into the box and the slew on each move. It and the ten-move MPC run outside the application.

revealed for comparison. The challenge is judged on the case as given, so a changed model parameter or disturbance voids an attempt. Read at $t = 12$ s by the challenge's own rule, the shipped tuning leaves a worst gap error of 40.00 mm after the step, 67.0 % of the run at a current limit and a peak of 3.00 A. The suggested tuning, $K_c = 400\,\mathrm{A/m}$, $\tau_I = 0.5$ s and $\tau_D = 20$ ms, keeps the worst error to 0.18 mm, 0.82 mm inside the band, with no time at a limit and a peak of 1.87 A.

Pause. Try Exercise 6.3 before reading on.

6.8.1 Explicit MPC Against the PID

Scenario 9 steps the setpoint from 20 to 26 mm at $t = 1$ s under the MPC, where the declared equilibrium current is 2.31 A of the 3 A ceiling, and it is credited when the gap is within 0.5 mm of 26 mm at $t = 3$ s and the step has then been run under the PID after Reset. Every row of Tables 6.6 and 6.7 uses one seed, sensor, amplifier, plant and duration, the PID at its default tuning (the challenge at both of its tunings) and the MPC at scenario 9's settings, with nothing retuned per case. The step was chosen from twelve candidate cases after both controllers had run on them, so the neighboring steps are reported with it. Table 6.6 adds the online solve, the PID retuned once by the MPC's own rule (the LQR's gap and rate gains as $K_c = 1.356\,\mathrm{A/mm}$ and $\tau_D = 17.6$ ms), the PID with its proportional term on the measurement (I-PD), the classical remedy for a setpoint step into a limit, and four counterfactual runs.

Under the MPC the command falls to 0 A by 1.005 s, stays there to 1.016 s and sits on the ceiling from 1.029 to 1.097 s, while the ball sinks to 28.52 mm and is within 0.1 mm of 26 mm from 1.167 s, an IAE of 0.319 mm s, as shown in Figure 6.5. In the window of the figure the table visits 14 regions and a constraint shapes 92 of its 351 plans, and the lookup and the online solve give the same current at every decision to 2.7×10^{-15} A with no miss. Under the PID, K_c times the 6 mm error removes 2.4 A from the new 2.31 A feedforward, and the command sits at 0 A from 1.005 to 1.007 s. It then climbs more slowly than the MPC's and reaches the ceiling at 1.037 s, 8 ms later, with 2.08 A in the coil against the MPC's 2.39 A, and capture is lost at 1.225 s. With the LQR's gains the PID's kick is 8.1 A and the ball is lost at 1.144 s.

Each run that removes one ingredient holds the ball. The same PID holds the step at 0.405 mm s with the ceiling raised to 6 A, and at 0.337 mm s with the step taken as a 50 ms ramp of setpoint moves, while a 25 ms ramp gives 1.356 mm s, a 10 ms ramp still loses the ball, and the LQR's gains on the 50 ms ramp give 0.191 mm s. The MPC's own law clipped into the box and the slew, with no look-ahead, holds at 0.325 mm s against the table's 0.319. The headline result is therefore the PID's proportional kick into the 3 A ceiling, met on the MPC side by state feedback with three times the PID's rate gain, and not the table's look-ahead, which decides only at the edge. A step to 26.5 mm is held at 0.637 mm s by the table and 0.689 by the clipped law, a step from 20 to 10 mm is held by the

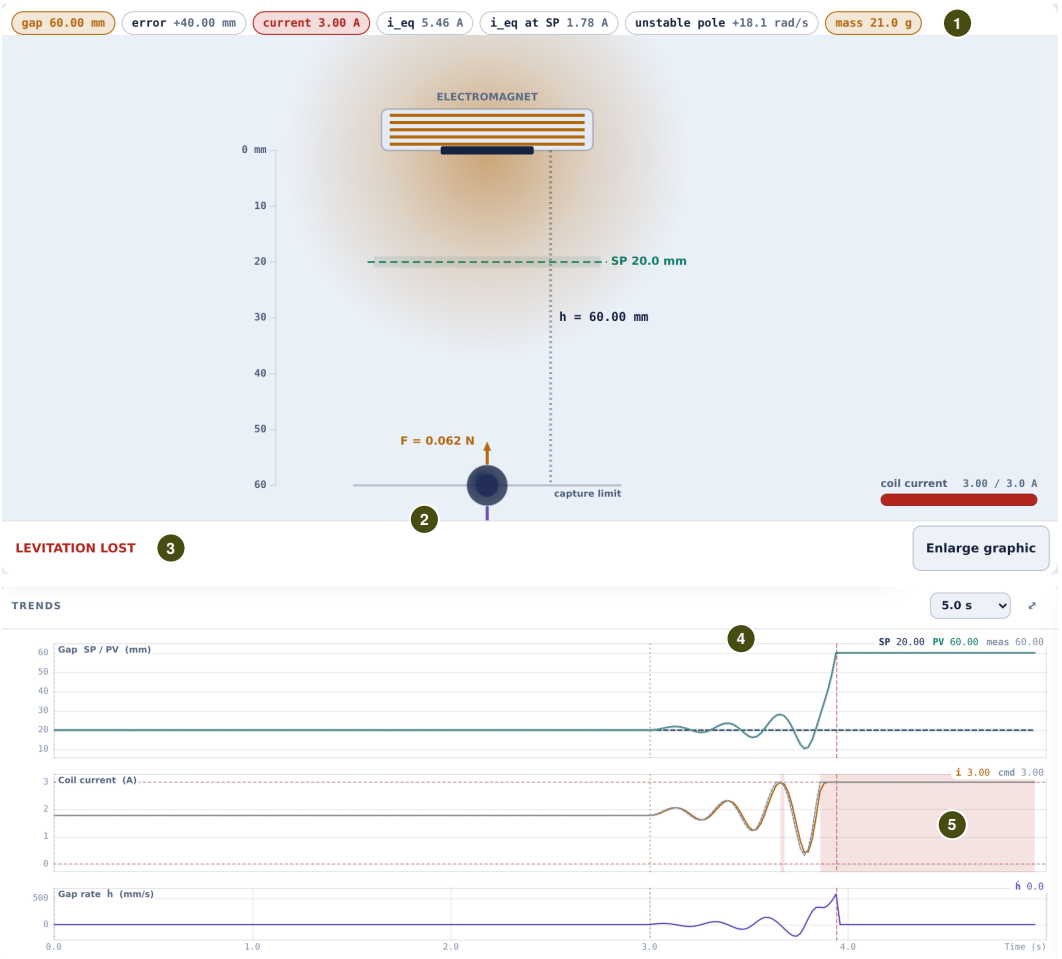


Figure 6.4 The graded scenario at $t = 4.95\,\mathrm{s}$, two seconds after a loop that had held a steady gap for three seconds diverged, in a composite of the screen: the plant view above the trends, with the header that prints the time, the control panel and the scorecard cut. The shipped tuning $K_c = 0.150\,\mathrm{A/mm}$, $\tau_I = 5\,\mathrm{s}$, $\tau_D = 6\,\mathrm{ms}$ satisfies the gain condition (6.17) and violates $\tau_D > \tau_c$ by 4 ms. (1) The badge row: the gap error reads 40 mm, and at the 60 mm radius holding station would need 5.46 A against a 3 A amplifier, with the open-loop pole slowed to $+18.1\,\mathrm{rad/s}$. (2) The ball at the capture limit, where a loss-of-capture fault is latched, and (3) its banner. (4) The gap strip is flat at the 20 mm setpoint until the $+5\,\%$ mass step at $t = 3\,\mathrm{s}$, after which the divergent $+3.18 \pm 25.0\,\mathrm{rad/s}$ mode grows visibly over three cycles and the ball departs on the fourth. (5) The current strip shows the same envelope reaching the dashed 3 A ceiling, with the shaded intervals marking the time spent against it. The gap-rate strip below it spikes as the ball departs, to the 0.59 m/s of the run's record, which the strip does not print, and the cut scorecard reports an accumulated current travel of 9.39 A and 19.8 % of the run at the ceiling.

table while the clipped law touches the magnet at 1.069 s and the PID at 1.077 s, and a step to 27 mm defeats all three. The ten-move online MPC, which the region budget excludes, holds that step at 1.06 mm s, and the I-PD holds every step from 8 to 27 mm, at 2.98 mm s on this one, nine times the table's IAE.

On the mass steps, the drops and the noise of Table 6.7 the table and the clipped law give identical

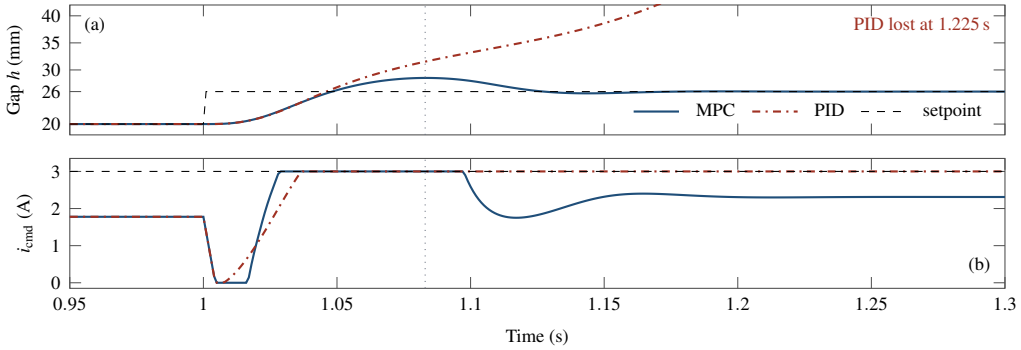


Figure 6.5 Scenario 9 under the explicit MPC and, after Reset, under the PID, from 0.95 to 1.30 s with the setpoint step at 1 s. (a) Gap. (b) Current command against the 3 A ceiling. The MPC is solid, the PID dash-dotted, and the setpoint and the ceiling are thin dashed lines. At the dotted line, $t = 1.083$ s, the MPC's ball is at its deepest, 28.52 mm, and the PID's at 31.53 mm and falling, both with the command on the ceiling.

Table 6.7 Disturbances and Losses, the PID at Its Default Tuning Against the Explicit MPC. IAE From the Event and Largest Gap Error

	IAE (mm s)		e_{max} (mm)	
Case	PID	MPC	PID	MPC
Challenge, +5 % mass[a]	0.055	0.006	0.19	0.15
Challenge, shipped PID tuning[b]	326.6	0.006	40.00	0.15
Scenario 4, +10 % mass	0.108	0.014	0.38	0.29
+10 % mass at 28 mm	0.152	0.019	0.55	0.34
1 mm drop at 20 mm	0.045	0.053	1.00	1.13
1 mm drop at 28 mm	0.059	0.044	1.00	1.00
Scenario 6, noise[c]	2.35	2.02	0.94	1.00
Drop, $R_i = 100$ A/s[b]	0.052	61.14	1.01	40.00
Drop, $m = 0.010$ kg	0.074	8.89	0.86	9.89
SP 28 to 8 mm[d]	2.09	1.49	20.00	20.00

IAE as in Table 6.6, so the PID's 0.045 after the drop at 20 mm is 0.033 on the scorecard of Table 6.4. [a]PID at the suggested tuning, 0.013 with the MPC's target current as its feedforward. [b]Ball lost, by the PID at 3.941 s and by the MPC at 1.640 s. [c]Over 9.3 s, current travel 114 against 212 A. [d]Both touch the magnet at 1.095 s and recover.

rows, so there the comparison is the LQR with its estimator against the PID. On the challenge's 5 % heavier ball the MPC reaches 0.006 mm s against 0.055 for the PID at the suggested tuning and 0.013 for the PID given the MPC's target as its feedforward, so the estimate does most of that work, and with the shipped tuning the PID loses the ball at 3.941 s. The MPC loses four cases. After a 1 mm drop at 20 mm its IAE is 0.053 against 0.045 mm s, with the command 10 ms on the ceiling and 8 ms at zero. Under the noise of scenario 6 its current travel is 212 against 114 A and its time on the ceiling 120 against 24 ms, for a lower IAE, a cost of the LQR's gains that the clipped law shares. After a 1 mm drop the PID holds at every slew tried down to 30 A/s, while the table holds at 120 A/s (0.288 mm s) and loses the ball at 100 and below, as the clipped law does, so that loss belongs to gains designed for the declared 400 A/s. A learner's 10 g ball is model mismatch, and after a drop it swings by 9.89 mm against the PID's 0.86 mm.

No scenario or hard case produces a miss, so 50 ms of misses were forced on scenario 9 with the noise off. At rest and on the ceiling the PID takes the coil within 25 mA of the MPC's command and the ball is held (0.319 to 0.394 mm s). Forced at 1.005 or 1.020 s the hand-over is bumpless, with first steps of 26 and 125 mA against a 400 mA slew step, but the PID, which loses this step by itself, loses the ball. The fallback is only as good as the shipped structure.

6.9 Numerical Implementation

The physics, the controller and the historian advance on separate periods, at 0.5 ms for classical fourth-order Runge-Kutta over the three states, 1 ms for the controller, and 50 samples per simulated second for the record. The engine of Section 1.6.1 keeps the computed trajectory independent of the display, so two machines produce identical exports. Within one step the controller and its sensor read run on every second step, and the hard-event correction follows the integration. The step size is chosen against the fastest element in the model rather than the slowest, because at the 8 mm setpoint limit the unstable pole reaches 49.5 rad/s while the coil contributes another at 100 rad/s. Among the checks that run with the application's test suite, four test the plant and the PID, on the algebraic equilibrium, on a coil current that never exceeds $i_{\max}$ in any scenario, on the sign of the force response to a current increase, and on the sign of the controller response to a growing gap. Further checks cover the feedforward's use of the declared constants and the explicit MPC.

The manual mode supplies the equilibrium current on request, so the existence of an equilibrium can be separated from its stability, and each run exports at 50 Hz for offline analysis.

6.10 Technology in Practice

What Measures It

The chapter's sensor is a light beam across the gap, read as the true gap plus colored noise. An industrial magnetic bearing measures its rotor with displacement probes instead, and a design study of a high-speed spindle lists what is fixed when they are chosen: measurement range, bandwidth, resolution, temperature rating and sensitivity to the magnetic field. The bandwidth caps the frequency up to which the position is measured accurately, because attenuation and phase lag grow as vibration approaches it, and the resolution caps the accuracy the spindle can reach. That spindle chose eddy-current probes whose response lies above its tooth-passing frequency [25]. A probe's lag would add to the filter and sample lags that already lift the simulator's stability boundaries above the predicted ones (Section 6.6).

> **TECHNOLOGY** **Sensor: An Infrared Beam the Ball Partly Blocks**
>
> One teaching rig finds its ball with an infrared beam that starts at a light-emitting diode and ends at a photodiode. The ball partly blocks the beam, so the light that reaches the detector, and the amplitude of its signal, follows the ball's position. The diode is pulsed 10,000 times a second, and an infrared beam rather than a visible one reduces disturbances from room light. With the ball levitating, the rig's filtered position voltage still carried about ±0.05 V of noise on 1.6 V [26]. The chapter's sensor is that beam with colored noise added and no bandwidth limit, drift or gap-dependent gain.

What Moves It

In an industrial bearing, sensors monitor the rotor's position, the signal is compared with a reference, and the controller commands power amplifiers that feed the electromagnets. Each degree of freedom needs two electromagnets, because they can only attract [20]. The chapter's single magnet has gravity as its second. Auxiliary bearings with a smaller clearance than the magnets carry the rotor at rest and let it decelerate safely if the main and redundant features fail together [20]. A high-friction contact with such a bearing can grow into a destructive backward whirl, with contact forces as high as 300 times the rotor's weight [25]. The chapter's pole face is a perfectly inelastic stop instead.

Where the Controller Runs

On one industrial compressor an analog PID loop set the bearing's stiffness and damping: the damping by passive elements of the compensation circuit, the gain by a potentiometer [20]. The teaching rig's analog proportional-derivative controller is built from operational amplifiers, and digital control is offered beside it as a quick way to test whether a set of components can levitate at all [26]. On an attractive-suspension vehicle individually controlled direct-current magnets in the undercarriage hold

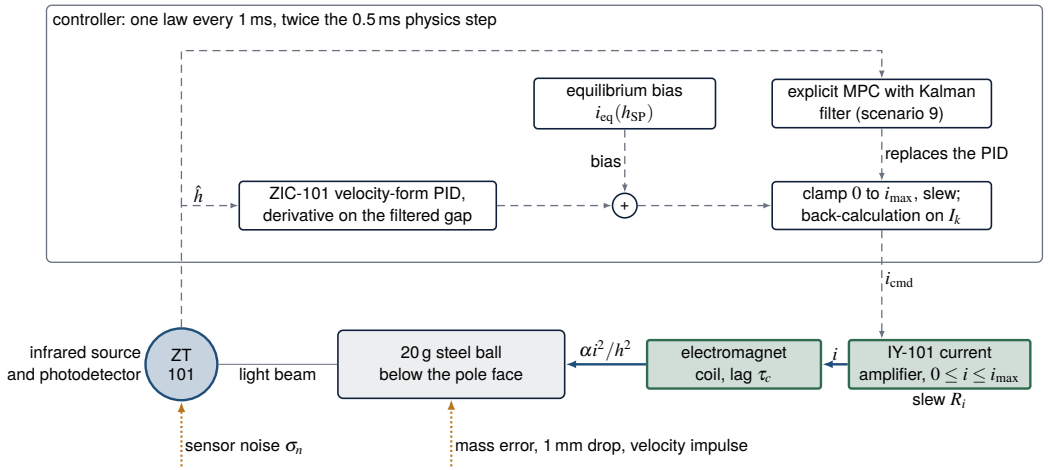


Figure 6.6 The suspension loop on real hardware. An infrared beam partly blocked by the ball, ZT-101, reports the gap $\hat{h}$ to the velocity-form PID ZIC-101, which takes its derivative on the filtered measurement. The equilibrium current at the setpoint enters as a bias, the demand is clamped to the amplifier's range and slew, and back-calculation corrects the integral alone. The current amplifier IY-101 drives the coil, whose lag τ_c the derivative term has to outrun. In scenario 9 the explicit MPC with its Kalman filter replaces the PID. Notice that only the gap is measured: under the PID the ball's speed reaches the loop through the derivative term alone.

the gap, fed through pulse-width-modulated power supplies, and the survey lists complex control systems for suspension and guidance among the drawbacks [19]. The chapter's PID runs every 1 ms, twice the physics step, in velocity form with the balancing current as its bias and back-calculation on the integral alone. Figure 6.6 draws the loop on hardware.

> **TECHNOLOGY** **Final Element: A Power Stage and a Coil That Lags**
>
> On the same rig operational amplifiers drive a power transistor that feeds the electromagnet from a 30 V, 1 A supply, and the ball levitated on 470 mA at an air gap of 7 mm. The coil's time constant, its inductance over its resistance, sets how fast the control signal can change the field in the gap, and a loop on the coil current would improve that response, which the rig did without [26]. Industrial bearings use switching amplifiers, which moved from two output states to three because the amplifier is then more efficient and the bearing has lower eddy-current losses and less acoustic noise [27]. The chapter's amplifier is a current loop of that kind: a lag τ_c, a slew limit and a ceiling.

What Is Modernizing It

Self-sensing removes the probe: the bearing's own coils report the gap through their inductance, and the method has reached turbomolecular pumps and elevator guideways [27]. Design frameworks now carry a bearing spindle from its requirements through the magnets, the backup bearings, the instruments and the control hardware to its assembly, because design knowledge had stayed fragmented across isolated prototypes [25]. The explicit predictive controller of Section 6.5.1, a stored table looked up in place of an online solve, is the chapter's own example of fitting optimization into a kilohertz loop.

IN PRACTICE **Proportional First, Then Derivative**

The teaching rig's analog controller is set on two potentiometers. The proportional gain is set first, with the derivative gain low, while the ball is held at the desired place, and too little gain cannot lift it. With enough proportional gain but too little derivative gain the ball may fly into the electromagnet and bounce off, and more derivative gain stops it. A levitating ball that vibrates also asks for more derivative gain. Settings found before the magnet and the power transistor warm up may need a final adjustment [26]. Condition (6.18) explains the order: the derivative term is what holds the ball up.

6.11 Limitations

This simulator teaches a loop and does not size a magnet. Its single degree of freedom excludes the guidance and lateral-stability problems that dominate real suspension vehicles [9], and its lumped force coefficient excludes the magnetic design questions that determine α [3]. The controllers are a fixed-structure PID and a three-move explicit MPC on one linearization, so the feedback-linearization and adaptive designs this plant motivated [13–16] cannot be demonstrated, although Table 6.3 showing one tuning producing three different closed-loop responses is exactly the evidence that makes those designs look necessary. Every disturbance except the sensor noise is applied on command rather than arising as stochastic plant behavior. A fixed integration step and a fixed control period give bit-identical repeats, which no field controller delivers. Section 6.3.4 states where each modeling assumption breaks down, and the three that bind hardest are the constant force coefficient at large gap, the absence of voltage saturation in the amplifier, and the continuous-time basis of the stability conditions.

The explicit MPC's model keeps the slopes of the 20 mm design gap at every setpoint, so the $1/h^2$ curvature enters only through the target of Equation (6.15). Its estimator treats the colored sensor noise as white, has no pole face or capture radius, and knows only the declared ball, magnet, gravity and coil lag, so a changed plant is absorbed at steady state but not in a fast transient, as the 10 g ball shows. Three 1 ms moves are the longest horizon whose table stays under 100 regions, too short to anticipate a slow amplifier, and a state outside the tabulated domain hands the coil to a PID that may not hold the ball.

OPEN CHALLENGE **Sensing the Gap With the Coil Itself**

A self-sensing magnetic bearing drops the position probe and estimates the gap from its coils' voltage and current, because a coil's inductance changes with the gap. A review found it in commercial use, with hurdles left for demanding products. Its robustness hinges on the size of the switching ripple, which three-state amplifiers cut, sometimes about tenfold, so a robust self-sensed bearing is likely to use somewhat more electrical power than one with a probe. Eddy currents and saturation of the iron also blur the coil's sensitivity to the gap [27]. Progress would be self-sensing that is both efficient and robust.

6.12 Discussion

Removing contact removes wear, lubrication, friction loss and a maintenance schedule, which is why active magnetic bearings hold rotors in turbomolecular pumps, gas compressors and flywheels, and why those rotors turn faster than any bearing that touches them permits [8]. The price is that the machine has no passive resting state. A power failure in an oil-film bearing is an interruption, while a power failure in a magnetic bearing drops the rotor onto its auxiliary bearings, and much of the design discipline of that field follows from the asymmetry. The same argument scales to electromagnetic-suspension vehicles, where the gap under a guideway rail is a few millimeters and the loop that maintains it carries the vehicle [9, 12].

The transferable lesson is broader than magnets. A right-half-plane pole at p imposes a floor rather than a preference, because the sensitivity integral is fixed at πp instead of zero, so the closed-loop bandwidth must exceed the instability and the disturbance amplification that appears elsewhere in frequency cannot be tuned away but only relocated [17]. Such plants must be respected rather than

merely tuned, because a control failure on them is qualitatively different from a poorly damped response [18]. Aircraft built with relaxed static stability, reactors held on an unstable branch for yield, the elongated plasma of a tokamak (Volume 18, Chapter 12) and this suspension all share that structure, and a stability margin on such a plant bounds the time available to act rather than the quality of a response.

6.13 Conclusion

An electromagnet and a steel ball make a plant with an equilibrium at every gap and stability at none, and that property reorders what a PID loop is understood to be for. The simulation of this chapter integrates three states under an inverse-square force with a first-order current amplifier, a 3 A ceiling, a 400 A/s slew limit and a gap sensor with a 4 ms correlation time, and closes a kilohertz digital PID in velocity form with an equilibrium feedforward bias and back-calculation anti-windup on the integral alone. Linearization about a 20 mm gap gives an open-loop pole at $+31.3\,\mathrm{rad/s}$, and Routh analysis of the resulting fourth-order loop gives three inequalities that decide whether the ball floats. The simulator brackets the predicted critical gain of 107.0 A/m between 105 and 110 A/m and the predicted critical derivative time of 10.30 ms between 10 and 11 ms. The order of those inequalities is the teaching point. Proportional gain above the magnetic stiffness of 88.9 A/m only converts a runaway into an undamped oscillation, integral action cannot stabilize the plant under any setting whatsoever, and derivative action longer than the 10 ms coil time constant is what holds the ball up, bounded from above by the amplifier ceiling and by the sensor noise it differentiates. An explicit MPC of 91 regions holds a step into the 3 A ceiling that the PID loses (0.319 mm s against a lost ball), although the same law clipped without look-ahead holds it at 0.325 mm s, so the PID's proportional kick, not the optimization, decides that case. A fuller model would add a gap-dependent coil inductance with a voltage limit, an eddy-current drag term, a second translational axis for lateral stability, and a sampled-data version of the three conditions. The suspension is an accessible browser platform for studying a loop that manufactures its own operating point.

KEY POINTS

- A steel ball below an electromagnet has a balancing current at every gap and is stable at none, so even the exact current loses the ball.
- Proportional gain alone gives an undamped oscillation at best, and integral action cannot stabilize the plant.
- Derivative action longer than the 10 ms coil lag holds the ball up at any gain.
- The amplifier ceiling and the sensor noise that derivative action amplifies bound the gain from above.

Where else this applies. Volume 18, Chapter 12 holds a plasma with the same open-loop instability.

References

[1] T. H. Wong, "Design of a magnetic levitation control system—an undergraduate project," *IEEE Transactions on Education*, vol. E-29, no. 4, pp. 196–200, 1986.

[2] P. S. Shiakolas and D. Piyabongkarn, "Development of a real-time digital control system with a hardware-in-the-loop magnetic levitation device for reinforcement of controls education," *IEEE Transactions on Education*, vol. 46, no. 1, pp. 79–87, 2003.

[3] W. G. Hurley and W. H. Wolfle, "Electromagnetic design of a magnetic suspension system," *IEEE Transactions on Education*, vol. 40, no. 2, pp. 124–130, 1997.

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

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

[8] E. H. Maslen and G. Schweitzer, Eds., *Magnetic Bearings: Theory, Design, and Application to Rotating Machinery*. Berlin, Heidelberg: Springer, 2009.

[9] H.-W. Lee, K.-C. Kim, and J. Lee, "Review of maglev train technologies," *IEEE Transactions on Magnetics*, vol. 42, no. 7, pp. 1917–1925, 2006.

[10] W. Braunbek, "Freischwebende Körper im elektrischen und magnetischen feld," *Zeitschrift für Physik*, vol. 112, no. 11-12, pp. 753–763, 1939.

[11] M. H. Tuttle, R. A. Kilgore, and R. P. Boyden, "Magnetic suspension and balance systems: A selected, annotated bibliography," NASA Langley Research Center, Hampton, Virginia, NASA Technical Memorandum TM-84661, Jul. 1983. [Online]. Available: https://ntrs.nasa.gov/citations/20080047362

[12] R. D. Thornton, "Efficient and affordable maglev opportunities in the United States," *Proceedings of the IEEE*, vol. 97, no. 11, pp. 1901–1921, 2009.

[13] W. Barie and J. Chiasson, "Linear and nonlinear state-space controllers for magnetic levitation," *International Journal of Systems Science*, vol. 27, no. 11, pp. 1153–1163, 1996.

[14] D. L. Trumper, S. M. Olson, and P. K. Subrahmanyan, "Linearizing control of magnetic suspension systems," *IEEE Transactions on Control Systems Technology*, vol. 5, no. 4, pp. 427–438, 1997.

[15] A. El Hajjaji and M. Ouladsine, "Modeling and nonlinear control of magnetic levitation systems," *IEEE Transactions on Industrial Electronics*, vol. 48, no. 4, pp. 831–838, 2001.

[16] Z.-J. Yang and M. Tateishi, "Adaptive robust nonlinear control of a magnetic levitation system," *Automatica*, vol. 37, no. 7, pp. 1125–1131, 2001.

[17] 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.

[18] G. Stein, "Respect the unstable," *IEEE Control Systems Magazine*, vol. 23, no. 4, pp. 12–25, 2003.

[19] J. L. He, D. M. Rote, and H. T. Coffey, "Survey of foreign maglev systems," Argonne National Laboratory, Center for Transportation Research, Argonne, Illinois, Tech. Rep. ANL/ESD-17, Jul. 1992. [Online]. Available: https://www.osti.gov/servlets/purl/10134413

[20] R. G. Kirk, J. F. Hustak, and K. A. Schoeneck, "Active magnetic bearings applied to industrial compressors," in *Magnetic Suspension Technology Workshop*, ser. NASA Conference Publication, no. 3202. Hampton, Virginia: NASA Langley Research Center, 1993. [Online]. Available: https://ntrs.nasa.gov/citations/19930018381

[21] 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.

[22] A. Alessio and A. Bemporad, "A survey on explicit model predictive control," in *Nonlinear Model Predictive Control*, ser. Lecture Notes in Control and Information Sciences, L. Magni, D. M. Raimondo, and F. Allgöwer, Eds. Berlin, Heidelberg: Springer, 2009, pp. 345–369.

[23] 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.

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

[25] K. S. Ahmed and B. Bediz, "Design framework and manufacturing of an active magnetic bearing spindle for micro-milling applications," arXiv:2603.00169v2 [eess.SY], 2026.

[26] S. Barker and R. C. Matusiak, "A technology approach to magnetic levitation," in *2000 ASEE Annual Conference & Exposition*, St. Louis, Missouri, 2000.

[27] E. H. Maslen, "Self-sensing for active magnetic bearings: Overview and status," in *Proceedings of ISMB10, International Symposium on Magnetic Bearings*, 2006. [Online]. Available: https://www.magneticbearings.org/app/uploads/publications/ismb10/ismb10_submission_key03.pdf

6.14 Exercises

6.1 Concept check · 1 min

The suspension has one loop. Which pair names what the controller reads and what it adjusts?

(a) It reads the coil current and adjusts the gap setpoint
(b) It reads the measured gap and adjusts the coil current command
(c) It reads the ball mass and adjusts the force coefficient α
(d) It reads the magnetic force and adjusts the amplifier ceiling

6.2 Concept check · 2 min

Scenario 1 starts the ball at 20 mm with the balancing current. The reader switches to Manual before Run and touches nothing. What happens?

(a) The ball stays at 20 mm, because the net force on it is zero
(b) The ball drifts to a nearby gap where the magnetic force again balances
(c) The ball leaves within a second or so, because any error grows by e every 32 ms
(d) The ball swings about 20 mm, because the magnet pulls back like a spring

6.3 Concept check · 2 min

The challenge ships $K_c = 0.150\,\mathrm{A/mm}$, $\tau_I = 5\,\mathrm{s}$ and $\tau_D = 0.006\,\mathrm{s}$, with the coil time constant $\tau_c = 10\,\mathrm{ms}$, and the ball is lost after the 5 % mass step. Which single change keeps the ball?

(a) Raise K_c to 1.000 A/mm, far above the floor of 0.089 A/mm
(b) Shorten τ_I to 0.50 s so the integral removes the sag sooner
(c) Raise the derivative filter to 0.010 s to quiet the derivative term
(d) Raise τ_D to 0.020 s, above the coil time constant

6.4 Theory and calculation · 6 min

Use the declared ball and magnet: $m = 0.020\,\mathrm{kg}$, $\alpha = 2.4832 \times 10^{-5}\,\mathrm{N\,m^2/A^2}$ and $g = 9.81\,\mathrm{m/s^2}$.

(a) Find $K_{\min} = \sqrt{mg/\alpha}$ and the balancing current at gaps of 20 and 28 mm. *Checkpoint:* 88.9 A/m, 1.778 A and 2.489 A.

(b) Find the unstable pole p, its time constant $1/p$ and the control effectiveness B at both gaps, and show that p^2/B is the same at both. *Checkpoint:* 31.3 rad/s, 31.9 ms, 11.04 m/(s^2 A) at 20 mm, and 26.5 rad/s, 37.8 ms, 7.88 m/(s^2 A) at 28 mm. $p^2/B = 88.9\,\mathrm{A/m}$ at both.

(c) At the stored manual current the ball starts 1.4×10^{-13} m from its equilibrium at rest, so the linear model gives $\delta h = \delta h_0 \cosh pt$. How long does it take to cover the 40 mm to the capture radius at 20 mm, and by what factor does an error that is already growing, $\delta h \propto \mathrm{e}^{pt}$, grow during a 250 ms human reaction? *Checkpoint:* 0.86 s, and about 2500 ($pt = 7.8$).

6.5 Theory and calculation · 8 min

The closed loop of the PID on the linearized plant has the characteristic polynomial (6.16). Work at 20 mm with $p^2 = 981/\mathrm{s^2}$, $B = 11.04\,\mathrm{m/(s^2\,A)}$ and $\tau_c = 10\,\mathrm{ms}$.

(a) Set $\tau_c = 0$ and $\tau_D = 0$. Write the polynomial for proportional action alone, with $K_c = 400\,\mathrm{A/m}$, and for PI action. What does each loop do? *Checkpoint:* $s^2 + (BK_c - p^2) = 0$, an undamped oscillation at 58.6 rad/s. The PI polynomial has no s^2 term and is unstable for every setting.

(b) With $\tau_I = 0.5\,\mathrm{s}$ and $\tau_D = 20\,\mathrm{ms}$, find the smallest K_c that satisfies condition (6.19), and the size of the budget at the default $K_c = 400\,\mathrm{A/m}$. *Checkpoint:* 107.0 A/m, and 34.3/s against the 2/s needed.

(c) With $K_c = 150\,\mathrm{A/m}$ and $\tau_I = 5\,\mathrm{s}$, find the smallest derivative time. *Checkpoint:* 10.30 ms.

6.6 Theory and calculation · 6 min

The amplifier holds the current between 0 and $i_{\max} = 3.0\,\mathrm{A}$ and slews at most $R_i = 400\,\mathrm{A/s}$. Use $K_{\min} = 88.9\,\mathrm{A/m}$ and the default tuning $K_c = 400\,\mathrm{A/m}$, $\tau_I = 0.5\,\mathrm{s}$, $\tau_D = 20\,\mathrm{ms}$, $\tau_f = \Delta t_c = 1\,\mathrm{ms}$.

(a) How far below its setpoint can the ball sit at rest and still be lifted by the full current, at 20 and at 28 mm? What is the highest gap a 2.4 A amplifier can hold? *Checkpoint:* 13.75 mm and 5.75 mm (both to 33.75 mm), and 27.0 mm.

(b) A 1 mm drop lands between two scans. On the next scan the filtered measurement moves by $(1 - \mathrm{e}^{-1})\,1\,\mathrm{mm}$. Find the current the PID demands, c_k, at 20 and 28 mm, starting from the balancing current with $I = 0$. *Checkpoint:* 7.23 A and 7.95 A.

(c) The slew limits the applied current to $i_0 + R_i \Delta t_c$. Find the excess $c_k - u_k$ and the change it makes in I_k through Equation (6.13), against a recursion that absorbs the whole excess in one scan. *Checkpoint:* 5.06 A at both gaps, and -0.0101 A against -5.06 A.

6.7 App challenge · 7 min

Load scenario 2, +1 mm displacement (dropped), which opens in Manual with the balancing current for 20 mm and the ball 1 mm lower. Set the speed to 0.01, so 0.01 s of simulated time passes per second, and use the default seed (the scenario has no noise). Read times and gaps from the gap trend.

(a) Run until LEVITATION LOST is marked and read the time of the mark. Then load scenario 3, −1 mm displacement (pulled up), and do the same for MAGNET CONTACT. *Checkpoint:* 0.17 s and 0.10 s (±0.01 s).

(b) Reload scenario 2 and press Run. Switch to Auto when the clock reads between 0.04 and 0.06 s, and let it run another 1 s. Read the gap at the switch and the Peak current tile. *Checkpoint:* The ball is caught. Gap 21.8 to 23.1 mm, peak 2.4 to 2.9 A.

(c) Reload and repeat with the switch at 0.09 s. *Checkpoint:* The ball is lost. The gap at the switch is about 27 mm.

6.8 App challenge · 7 min

Use the default tuning, the default seed and speed 1. In each run press Run, fire Drop ball 1 mm at any time between 0.5 and 1.5 s, and pause at about 4.7 s. Read the Peak current and MV travel tiles.

(a) Scenario 8, Low levitation position (11 mm). *Checkpoint:* Peak 1.27 A, MV travel 1.46 A.

(b) Scenario 1, Equilibrium without stability, which holds 20 mm in Auto, and scenario 7, High levitation position (28 mm). *Checkpoint:* 2.08 A and 1.25 A, then 2.78 A and 1.44 A.

(c) Using the balancing currents $i_0 = K_{\min} h_{\mathrm{SP}}$, find each peak as a share of the 3 A ceiling and the current the loop added above i_0. What uses up the headroom? *Checkpoint:* 42, 69 and 93 %. Added: 0.29, 0.30 and 0.29 A.

6.9 App challenge · 8 min

Load the challenge scenario, Challenge, reject a 5% mass increase, and set the speed to 2 (a 12 s run takes 6 s) and use the default seed. Make each setting before Run, run to 12 s and read the challenge's result line (worst gap error, time saturated, peak current). Reload before each part.

(a) Run as shipped. Then set Proportional gain to 1.000 A/mm and run again. Read the time of LEVITATION LOST in each run. *Checkpoint:* Lost at 3.94 s, then lost at 3.68 s.

(b) Set only Derivative time to 0.020 s. *Checkpoint:* Held: 1.01 mm, 0.0 %, 1.94 A. The target fails by 0.01 mm.

(c) Set Derivative time to 0.020 s and Integral time to 0.50 s. Then press Reveal suggested tuning, Reset and run once more. *Checkpoint:* 0.92 mm, 0.0 %, 1.94 A, met. Suggested: 0.18 mm, 0.0 %, 1.87 A.

6.10 Application area · 4 min

A turbomolecular pump runs its rotor on active magnetic bearings.

(a) Compare what a power failure does to an oil-film bearing and to a magnetic bearing, and name the equipment the magnetic bearing needs as a result. *Checkpoint:* An interruption against a dropped rotor, and auxiliary (touchdown) bearings.

(b) The chapter's loop runs at 1 kHz, and the best derivative filter is near $\tau_D/10 = 2\,\text{ms}$. Why do physical suspensions run their loops at 10 kHz? *Checkpoint:* The 1 ms sample period is within a factor of two of the filter, and both lags sit in the path that holds the ball.

6.11 Application area · 4 min

The explicit MPC of the chapter stores 91 regions for three 1 ms moves. One to five moves give 5, 23, 91, 283 and 710 regions. In Node.js a lookup takes at most 0.93 µs and an online solve of the same program a median of 7.6 µs.

(a) A bearing controller must decide within 100 µs on a processor 50 times slower than that host. Which evaluation fits? *Checkpoint:* The lookup, at about 47 µs. The online solve needs about 380 µs.

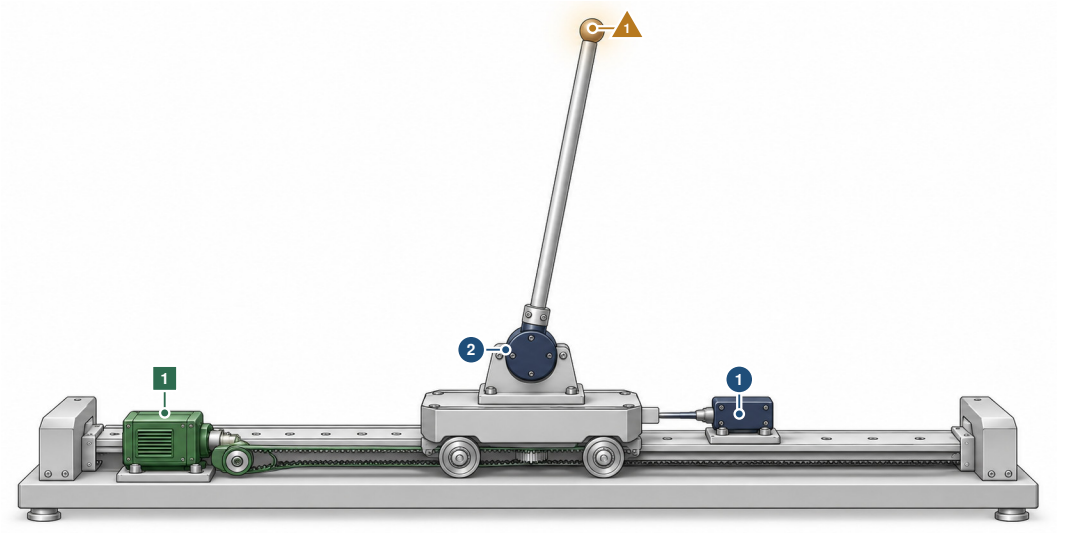
(b) What does the short horizon cost on this plant? Give two cases from the chapter. *Checkpoint:* A slow amplifier (lost at 100 A/s) and the step to 27 mm, which a ten-move online MPC holds.

6.12 Review the agent · 5 min

For scenario 6, Noisy position sensor ($\sigma_n = 0.15\,\text{mm}$), an AI assistant writes: *"The coil current chatters by 0.44 A because the derivative term amplifies the noise. Raise the derivative filter from 0.001 s to $\tau_D/2 = 0.010\,\text{s}$. That cuts the noise bandwidth tenfold and the MV travel with it, and since the derivative only adds damping, the gap will barely change."* Find the flaw, and say how you would check it in the app.

CHAPTER 7

Inverted Pendulum Simulator for Balancing a Pole Without Losing the Cart



The controller moves 1 motor and belt that push the cart

The controller reads 1 where the cart is on the rail 2 the pole's angle at its hinge

What disturbs it 1 a tap on the pole

Notice that the only way to catch the leaning pole is to drive the cart under it, while the cart must stop short of the bumpers at both ends of the rail. Illustration, not to scale.

IN PLAIN TERMS Keeping a Pole Up Without Losing the Cart

The main idea. A motor pushes a small cart along a short rail, with a pole standing on it on a free hinge. The controller keeps the pole upright by moving the cart under it, like a hand under a stick balanced on a fingertip.

The challenge. Left alone, the pole falls over in less than a second. A controller that watches only the pole keeps it up while the cart drifts into the end of the rail.

Why it matters. Balancing robots, walking machines and rockets standing on their thrust all share this problem.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact on a launch vehicle in Section 7.1, the history in Section 7.2, the try-its and the fact on a longer pole in Section 7.6, and the key points
Study	65 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 7.6), Technology in Practice (Section 7.8), Limitations (Section 7.9) and the open challenge

The control question One motor pushes the cart, and the pole responds only to its acceleration, so stopping the fall and keeping the cart on a finite rail share one input. Which loops hold the pole without losing the cart, and which gains, delays and pole changes defeat them?

Learning objectives

- Derive the unstable root, the wrong-way zero and the angle-gain floor $(M+m)g$ from the linearized cart and pole.
- Explain why rail friction makes an angle loop alone run the cart into a bumper, and why every cart move starts the wrong way.
- Use the characteristic polynomial of the cascade to predict which coefficient a longer or a heavier pole attacks, and which fix restores it.
- Relate the stable window of the angle gain to the measurement delay and sample interval.
- Compare the regulator with a predictive controller on the same model, and explain what the terminal cost does for a short horizon.

Prerequisites Sections 1.4.4 and 1.4.5 (state space, LQR, Bryson's rule), 1.4.7 (unstable pole, RHP zero, delay window), 1.6.1 (engine) and 1.6.3 (predictive control). Chapter 6 for a gain floor and ceiling on a plant with one unstable pole.

The app https://apmonitor.com/apps/inverted-pendulum/

See also Volume 3, Chapter 10 and Volume 3, Chapter 9 for a position loop that commands an angle and for a right-half-plane zero that bounds the bandwidth. Volume 9, Chapter 4 for an offset-free predictive controller with a disturbance estimate, and Volume 9, Chapter 2 for predictive control of a delayed integrating plant with a terminal constraint.

7.1 Introduction

Balancing is a control task that engineering keeps meeting in new machines. A two-wheeled robot carries its load above the axle [1], a walking pattern generator treats a biped as a pendulum standing on its support foot [2], a rocket in the first seconds of launch has to be held upright on its thrust before the airstream can steady it [3], and most of the corrections a person makes while balancing a stick on a fingertip are faster than the measured neural delay [4]. The inverted pendulum on a cart is the reduced model behind all of them, and its literature as a control benchmark reaches back to 1960 [5].

> ✦ DID YOU KNOW? **A Cart and Pole Standing In for a Rocket**
>
> In 2011 a space flight center used a laboratory cart and inverted pendulum to simulate a launch vehicle during liftoff. A fluid slosh tank, built to ride on the pendulum, was to test a controller's robustness to slosh that it does not know [6]. Volume 3, Chapter 10 lands a rocket with the same structure, a position loop that commands an attitude.

The pole draws the eye, and the natural conclusion is that a loop which keeps it upright has done the work. It has not. The cart has a position of its own that the pole angle does not reveal, and a loop closed on the angle alone will hold the pole steady over a cart that is creeping toward the end of its travel. In the simulator of this chapter, once that loop has caught a 2° start it keeps the pole within 1.53° of upright while rail friction accelerates the cart into a bumper at 4.11 s. The omission is as old as the apparatus, since two of the first textbook treatments stabilized the angle and left the cart free to drift [3].

The difficulty belongs to a class. One actuator serves two outputs. The pole is open-loop unstable at a rate set by gravity and geometry, and the transfer function from force to cart position carries a right-half-plane zero, so every move of the cart begins in the wrong direction. The unstable pole puts a floor under the angle gain, the lags of the drive and the sensors put a ceiling above it, and the zero limits how fast the cart may be steered [7, 8]. Delay narrows the window between floor and ceiling, which is why the time delays of the sensory and motor pathways limit human balance [9]. Balancing

robots, legged machines and vehicles standing on thrust share that shape, a fast loop against the fall nested inside a slower loop that places the base.

Pendulum rigs are a staple of control teaching [3], and assessments of control education keep returning to laboratory work and to the gap between theory and practice [10, 11]. However, a hardware rig is protected against exactly the failures that teach the most, such as a pole dropped by hand, a cart driven into its end stop and a loop that rocks after a change of pole. Each of those is a limit of the class rather than a mistake in the gains.

This chapter explains which configurations of a cart and pole loop settle and why, using a simulator described closely enough to be rebuilt from the text. The model, the cascade, the regulator and the energy swing-up are all established methods. Section 7.2 traces the history of the apparatus. Section 7.3 gives the nonlinear model, with rail impacts and a deck stop, and the chain of quantizers, delay and filtered differences through which the controller sees it. Section 7.4 derives three results. A closed-form characteristic polynomial of the cascade shows the cart loop raising the angle-gain floor from 11.3 to 17.4 N/rad and separates the two ways a new pole defeats an old tuning. A friction root makes an angle loop alone run away. An exact sampled-data linearization of the implemented loop gives a stable window that agrees with 39 of 40 nonlinear runs. Section 7.5 describes the five control structures, among them a discrete Riccati solution computed in the browser and a model predictive controller placed beside the regulator. That controller shares the regulator's model, weights and Riccati solution, and a completed square shows that the two coincide until a drive, rail or lean limit binds. A second setting removes its terminal weight, which leaves a finite-horizon law whose stability depends on the horizon, and Section 7.5.6 computes where it becomes stable. Section 7.6 runs the scenario sequence, in which each case changes one quantity. It measures what the limits are worth on this plant, as the difference between the regulator and the predictive controller (Section 7.6.8), and how long a horizon the nonlinear rig needs without the terminal weight (Section 7.6.9). Section 7.7 explains why a quantized loop replays exactly and yet depends on the last bit of its arithmetic. Section 7.8 describes the sensors, drives and controllers of real rigs, Sections 7.9 and 7.10 state the limitations and the conclusions, and the exercises follow in Section 7.11.

7.2 Balancing a Pole on a Moving Base

The first known solution was a bachelor's thesis written at M.I.T. in 1960 under the apt title "The Mechanical Seal" [3]. A broom stood on a platform driven through electromechanical integrators on a plotting table, and without position feedback the platform drifted. Its author estimated that a misalignment of one second of arc would carry the base to the end of its travel in about 100 s, and he added a position loop whose polarity was chosen to be positive, so that a base displaced to one side moves slightly further out, the broom tips inward, and the broom falls back toward the center of the table [3]. Work at Stanford on systems of several independent pendula followed in 1963, and its analysis arrived at the same positive feedback around the cart position [3].

By the end of the 1960s the single pendulum had entered the standard textbooks. Two of them stabilized the angle alone and left the cart position unregulated, one with a gain condition that matches the floor $(M+m)g$ derived in Section 7.4.2, while a third set the cart loop as an exercise whose answer is that the cart corrects a position error by first backing up [3]. Both observations reappear below, the first as the runaway of case 4 and the second as the wrong-way start of every cart move.

Swinging the pendulum up from hanging and then stabilizing it was demonstrated in 1976 [12]. Swing-up laws were developed further in the decades that followed [13], and the energy formulation, which drives the energy of the pendulum toward its upright value with a pivot acceleration signed by $\dot{\theta}\cos\theta$, gave the method the form used in this application [14]. The cart and pole also became a test problem for learning control, in which adaptive elements learned to balance it from a failure signal alone [15].

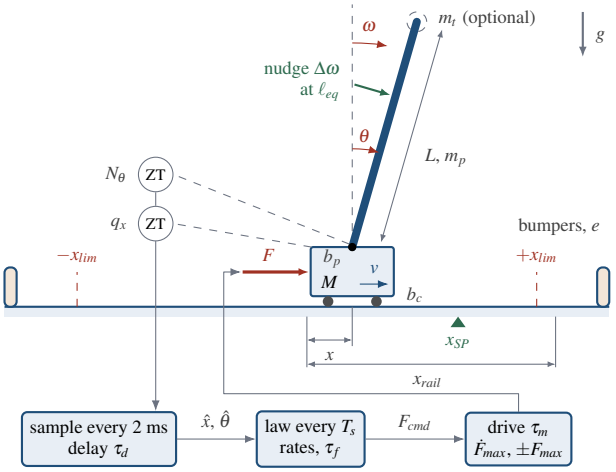


Figure 7.1 Cart and pole on a rail with bumpers, and the chain from the encoders to the drive. The cart of mass M on viscous friction b_c is pushed by the developed force F. The rod of length L and mass m_p, with an optional tip mass m_t, turns on a pivot with damping b_p, and its angle θ and rate ω are positive toward $+x$. A nudge changes ω by $\Delta\omega$ at the center of percussion, a distance ℓ_{eq} from the pivot. The cart center meets the bumpers at $\pm x_{rail}$, and they return it with restitution e, while the dashed lines at $\pm x_{lim}$ are the challenge limit. The encoders ZT, of resolution q_x and N_θ counts per turn, are sampled every 2 ms and delayed by τ_d, the law runs every T_s with rates filtered by τ_f, and the drive follows F_{cmd} with lag τ_m, slew limit $\dot{F}_{max}$ and limit $\pm F_{max}$.

> ⧖ HISTORY **Swinging Up by Energy, 1992 to 1996**
>
> Inverted pendulums had been classic tools in control laboratories since the 1950s, first to show how feedback stabilizes an unstable system [14]. A swing-up design published in 1992 noted that the optimal control found by the maximum principle is a function of time, applied as feedforward, so it is not robust and the transfer to upright is not assured [13]. An energy view, presented at a world congress in San Francisco in July 1996, treated the pendulum's energy as an integrator with varying gain, which loses control only when the pendulum is horizontal or reverses its swing. One ratio, the pivot's largest acceleration over gravity, sets the behavior: above four thirds one swing is enough, and below 0.577 it takes three swings or more. Energy control leaves the upright point a saddle, so another strategy must catch the pendulum [14].

The same structure moved into machines and into physiology. A two-wheeled mobile robot was designed and controlled as an inverted pendulum on a moving base [1], biped walking patterns are generated from a pendulum whose pivot is the support foot [2], and models of human balance use inverted pendula whose stabilization is limited by neural delay [4, 9]. Experiments on a cart and pole also found that a regulator designed with a viscous friction model left small oscillations that stick-slip friction explained [16], which bears on the friction model adopted below.

7.3 Cart and Pole Model

The apparatus in the application is a cart on a horizontal rail carrying a uniform rod on a free pivot, pushed by one force actuator, as shown in Figure 7.1. It is the upright cart of Section 1.4.1 with a force in place of a commanded acceleration, and the same cart with the pendulum hanging is the crane of Chapter 12. The cart position x is measured from the center of the rail and the pole angle θ from upright, both positive toward $+x$, so a pole leaning toward $+x$ has $\theta > 0$. The six states are the cart position and velocity v, the pole angle and rate ω, the drive force F as delivered so far, and a latch for a pole lying on the deck stop. Table 7.1 sorts the signals by role.

Table 7.1 Classification of the Signals

Symbol	Role	Nominal	Range
θ	CV, pole angle	0°	SP 0, band ±2°
x	CV, cart position	0 m	SP ±0.6, band ±0.05
F_{cmd}	MV, drive command	0 N	±20
x, v	states, cart	0	$\lvert x\rvert \le 1.0$ m
θ, ω	states, pole	0	full circle
F	state, developed force	0 N	±20
—	state, deck latch	0	0 or ±1
$\hat{x}$, $\hat{\theta}$	measurements	—	steps q_x, $2\pi/N_\theta$
$\hat{v}$, $\hat{\omega}$	rate estimates	0	filtered
$\Delta\omega$	DV, nudge	60 °/s	0 to 300
L, m_p	DV, rod	0.5 m, 0.15 kg	0.3–1.0, 0.05–1.0
$\theta(0)$	DV, initial tilt	0°	per scenario
θ_{lean}	output, lean asked	0°	±6°
E	output, pole energy	0 J	−0.736 J hanging
λ	output, unstable root	5.58 s^{-1}	—

7.3.1 Equations of Motion

Lagrange's equations for the cart and the pole, with the first moment $m\ell$ of the pole and its inertia J about the pivot, are given in Equation (7.1).

$$\begin{aligned}(M+m)\ddot{x} + m\ell\cos\theta\,\ddot{\theta} - m\ell\sin\theta\,\dot{\theta}^2 &= F - b_c\dot{x}\\ m\ell\cos\theta\,\ddot{x} + J\ddot{\theta} - m\ell g\sin\theta &= -b_p\dot{\theta}\end{aligned} \tag{7.1}$$

where $m = m_p + m_t$, $m\ell = m_pL/2 + m_tL$ and $J = m_pL^2/3 + m_tL^2$. The first row is a force balance in newtons and the second a moment balance about the pivot in newton meters. For the default rod $m\ell = 0.0375\,\mathrm{kg\,m}$ and $J = 0.0125\,\mathrm{kg\,m^2}$. Nothing is linearized. The two accelerations are solved from the 2×2 system at every evaluation of the right-hand side, and its determinant $\Delta(\theta) = (M+m)J - (m\ell\cos\theta)^2$ is never smaller than MJ, because $J \ge (m\ell)^2/m$ for any rigid pole, so the solve cannot fail anywhere on the circle that the swing-up visits. With no force and no friction the total energy $\frac{1}{2}(M+m)v^2 + m\ell v\omega\cos\theta + \frac{1}{2}J\omega^2 + m\ell g\cos\theta$ is conserved, and the verification suite of the application confirms it to one part in 10^6 over 10 s.

7.3.2 Drive, Rail and Deck Stop

The drive is a force source. The developed force follows the command through a first-order lag with a slew limit, as shown in Equation (7.2).

$$\dot{F} = \mathrm{sat}_{\dot{F}_{max}}\left(\frac{F_{cmd} - F}{\tau_m}\right), \qquad \lvert F_{cmd}\rvert \le F_{max} \tag{7.2}$$

where $\tau_m = 30\,\mathrm{ms}$ lumps the current loop, the motor and the belt, $\dot{F}_{max} = 1000\,\mathrm{N/s}$ and $F_{max} = 20\,\mathrm{N}$, which gives the bare cart $20\,\mathrm{m/s^2}$. The slew limit binds only when the command moves by more than $\dot{F}_{max}\tau_m = 30\,\mathrm{N}$, which within ±20 N requires a command swinging from near one limit toward the other, and the lag cannot carry F past a clipped command.

The cart meets rubber bumpers when its center reaches $x = \pm x_{rail} = \pm 1.0\,\mathrm{m}$. Contact is handled as a hard event between integration steps and never inside the derivative evaluation. A cart found beyond a stop is returned to it, and if it was moving into the stop its velocity is reversed with restitution $e = 0.4$, as shown in Equation (7.3).

$$\Delta v = -(1+e)\,v, \qquad \Delta\omega = -\frac{m\ell\cos\theta}{J}\,\Delta v \tag{7.3}$$

The second relation follows from applying the impulse to the cart alone through the mass matrix of Equation (7.1), so the pivot carries part of it into the pole, and a cart that strikes the right bumper throws its pole toward the right. For the default rod $m\ell/J = 3\,\mathrm{m}^{-1}$, so an arrival at 0.55 m/s turns the pole by 132 °/s. Below a resting-contact speed $v_{rest} = 0.02\,\mathrm{m/s}$ the restitution is zero, which also holds a cart that the drive keeps pressing into the stop, and with $e = 1$ an impact conserves kinetic energy exactly.

A pole that crosses $\pm 90^\circ$ outward lands on a stop on the cart deck, its rate is set to zero, and it then rides with the cart as one body of mass $M + m$. The stop is removed in the swing-up structure, whose pole must pass through the bottom. A nudge adds $\Delta\omega$ to the pole rate and leaves the cart velocity unchanged, which is what a tap at the center of percussion produces, a distance $\ell_{eq} = J/m\ell$ from the pivot ($2L/3 = 0.333\,\mathrm{m}$ for the bare rod). The pole energy relative to upright rest, Equation (7.4), is zero upright and $-2m\ell g = -0.736\,\mathrm{J}$ hanging.

$$E = \tfrac{1}{2} J\omega^2 + m\ell g\,(\cos\theta - 1) \tag{7.4}$$

7.3.3 What the Controller Receives

No control law reads the true state. Every $\Delta t_c = 2\,\mathrm{ms}$ the controller clock reads the two encoders, rounds the cart position to $q_x = 0.02\,\mathrm{mm}$ and the angle to $2\pi/N_\theta$ with $N_\theta = 4096$ counts per turn (0.088°), and pushes the pair into a delay line. When the law executes at t_k, once every sample interval T_s, it takes the sample from $t_k - \tau_d$, as shown in Equation (7.5), and forms the rates as filtered first differences, as shown in Equation (7.6).

$$\begin{aligned} \hat{x}_k &= q_x\,\mathrm{rnd}\left(\frac{x(t_k - \tau_d)}{q_x}\right), \\ \hat{\theta}_k &= \frac{2\pi}{N_\theta}\,\mathrm{rnd}\left(\frac{N_\theta\,\theta(t_k - \tau_d)}{2\pi}\right) \end{aligned} \tag{7.5}$$

$$\hat{v}_k = \hat{v}_{k-1} + a\left(\frac{\hat{x}_k - \hat{x}_{k-1}}{T_s} - \hat{v}_{k-1}\right) \tag{7.6}$$

where rnd rounds to the nearest integer, $a = 1 - \mathrm{e}^{-T_s/\tau_f}$, and $\hat{\omega}_k$ is formed from $\hat{\theta}$ in the same way. The filter time constant is $\tau_f = 15\,\mathrm{ms}$, so $a = 0.487$ at the default $T_s = 10\,\mathrm{ms}$, and one angle count differenced over one sample moves $\hat{\omega}$ by 0.075 rad/s. The delay is a multiple of the 2 ms clock between 0 and 120 ms, and the law holds its force between executions, so on average it sees the pole $\tau_d + T_s/2$ in the past. The encoders carry no random noise, and the application draws no random numbers, so the seed of Section 1.6.1 changes nothing in this app.

7.3.4 Parameters and Numerics

Table 7.2 lists every plant constant with the class the application gives it. Adjustable constants appear as sliders in the model dialog, the rod length and mass are disturbance sliders on the main panel, and hidden constants are fixed and shown only through their effect. The engine of Section 1.6.1 integrates the equations by the fourth-order Runge-Kutta method on a 0.5 ms step, the controller clock runs every 2 ms and the law every T_s, and the log samples at 50 Hz, with speeds from 0.1 to 2 seconds of simulated time per second of real time. The contact and deck corrections follow each integration step, before the outputs and the log sample. Halving the step to 0.25 ms moves the peak angle and the cart travel of case 1 by less than 0.5 %, which the verification suite checks.

7.3.5 Assumptions

The pole is a rigid uniform rod with viscous pivot damping. A real pole bends, and its first bending mode, far above the loop bandwidth for a short aluminum rod, approaches it for a long and light one.

Cart friction is viscous only. Stick-slip friction makes a balanced cart hunt in a small limit cycle about its target, which is what experiments on a cart and pole found when a regulator was designed

Table 7.2 Plant Parameters. Class a Is Adjustable in the Model Dialog, s Is a Disturbance Slider, h Is Hidden and Fixed.

Symbol	Meaning	Value	Range	
M	cart mass	1.0 kg	0.4–3	a
b_c	cart friction	1.0 N s/m	0–8	a
L	rod length	0.5 m	0.3–1.0	s
m_p	rod mass	0.15 kg	0.05–1.0	s
m_t	tip mass	0 kg	0–0.5	a
b_p	pivot damping	0.002 N m s/rad	0–0.05	a
g	gravity	9.81 m/s^2	1.6–24.8	a
F_{max}	drive force limit	20 N	5–40	a
τ_m	drive lag	30 ms	5–100	a
$\dot{F}_{max}$	drive slew limit	1000 N/s	100–5000	a
x_{rail}	bumper position	1.0 m	fixed	h
x_{lim}	challenge line	0.8 m	fixed	h
e	restitution	0.4	0–1	a
v_{rest}	resting-contact speed	0.02 m/s	fixed	h
q_x	cart encoder step	0.02 mm	0.001–1	a
N_θ	pole encoder counts	4096/turn	256–65536	a
$m\ell$, J	first moment, inertia	0.0375, 0.0125	derived	
ℓ_{eq}	$J/m\ell$	0.333 m	derived	
λ	unstable root	5.58 s^{-1}	derived	
z	wrong-way zero	5.35 rad/s	derived	

without it [16]. It would also alter the runaway of Section 7.4.2, because a lean too small to overcome breakaway would not move the cart.

The drive is a force source with one lag and one slew limit. Back electromotive force, current limits inside the amplifier, belt stretch and gear backlash are lumped into the lag, and a belt drive with backlash is harder to balance than this one.

The bumpers are instantaneous impacts with a fixed restitution, and the cart never lifts or rocks on its wheels. A real impact lasts milliseconds and can unseat the pole.

The encoders are exact apart from their resolution, with no noise and no missed counts, and the delay is a pure transport delay. A camera would add noise and dropped frames on top of the delay of case 8.

The regulator is designed with the true pole and redesigned at once after every change, which no engineer achieves in practice. Gravity, the masses and the frictions are constant during a run and the rail is level, so no constant force acts on the cart.

7.4 Linear Analysis of the Balancing Problem

Section 1.4 sets out the tools used here on a simple pendulum. Three models are kept apart below. The nonlinear equations of Section 7.3 are the simulation model, and the linearizations of this section serve analysis only. The controllers carry models of their own, the four-state linearization with friction on which the regulator is designed and the cart equation that the swing-up inverts, both evaluated with the true parameters rather than identified from data.

7.4.1 The Unstable Root and the Wrong-Way Zero

Linearized about upright rest with F as the input and no friction, Equation (7.1) becomes Equation (7.7).

$$\Delta_0\,\ddot{\theta} = (M+m)\,m\ell g\,\theta - m\ell F, \quad \Delta_0\,\ddot{x} = JF - (m\ell)^2 g\,\theta \tag{7.7}$$

where $\Delta_0 = (M+m)J - (m\ell)^2 = 0.01297\,\mathrm{kg^2\,m^2}$. The tilt grows at $\lambda_0 = \sqrt{(M+m)m\ell g/\Delta_0} = 5.71\,\mathrm{s^{-1}}$, and with the cart friction and the pivot damping the root is $5.58\,\mathrm{s^{-1}}$, one e-fold every 179 ms and a doubling every 124 ms. The remaining open-loop roots are 0 for the cart position, $-0.87\,\mathrm{s^{-1}}$ and $-5.86\,\mathrm{s^{-1}}$. Holding a tilted pole still requires accelerating the whole assembly at g per radian of tilt, so the force needed per radian is $(M+m)g = 11.3\,\mathrm{N/rad}$, a figure that recurs below as a gain floor.

The transfer function from force to cart position has the numerator of the pole equation, as shown in Equation (7.8).

$$\frac{x(s)}{F(s)} = \frac{Js^2 + b_p s - m\ell g}{D(s)}, \qquad z \approx \sqrt{g/\ell_{eq}} \tag{7.8}$$

where $D(s)$ is the open-loop characteristic polynomial. Its positive root z is 5.42 rad/s without pivot damping and 5.35 rad/s with it. It is the rate at which the pole would fall if the cart were clamped, since ℓ_{eq} is the length of the equivalent simple pendulum, and it lies only 4 % below the unstable root. The sensitivity function of any stable loop must equal one at a right-half-plane zero of the plant and vanish at a right-half-plane pole, so a maximum modulus argument bounds its peak below by $(z+\lambda)/(\lambda - z) = 46$ for a controller that measures only the cart position [7]. This is the bound of Section 1.4.7, which is infinite on the running example because a drive that sets the acceleration puts the zero on the pole. The force drive here separates them by only 4 %, so the bound is finite but far too large for a usable loop. Every structure in the application therefore reads the angle, and the cart position enters through it.

7.4.2 The Angle Loop Alone

With $F = K_{p\theta}\theta + K_{d\theta}\omega$ and the cart left out of the law, the linearized loop has the characteristic polynomial of Equation (7.9) when the small pivot damping is neglected.

$$s\big[\Delta_0 s^3 + (b_c J + m\ell K_{d\theta})s^2 + m\ell\,(K_{p\theta} - (M+m)g)\,s - b_c m\ell g\big] = 0 \tag{7.9}$$

Without friction the bracket reduces to $s[\Delta_0 s^2 + m\ell K_{d\theta} s + m\ell(K_{p\theta} - (M+m)g)]$, so the pole is held whenever $K_{p\theta} > (M+m)g = 11.3\,\mathrm{N/rad}$ and $K_{d\theta} > 0$, a floor that depends on the total mass and not on the length or the inertia of the pole. Chapter 6 derives the same kind of floor for a levitated ball. The cart is left with a double root at the origin and drifts. With friction the constant term is negative, so by Descartes' rule of signs one root is real and positive whatever the gains. For small friction it is close to $b_c g/(K_{p\theta} - (M+m)g)$, because a cart moving at speed v needs the force $b_c v$, which the loop can supply only by letting the pole lean forward by $b_c v/(K_{p\theta} - (M+m)g)$, and that lean accelerates the cart at g per radian. The estimate is $0.52\,\mathrm{s^{-1}}$ at the defaults, the exact root is $0.46\,\mathrm{s^{-1}}$, and it is $0.47\,\mathrm{s^{-1}}$ with the drive lag, the rate filter and the hold included. Friction, which damps most mechanical systems, destabilizes this one.

Pause. Try Exercise 7.1 before reading on.

7.4.3 The Cascade as Four-State Feedback

The cascade of Section 7.5 multiplies out, inside its lean clamp, to $F = K_{p\theta}\theta + K_{d\theta}\omega + k_x(x - x_{SP}) + k_v v$ with $k_x = K_{p\theta}K_{px} = 3.67\,\mathrm{N/m}$ and $k_v = K_{p\theta}K_{dx} = 5.24\,\mathrm{N\,s/m}$ at the defaults. With the frictionless model, that law closes into the quartic of Equation (7.10).

$$\frac{\Delta_0}{m\ell}s^4 + (K_{d\theta} - \ell_{eq}k_v)\,s^3 + (K_{p\theta} - (M+m)g - \ell_{eq}k_x)\,s^2 + gk_v s + gk_x = 0 \tag{7.10}$$

At the defaults its coefficients are 0.346, 3.25, 17.5, 51.4 and 36.0 in SI units, and its roots are $-0.97\,\mathrm{s^{-1}}$, $-2.04 \pm 4.52i\,\mathrm{s^{-1}}$ and $-4.37\,\mathrm{s^{-1}}$. Three consequences follow. The two lowest coefficients

Table 7.3 Dominant Modes of the Sampled-Data Linearization Against the Measured Behavior

Case, configuration	Mode (s^{-1})	Measured
1, cascade	-1.56	settles
1, regulator	$-1.96 \pm 1.61i$	settles
3, 75 N/rad	$+2.52 \pm 12.06i$	$\pm 16.7°$ at 2.10 Hz
3, 55 N/rad	$+0.55 \pm 10.80i$	$\pm 9.3°$ at 1.75 Hz
3, 40 N/rad	$-1.38 \pm 0.75i$	$\pm 0.08°$ residual
3, $\tau_m = 10$ ms	$+1.02 \pm 12.95i$	$\pm 8.8°$
4, angle loop alone	$+0.471$	speed grows at 0.47 s^{-1}
4, weak cart loop	$+0.024 \pm 0.531i$	growing swing, 11.9 s period
5, cascade	$+0.50 \pm 4.61i$	$\pm 8.3°$ at 0.70 Hz
5, 40 and 7 N s/rad	$-0.061 \pm 5.76i$	$\pm 1.00°$ from 30 to 60 s
5, soft cart loop	$-0.96 \pm 4.49i$	settles
6, cascade	$+0.022 \pm 3.92i$	$\pm 9.8°$ at 0.60 Hz
6, 40 N/rad	$-0.42 \pm 6.34i$	settles
8, cascade	$+1.05 \pm 8.04i$	falls at 6.84 s
8, 20 N/rad	$-0.26 \pm 2.92i$	holds
8, 15 N/rad	$+0.47 \pm 2.30i$	falls at 4.01 s

are positive only if k_x and k_v are, so the loop must push a cart that is right of its target further to the right, the polarity chosen deliberately in 1960 and the signature of the zero at z. The outer loop consumes the angle loop's damping in proportion to $\ell_{eq} k_v$ and its stiffness in proportion to $\ell_{eq} k_x$, so the Routh-Hurwitz conditions put the floor of the cascade at 17.4 N/rad rather than 11.3 N/rad. Because k_v grows with $K_{p\theta}$, the same conditions give the cascade a ceiling even without lags, at 73.2 N/rad. A changed pole also attacks a particular coefficient. A 1.0 m rod doubles ℓ_{eq} to 0.667 m and cuts the s^3 coefficient from 3.25 to 1.51, while a 1.0 kg rod raises $(M+m)g$ to 19.6 N/rad and cuts the s^2 coefficient from 17.5 to 9.16. More angle gain restores the second but not the first, because k_v grows with $K_{p\theta}$ in a cascade, while a softer cart loop restores the first.

7.4.4 The Sampled-Data Loop

Equations (7.9) and (7.10) leave out the drive lag, the rate filter, the hold and the delay. The linearized plant under the implemented law, without the quantizers, the clamps and the limits, is a linear periodic system, so its small-signal modes can be computed exactly. The plant with friction and the drive lag is discretized with a zero-order hold over one 2 ms clock period, the delay line and the filtered differences of Equation (7.6) are written as states, and the product over one sample interval is a transition matrix whose eigenvalues μ, mapped by $s = \ln(\mu)/T_s$, are the closed-loop modes. With the quantizers removed from the simulator and a 0.001 °/s tap that keeps the response linear, the model reproduces the growth and the frequency of the rocking mode at 55 N/rad (0.547 against 0.545 s^{-1}, at 1.718 Hz in both), of the heavy rod of case 6 (0.022 s^{-1} at 0.624 Hz in both) and of the stiffened loop on the long rod of case 5 (–0.061 s^{-1} at 0.916 Hz in both). Table 7.3 sets its predictions beside the measured outcomes.

At the default 10 ms sample interval and no delay the sampled cascade is stable from 18.6 to 50.7 N/rad, so the lags cut the lag-free ceiling of 73.2 N/rad by almost a third, while the lags and the friction raise the floor from 17.4 to 18.6 N/rad. At the 20 ms sample interval of case 8, the floor stays between 18.3 and 18.6 N/rad for every delay, 65 % above $(M+m)g$, while the ceiling falls from 47.8 N/rad with no delay to 38.6, 31.7, 26.3 and 21.6 N/rad at delays of 20, 40, 60 and 80 ms, as shown in Figure 7.2. The window closes between 94 and 96 ms, an average sensing lag of about 105 ms, or 0.59 of the e-folding time. Proportional feedback on the unstable mode alone keeps a window up to the bound $p\,\tau_d < 1$ of Section 1.4.7, with p the unstable root, which is a delay of one full e-folding time. The cascade, with its lags and its cart loop, loses the window well before that. Forty nonlinear runs with the 60 °/s tap agree with the window at 39 points. The exception is 20 N/rad

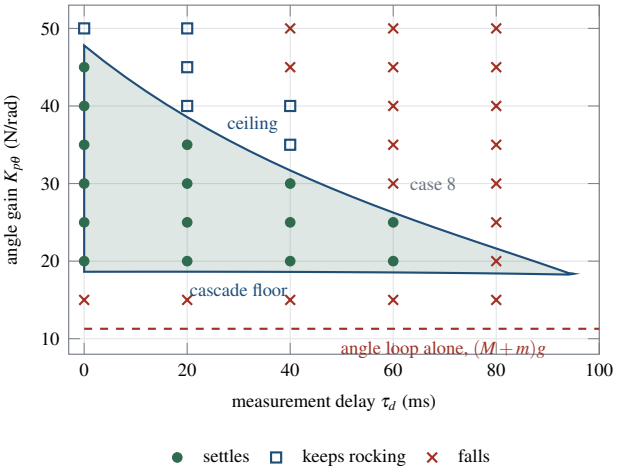


Figure 7.2 Stable window of the angle gain against the measurement delay for the default cascade at the 20 ms sample interval of case 8, from the sampled-data linearization. Markers are nonlinear runs with the 60 °/s tap, classified from 35 to 40 s.

at 80 ms, which is linearly stable with a pair at $-0.33 \pm 8.26i\,\mathrm{s}^{-1}$ and holds a 5 °/s tap, but which the full tap drives onto the lean clamp for 1.75 s and into the bumper at 4.77 s. A linear window is a necessary condition for recovering a disturbance and not a sufficient one.

7.5 Implemented Control Structures

7.5.1 The Sampled Law and the Drive Command

The law executes once every sample interval T_s, 10 ms by default and adjustable from 2 to 40 ms in steps of 2 ms, on the measurements of Equations (7.5) and (7.6), and it holds its force between executions. Every law computes a force toward $+x$. The command is clipped at $\pm F_{max}$, and the saturation flag compares the unclipped command with that limit and is cleared on every scan, so a law asking for 35 N reads as saturated while the developed force is still rising. A watchdog forces the command to zero once the measured angle reaches 90°, except in the swing-up. None of the five laws carries integral action, so the velocity-form PID of Section 1.6.2 is not used. There is no integrator to wind up, and the only limits are the command clamp, the drive slew limit of Equation (7.2) and, in the cascade, the lean clamp. The settings are collected in Table 7.4.

7.5.2 Angle Loop and Cascade

The angle loop pushes the cart toward the side to which the pole leans, as shown in Equation (7.11), with the defaults $K_{p\theta} = 30\,\mathrm{N/rad}$ and $K_{d\theta} = 5\,\mathrm{N\,s/rad}$.

$$F = K_{p\theta}\,\hat{\theta} + K_{d\theta}\,\hat{\omega} \tag{7.11}$$

The cascade wraps a slow cart loop around it that asks the pole for a lean toward the target, as shown in Equation (7.12).

$$\begin{aligned} \theta_{lean} &= \mathrm{sat}_{\theta_{lean,max}}\big(K_{px}(x_{SP} - \hat{x}) - K_{dx}\hat{v}\big) \\ F &= K_{p\theta}(\hat{\theta} - \theta_{lean}) + K_{d\theta}\,\hat{\omega} \end{aligned} \tag{7.12}$$

with $K_{px} = 7^\circ$ per meter, $K_{dx} = 10^\circ$ per m/s and $\theta_{lean,max} = 6^\circ$. The angle loop's only way to lean the pole toward the target is to pull the cart away from it first. This is the right-half-plane zero of Section 1.4.7, which asks for a fast loop on the angle inside a slow loop on the cart. The rocket of Volume 3, Chapter 10 has the same structure, a position loop that commands an attitude, and the

Table 7.4 Controller Settings

Symbol	Meaning	Default	Range
T_s	sample interval	10 ms	2–40
τ_d	measurement delay	0 ms	0–120
τ_f	rate filter	15 ms	0–50
$K_{p\theta}$	angle gain	30 N/rad	0–150
$K_{d\theta}$	angle damping	5 N s/rad	0–15
K_{px}	cart gain	7° per m	0–30
K_{dx}	cart damping	10° per m/s	0–30
$\theta_{lean,max}$	lean clamp	6°	1–20
x_s	regulator cart scale	0.5 m	0.05–2
θ_s	regulator angle scale	20°	1–45
F_s	regulator force scale	5 N	1–40
T_p	MPC prediction horizon	1 s	0.3–2
x_{max}	MPC rail limit	0.8 m	0.3–0.95
θ_{max}	MPC lean limit	15°	5–45
P	MPC terminal weight	LQR	LQR, Off
k_E	energy gain	100 m/s^2 per J	0–200
a_{sw}	pump acceleration limit	4 m/s^2	0.5–10
θ_{cap}	capture angle	20°	5–45
ω_{cap}	capture rate	180 °/s	30–600
$\Delta\theta_{rel}$	release margin	15°	5–45
ω_c, ζ_c	pump centering	1.2 rad/s, 0.9	fixed
F_{push}	manual push	2 N	0.5–20

same kind of zero. Written as $F = -K\tilde{x}$ with $\tilde{x} = (\hat{x} - x_{SP}, \hat{v}, \hat{\theta}, \hat{\omega})$, the cascade inside its clamp is $K = [-3.67, -5.24, -30, -5]$ in N/m, N s/m, N/rad and N s/rad, all four entries negative.

7.5.3 Discrete Linear Quadratic Regulator

The regulator applies $F = -K\tilde{x}$ with K from the discrete algebraic Riccati equation, the sampled counterpart of Equation (1.6), on the model linearized at upright. The design model is the four-state linearization with friction and without the drive lag, discretized with a zero-order hold at T_s through one matrix exponential of the augmented matrix $\left[\begin{smallmatrix} A & B \\ 0 & 0 \end{smallmatrix}\right] T_s$, evaluated by scaling and squaring with a 14-term Taylor series. The weights follow Bryson's rule [17] of Section 1.4.5, one over the square of the acceptable excursion, as shown in Equation (7.13).

$$Q = \operatorname{diag}\left(\frac{1}{x_s^2}, 0, \frac{1}{\theta_s^2}, 0\right), \qquad R = \frac{1}{F_s^2} \tag{7.13}$$

with $x_s = 0.5\,\text{m}$, $\theta_s = 20^\circ$ and $F_s = 5\,\text{N}$. The recursion $P \leftarrow Q + \Phi^T P \Phi - \Phi^T P \Gamma (R + \Gamma^T P \Gamma)^{-1} \Gamma^T P \Phi$ runs from $P = Q$ until the relative change is below 10^{-11}, which takes 697 sweeps at the defaults, and gives $K = [-9.26, -9.44, -48.60, -8.28]$. The single input makes the inverse a scalar division. The gain is recomputed whenever the rod length or mass, the tip mass, the cart mass, either friction, gravity, the sample interval or a weight changes, as if the engineer redesigned after every change with a perfect model of the new pole. A change of any of these constants is therefore a redesign for a different rig, and not a test of the regulator's robustness, and the same holds for the predictive controller and the swing-up catch, which receive this design, and for the pump force of Equation (7.15), which uses the same constants. The regulator's design poles are $-1.75 \pm 1.51i\,\text{s}^{-1}$ and $-6.53 \pm 3.21i\,\text{s}^{-1}$, and with the drive lag, the rate filter and the hold the dominant pair moves to $-1.96 \pm 1.61i\,\text{s}^{-1}$. The angle entry becomes −61.1 N/rad for the 1.0 m rod and −63.8 N/rad for the 1.0 kg rod.

7.5.4 Energy Swing-Up and Handover

The swing-up pumps the energy of Equation (7.4), evaluated on the measurements, toward zero. Following the energy-control law [14], the pump asks for a cart acceleration, as shown in Equation (7.14).

$$a = \mathrm{sat}_{a_{sw}}\big(k_E E\,\mathrm{sgn}(\hat{\omega}\cos\hat{\theta})\big) - \omega_c^2(\hat{x} - x_{SP}) - 2\zeta_c\omega_c\hat{v} \tag{7.14}$$

with $k_E = 100\,\mathrm{m/s^2}$ per J, $a_{sw} = 4\,\mathrm{m/s^2}$ and a centering spring of $\omega_c = 1.2\,\mathrm{rad/s}$ and $\zeta_c = 0.9$ that keeps the pumping cart off the bumpers. At rest hanging, where $\hat{\omega}\cos\hat{\theta}$ is zero, the sign is taken as positive so that the swing starts itself. The acceleration becomes a force through the cart equation solved at the measured state, as shown in Equation (7.15).

$$F = \frac{\Delta(\hat{\theta})\,a + m\ell\cos\hat{\theta}\,(m\ell g\sin\hat{\theta} - b_p\hat{\omega})}{J} + b_c\hat{v} - m\ell\hat{\omega}^2\sin\hat{\theta} \tag{7.15}$$

This cancels the reaction of the swinging pole on the cart and the cart friction that would otherwise absorb the pumped energy. The regulator takes over when $|\hat{\theta}| < \theta_{cap} = 20^\circ$ and $|\hat{\omega}| < \omega_{cap} = 180^\circ/\mathrm{s}$, and the pump resumes only past $\theta_{cap} + \Delta\theta_{rel} = 35^\circ$, so the handover cannot chatter.

7.5.5 Model Predictive Control on the Regulator's Model

The fifth structure is a model predictive controller (MPC), an option that no case before case 10 selects. At each decision it plans the force over a horizon of $T_p = 1\,\mathrm{s}$, applies the first move and plans again from the next measurement, the receding-horizon law of Section 1.6.3, whose constrained form, with its stability and optimality conditions, is reviewed in [18]. Its model is that of the linear quadratic regulator (LQR) of Section 7.5.3. The matrices Φ, Γ, K and P and the weights Q and R of Equation (7.13) are handed to it from the same design call, so a changed pole reaches both controllers at once and both are given the present plant. Nothing is identified. This departs from the general rule of Section 1.6.3, under which a predictive controller never reads the parameters of the run, and it makes a changed pole a redesign rather than a model error. The MPC reads what the regulator reads and nothing else, the state $\tilde{x}$ formed from the delayed, quantized readings and filtered rates of Equations (7.5) and (7.6), the cart target and the drive limit F_{max}. The plant with its drive lag, the instruments with their quantizers, delay line and filters, and the controller's model, the four-state linearization held at T_s without the drive lag, therefore remain three separate things.

The decisions are offsets from the regulator's move, $F_k = -K\tilde{x}_k + c_k$, so the free variables perturb a fixed state feedback, a structure used in [19]. With the force carried as a fifth state, $\xi_k = (\tilde{x}_k, F_{k-1})$, the plan predicts the loop that the regulator closes, as shown in Equation (7.16). Chapter 8 plans departures from its regulator's force in the same way.

$$\xi_{k+1} = \begin{bmatrix} \Phi - \Gamma K & 0 \\ -K & 0 \end{bmatrix} \xi_k + \begin{bmatrix} \Gamma \\ 1 \end{bmatrix} c_k \tag{7.16}$$

where k counts intervals of T_s from the present state $\tilde{x}_0$ and $N = T_p/T_s$ steps, 100 at the defaults, span the horizon. The prediction thus starts on the stable closed loop rather than on the root at $+5.58\,\mathrm{s^{-1}}$, and the predicted force is an output that a limit can act on. The objective is the regulator's cost over the horizon with P as the terminal weight, plus the prices of the limits described below, as shown in Equation (7.17).

$$J = \sum_{k=0}^{N-1}\left(\tilde{x}_k^T Q\tilde{x}_k + RF_k^2\right) + \tilde{x}_N^T P\tilde{x}_N + \sum_i\left(\rho_i\varepsilon_i^2 + \sigma_i\varepsilon_i\right) \tag{7.17}$$

where $\varepsilon_i \geq 0$ are the slacks of the three limits and ρ_i and σ_i their quadratic and linear prices. The state term at $k = 0$ is fixed by the measurement and is omitted in the implementation. With $W = R + \Gamma^T P \Gamma$ and the gain $K = W^{-1} \Gamma^T P \Phi$, every state and force satisfies Equation (7.18).

$$\tilde{x}^T Q \tilde{x} + RF^2 + (\Phi\tilde{x} + \Gamma F)^T P (\Phi\tilde{x} + \Gamma F) - \tilde{x}^T P \tilde{x} = W(F + K\tilde{x})^2 \quad (7.18)$$

Expanded, the left side is $\tilde{x}^T (Q + \Phi^T P \Phi - P)\tilde{x} + 2F\Gamma^T P \Phi \tilde{x} + WF^2$. The Riccati equation turns its first term into $W(K\tilde{x})^2$ and the gain turns its second into $2WFK\tilde{x}$. Summed along any predicted trajectory, the P terms telescope, and the cost without slacks becomes Equation (7.19) for every horizon and every sequence of offsets.

$$J_0 = \tilde{x}_0^T P \tilde{x}_0 + W \sum_{k=0}^{N-1} c_k^2 \quad (7.19)$$

Three consequences follow. With no limit binding the best offsets are zero for any horizon and any blocking, so the first move is the regulator's, and the regulator is this MPC with its limits removed. The Hessian in the offsets is diagonal, $2Wn_j$ for an offset held over n_j steps. With only the hard limit on the first move, the plan is the regulator's move clipped at $\pm F_{max}$ with every later offset zero. Within its limits the plan is a finite-horizon, move-blocked form of the constrained linear quadratic regulator [20]. At the defaults $W = 0.04667$ against $R = 0.04$. Equation (7.19) holds to a relative error of 2.3×10^{-10} on 200 random states and offset sequences, the first move equals $-K\tilde{x}$ to 1.7×10^{-9} N over the 1599 plans of 16 s of case 1, and the clipped move is reproduced to 4.7×10^{-9} N.

Pause. Try Exercise 7.3 before reading on.

A plan is made at the regulator's own sample instants but at most every 10 ms, once in $n_d = \lceil 10\,\text{ms}/T_s \rceil$ executions, and between plans each execution applies $-K\tilde{x}$ plus the plan's offset for that instant. The offsets are held over six blocks of n_d times 1, 2, 4, 8 and 16 steps and the remainder of the horizon, which are 1, 2, 4, 8, 16 and 69 steps at the defaults. The horizon is 1.75 time constants of the slowest design pole, $-1.75 \pm 1.51i\,\text{s}^{-1}$, and P prices what lies beyond it. Horizons from 0.5 to 2 s stop the cart of case 10 within 1 mm of 0.768 m, and 0.3 s lets it run to 0.819 m.

Two hard rows hold the first move inside the drive limit, $|-K\tilde{x}_0 + c_0| \leq F_{max}$, with F_{max} read as the drive clamp reads it. Three limits are soft, each with one slack over the whole horizon, and they are judged every 50 ms of the plan, as shown in Equation (7.20).

$$|F_k| \leq F_{max} + \varepsilon_F, \qquad |x_k| \leq x_{max} + \varepsilon_x, \qquad |\theta_k| \leq \theta_{max} + \varepsilon_\theta \quad (7.20)$$

where $x_k = x_{SP} + \tilde{x}_{1,k}$ and θ_k are the predicted cart position from the center of the rail and the predicted angle. The rail limit $x_{max} = 0.8$ m is the challenge line, 0.2 m inside the bumpers, and the lean limit $\theta_{max} = 15°$ keeps the plan where the linearization holds, since up to 15° the sine of the angle stays within 1.1 % of the angle. The force is priced at $\rho_F = 10^4$ per N^2 and $\sigma_F = 10^4$ per newton, so its limit binds wherever it can. The rail and the lean are priced at $\rho = 10^3$ per m^2 or rad^2 and $\sigma = 10^2$ per meter or radian, which is 250 and 122 times the regulator's own weights on the cart and the angle. These prices and θ_{max} were chosen once, on case 10, from ρ in decades from 10 to 10^5 with $\sigma = \rho/10$ and θ_{max} of 15, 20, 30 and 45°, as the smallest cart excursion among the 14 of 20 settings that keep the pole up and the cart off the bumpers, while $\rho \leq 10^2$ behaves as the regulator. Each plan is a quadratic program in nine variables, the six offsets and three slacks, with 125 inequality rows at the defaults, solved by the dual active-set method of Section 1.6.3 with a ridge of $10^{-9}(1 + 2RN)$ on the diagonal of the offsets.

Offset-free predictive controllers usually add integrating disturbances to their model [21], as the Kalman filter of Section 1.6.3 does and as the boiler of Volume 9, Chapter 4 does for an unmeasured

fuel bias. This one has none, because no law of the application integrates, no constant force acts on the cart, upright rest on the target with zero force is an exact equilibrium of both the plant and the model, and an estimator would give the MPC dynamics that the regulator lacks and void Equation (7.19). The offset that results is shared. A constant force $d = 1\,\mathrm{N}$, added to the cart outside the application, leaves the cart 107.9 mm from the target under the regulator and 107.7 mm under the MPC, against $d/|K_1| = 108.0\,\mathrm{mm}$ from the cart entry of the gain.

A plan is discarded if the quadratic program ends in any status other than optimal, if the force it would apply at once is undefined or infinite, or if that force exceeds F_{max} in magnitude. The regulator then keeps the drive with the remaining offsets of the last accepted plan, alone once that plan has run out, a mark is written to the record, and control passes back to the MPC at the next accepted plan. Forced at each of the 101 decisions of case 10 between 2.00 and 3.00 s, a rejection leaves the command 0.21 N from the MPC's own on average and 2.97 N at worst, against 0.94 and 6.91 N for the bare regulator. No shipped case rejects a plan. The MPC also plans in manual operation. Handed the captured pole by the swing-up in the same capture region, it catches the pole of case 9 at 2.67 s, as the regulator does. Because a plan never asks past $\pm F_{max}$, the saturation flag counts the time the MPC's command sits at the limit.

7.5.6 Terminal Cost and the Stability of a Short Plan

The terminal weight is a setting of the MPC with two values. LQR, the default and the controller described above, prices the end of the plan with the regulator's P. Off gives the last step the stage weight instead, so that without slacks the plan minimizes Equation (7.21).

$$J_{\mathrm{off}} = \sum_{k=0}^{N-1} \left(\tilde{x}_k^T Q \tilde{x}_k + R F_k^2 \right) + \tilde{x}_N^T Q \tilde{x}_N \tag{7.21}$$

Nothing else changes. The model of Equation (7.16), the offsets, the blocks, the three soft limits, the hard rows on the first move and the rejection rule are the same objects under both values. The telescoping sum behind Equation (7.19) needs P at the last step, so under Off the zero offset is no longer optimal. With no limit binding the first planned force is still linear in the state, $F_0 = -K_H \tilde{x}_0$, with a gain K_H that depends on the horizon and the blocking, and the loop that the plan's own model predicts under that law is Equation (7.22).

$$\tilde{x}_{k+1} = (\Phi - \Gamma K_H)\, \tilde{x}_k \tag{7.22}$$

The receding-horizon law is stable on its model when the spectral radius $\rho(\Phi - \Gamma K_H)$, the largest eigenvalue magnitude, is below one.

With P at the last step $K_H = K$ at every horizon. The terminal weight is the value of the infinite-horizon problem from $\tilde{x}_N$, so the plan is the regulator's infinite-horizon problem cut at N with nothing lost, which is how a terminal cost equal to the cost-to-go makes a receding-horizon law stabilizing for stable and unstable linear plants alike [18, 22]. A condensed solution of the plan written independently of the application, with explicit predictions over the graduated blocks, gives $|K_H - K| \le 6.8 \times 10^{-9}$ for every N from 20 to 200 at $T_s = 10\,\mathrm{ms}$, and the application's own law agrees with that solution under Off to 2.3×10^{-6} in every gain entry. Without P the cart entry of the gain changes sign. At $T_p = 0.3\,\mathrm{s}$, $K_H = [+1.03, -1.45, -28.66, -4.36]$ against the regulator's $[-9.26, -9.44, -48.60, -8.28]$, so the plan holds the pole and pushes a displaced cart farther out, and $\rho = 1.00606$ per 10 ms, a real cart mode growing at $0.60\,\mathrm{s}^{-1}$.

The radius falls as the horizon grows, as shown in Figure 7.3. It is 1.00350 at 0.4 s, 1.000008 at 0.45 s and 0.99832 at 0.46 s, where the law first becomes stable, and every longer horizon to 2 s stays stable, with 0.99704 at 0.5 s and 0.98673 at 1 s against the regulator's 0.98261. With the drive lag and the rate filter added to the loop, as in the sampled-data model of Section 7.4.4, the radius moves by at most 2.0×10^{-3} and the crossing stays at 0.46 s, about 2.6 e-folding times of the unstable root and 0.8 time constants of the slowest design pole. A long enough horizon is the other route to

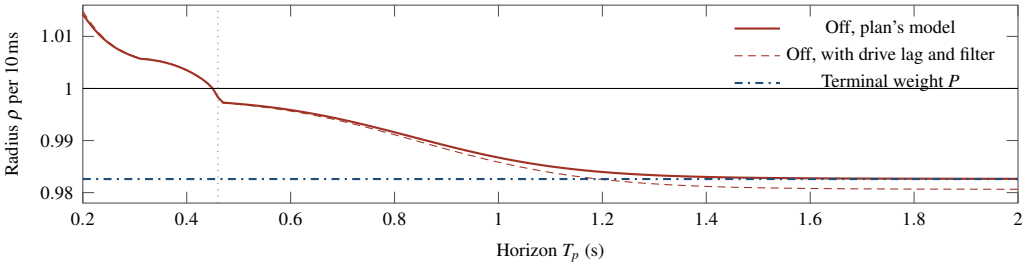


Figure 7.3 Spectral radius of the predictive law with no limit binding against the horizon, at $T_s = 10$ ms with the weights of case 10. Without the terminal weight (thick solid on the plan's model, dashed with the drive lag and the rate filter added) the law is unstable below the dotted line at 0.46 s. With it (dash-dotted) the law is the regulator's at every horizon. The thin solid line is a radius of one.

stability without a terminal cost [23], and 0.46 s is that length for this law. The pure finite-horizon law, with one-step blocks instead of the graduated ones, is a different gain, that of the backward Riccati recursion from $P_N = Q$. Its radius is 1.0105 at 0.3 s, and it crosses one at 0.48 s.

The linear crossing is necessary for a hold and not sufficient. Just past it the slowest mode is the cart, decaying at 0.296 s^{-1} at 0.5 s, a time constant of 3.4 s, so a push that sends the cart toward the end of the rail reaches the bumper before that mode brings the cart back. The drive limit and the 30 ms lag, which the plan's model lacks, take further margin, and the nonlinear rig needs a longer horizon than the linear crossing, as Section 7.6.9 measures.

7.5.7 Manual Operation

In manual operation the command is the drive-force slider, ±20 N in 0.5 N steps, plus pushes of $F_{push} = 2$ N that last 50 ms of simulated time from the buttons, or as long as an arrow key is held, so a push is the same at every speed. In automatic operation the slider follows the command, so a switch to manual hands over the force being applied. The law keeps executing in manual operation. With no integrator there is nothing to initialize, and on the switch to automatic the law's latest force replaces the operator's, softened only by the drive lag. The transfer is bumpless at the upright rest state that the verification suite checks, and not in general.

7.6 Results Across the Scenario Sequence

Eleven cases and a challenge make up the sequence in Table 7.5. Each of the first nine departs from case 1 in one setting, case 10 changes the structure and the push, and case 11 runs the controller of case 10 on the push of case 7 with a 0.3 s horizon and no terminal weight. The values in Table 7.6 were measured by running the published engine headlessly on each scenario, with the one change named made after the scenario is loaded and before the run starts, as a learner makes it, and reading the scorecard and the trajectories at the 0.5 ms physics step. The engine draws no random numbers, so every value repeats exactly on one JavaScript engine. The scenarios, their completion rules and the scorecard are those of Section 1.6.4.

7.6.1 Balance, the Wrong-Way Start and a Human

Case 1 runs the cascade at its defaults. The 60 °/s tap at 2 s leans the pole 4.68° toward $+x$ at 2.12 s, and the cart loop, braking the cart as it runs out, holds its lean request on the −6° clamp for 0.29 s. The pole overshoots that request to −6.98° at 2.66 s, the cart reaches 0.292 m on a 5.77 N peak, the pole is back inside ±2° 1.06 s after the tap and the cart is within 2 cm of center 2.46 s after it. At 8 s the target steps to 0.5 m. The lean request jumps to 3.5° and rises to 4.86° while the cart backs 21.3 mm the wrong way in the first 0.28 s, and the cart then reaches 90 % of the move 1.90 s after the step with a 1.0 mm overshoot (Table 7.6), and the cart trend of Figure 7.4 shows the move. The

Table 7.5 Scenario Sequence

#	Scenario	Change from case 1	Lesson
1	Tuned balance	—	catch, wrong-way start
2	Balance it yourself	manual, 1° start	the growth rate
3	Too much angle gain	$K_{p\theta} = 75\,\mathrm{N/rad}$	the ceiling
4	The angle loop alone	no cart loop, 2°	friction runaway
5	A longer pole	$L = 1.0\,\mathrm{m}$	damping lost
6	A heavier pole	$m_p = 1.0\,\mathrm{kg}$	stiffness lost
7	A hard push	tap 150 °/s	recoverable push
8	Delayed sensing	T_s 20, τ_d 60 ms	delay closes the window
9	Swing-up from hanging	hanging start	capture, handover
10	LQR against MPC on the rail	MPC, tap 205 °/s	worth of the limits
11	Terminal cost at a short horizon	MPC 0.3 s, Off, tap 150 °/s	worth of P
C	Challenge	5°, 125 °/s at 3 s	a limit off the CV

Table 7.6 Outcomes Read From the Running Engine

Case and condition	Result
1, tap 60 °/s at 2 s	4.68°, then −6.98°, 0.292 m, 5.77 N
1, back in ±2° / cart in 2 cm	1.06 s / 2.46 s after the tap
1, target 0 to 0.5 m at 8 s	21.3 mm back, 90 % at 1.90 s, 1.0 mm over
1, regulator	0.188 m, 0.83 s, 35.6 mm back, 1.42 s
2, no force, L 0.3 / 0.5 / 1.0 m	down at 0.741 / 0.944 / 1.326 s
2, automatic from 1°	peak 1.005°, 0.022 m
2, switch to automatic at 0.5 s	peak 10.8°, 0.464 m
3, 75 N/rad	±16.7° at 2.10 Hz, 43.3 % saturated
3, 55 / 40 N/rad	±9.3° / ±0.08°
3, $K_{d\theta}$ 8 / τ_m 10 ms	±0.08° / ±8.8°
4, angle loop alone	bumper at 4.11 s, down at 4.54 s
4, cascade / weak cart loop	0.046 m / bumper at 61.8 s
5, cascade	±8.3° at 0.70 Hz, first swing 0.970 m
5, 40 and 7 / soft cart loop	±1.00° / bumper at 3.28 s
5, regulator	1.01 s, 0.461 m
6, cascade	±9.8° at 0.60 Hz, cart ±0.178 m
6, 40 N/rad / regulator	3.28 s, 0.295 m / 0.87 s, 0.194 m
7, cascade	bumper at 2.88 s, down at 3.29 s
7, 15° clamp / regulator	0.745 m / 0.495 m, 19.0 N
7, largest tap without a bumper	130 / 183 / 229 °/s
8, cascade	bumper at 6.53 s, down at 6.84 s
8, $\tau_d = 20\,\mathrm{ms}$	peak 7.6°, 0.321 m
8, 40 / 20 / 15 N/rad	down 3.11 s / 14.9° / down 4.01 s
8, regulator / MPC	down 2.98 s / down 3.00 s
9, a_{sw} 4 / 6 / 2 m/s^2	2.67 s / 2.06 s / never caught
10, regulator / MPC	0.809 m, line at 2.46 s / 0.768 m
11, Off / terminal P	down 3.04 s / 0.509 m
C, shipped cascade	0.927 m, held 24.1 s, fails
C, regulator	0.392 m, held 25.4 s, passes
C, 40 and 6 N s/rad	0.721 m, passes

regulator on the same case catches the tap with 0.188 m of travel instead of 0.292 m, back inside ±2° in 0.83 s with an 8.48 N peak, and it makes the move faster, 90 % in 1.42 s, at the price of a deeper wrong-way start of 35.6 mm and a 10.0 mm overshoot. A faster cart loop pays for its speed in the depth of the wrong-way start, which is the right-half-plane zero at work.

🕮 **Pause.** Try Exercise 7.2 before reading on.

In case 2 the operator holds the drive with the pole 1° from upright. With no force applied the pole



Figure 7.4 The inverted pendulum application in scenario 1, Tuned balance, captured at $t = 12.00$ s with the cascade selected, a 10 ms sample interval and no measurement delay. The figure is a composite of the screen: the badge row and gauges, the scene and the three trends, with the header, the control panel and the scorecard cut. (1) The timing budget reads an e-fold of 179 ms against a sensing lag of 5 ms. (2) The cart stands at 0.498 m beside the target marker at 0.5 m with the pole upright. (3) The dashed ± 0.8 m challenge lines lie inside the bumpers at the ends of the rail. (4) The trends run from 2 to 12 s. At their left edge the 60 °/s tap leans the pole about 4.7° toward $+x$, the dashed lean request on the angle strip is pinned at −6° while the cart, pushed out to 0.29 m on a 5.8 N peak, is braked, and the pole swings to about −7°. (5) At 8 s the target steps to 0.5 m, the lean request jumps to 3.5°, and the cart dips below zero before it runs to the target. The force command carries spikes of a few tenths of a newton from single angle counts. The badge row reads 0.498 m of maximum travel and no rail hits.

lies on the deck at 0.944 s, at 1.326 s for a 1.0 m rod and at 0.741 s for a 0.3 m rod. At full speed every correction has to be seen, decided and applied inside one e-fold of 179 ms, while at the 0.1 speed setting the same fall lasts 9.44 s of wall time. A 2 N push lasting 50 ms, applied to the pole at rest upright, leaves it turning at 19.9 °/s toward $-x$ and the cart moving at 0.087 m/s 0.2 s after it starts. In automatic operation the cascade catches the same start with a 1.005° peak and 0.022 m of travel. A switch to automatic 0.5 s into the unaided fall, with the pole at 8.35° and 46 °/s, still recovers it with a 10.8° peak and 0.464 m of travel, and the latest switch that recovers it without touching a bumper is at 0.576 s, with the pole at 12.7°.

7.6.2 Floor, Ceiling and Delay

Case 3 raises the angle gain from 30 to 75 N/rad. The angle loop alone would stay stable at this gain, with its pendulum pair at $-2.42 \pm 17.42i\,\mathrm{s}^{-1}$ in the sampled-data model. In the cascade the cart loop's force gains grow with $K_{p\theta}$, so the s^3 coefficient of Equation (7.10) falls from 3.25 to 0.64 and

even the lag-free loop has a pair at $+0.14 \pm 13.21i\,\mathrm{s}^{-1}$. With the lags the pair is $+2.52 \pm 12.06i\,\mathrm{s}^{-1}$, and between 15 and 30 s the pole rocks $\pm 16.7°$ at 2.10 Hz, with the command beyond the 20 N limit 43.3 % of the time, the lean request on its clamp 73 % of the time and the cart shaking ± 0.129 m. Because the angle loop alone is stable, the clamp, which cuts off the outer loop, is what bounds the oscillation. At 55 N/rad the lag-free loop is stable at $-1.40 \pm 10.47i\,\mathrm{s}^{-1}$ and the sampled loop is not, at $+0.55 \pm 10.80i\,\mathrm{s}^{-1}$, so the pole rocks $\pm 9.3°$ at 1.75 Hz without saturating. Lowering the gain to 40 N/rad settles the loop to a $\pm 0.08°$ residual, and raising the damping to 8 N s/rad at 75 N/rad restores the s^3 coefficient to 3.64 and settles it to the same $\pm 0.08°$. A drive three times faster, 10 ms, does not, leaving $\pm 8.8°$. With the drive lag as the only lag, a 5 ms drive would bring the continuous loop to the edge at $+0.02 \pm 13.1i\,\mathrm{s}^{-1}$, but the rate filter and the hold keep the sampled loop unstable at $+0.52 \pm 13.0i\,\mathrm{s}^{-1}$, and the pole still rocks $\pm 7.9°$.

Case 8 replaces the encoder angle with a camera-like sensor that delivers a reading every 20 ms, 60 ms late, so the law sees the pole 70 ms in the past on average. The model puts an unstable pair at $+1.05 \pm 8.04i\,\mathrm{s}^{-1}$. After the tap the oscillation grows with a period of 0.68 s to a 32.7° swing, with the drive saturated 29 % of the time, the cart reaches the right bumper at 6.53 s, and the pole falls at 6.84 s. With a 20 ms delay the loop holds with a 7.6° peak and 0.321 m of travel. At 60 ms, 40 N/rad falls at 3.11 s, 20 N/rad holds with a 14.9° peak and 0.633 m of travel, and 15 N/rad falls at 4.01 s. The last failure is not a delay effect. The same gain falls at 5.03 s with the delay set to zero, because 15 N/rad is below the 18.6 N/rad floor that the cart loop imposes on the cascade although it is above the static floor of 11.3 N/rad.

7.6.3 The Cart Runaway

In case 4 only the angle loop runs, from a 2° start. The catch peaks at 2.01°, and from 1 s until the cart first meets the right bumper the pole leans between 0.38 and 1.53° forward, while the cart speed grows from 0.128 to 0.520 m/s between 1 and 4 s, a rate of $0.47\,\mathrm{s}^{-1}$ that matches the friction root of Section 7.4.2. The cart passes 0.5 m at 2.93 s and strikes the bumper at 4.11 s at 0.55 m/s. The 0.80 N s impulse through the pivot turns the pole by 132 °/s, the cart meets the bumper four more times in the next 0.14 s, and the pole falls at 4.54 s. The cascade from the same start keeps the cart within 0.046 m. A cart loop weakened to 1° per meter and 2° per m/s does not overcome the runaway. Its linearization has a pair at $+0.024 \pm 0.531i\,\mathrm{s}^{-1}$, and the cart swings with an 11.9 s period and a growing amplitude, 0.596 m by 40 s, crosses the 0.8 m line at 50.5 s, meets the bumper at 61.8 s and loses the pole at 62.5 s.

> **TRY IT IN THE APP** **4 · The Angle Loop Alone**
>
> Load the scenario and press Run. Watch the pole angle first: after the catch it never leans more than 1.53° forward. Then watch the cart. It creeps to the right, speeds up, strikes the right bumper at 4.11 s, and the jolt throws the pole down at 4.54 s.

7.6.4 A Longer and a Heavier Pole

Case 5 doubles the rod to 1.0 m at the same 0.15 kg. The unstable root falls from 5.58 to $3.98\,\mathrm{s}^{-1}$, so an unaided fall from 1° takes 1.326 s instead of 0.944 s, and yet the loop tuned for the short rod no longer settles. After the tap the cart swings out to 0.970 m, and from 20 to 60 s the pair rocks $\pm 8.3°$ at 0.70 Hz with the cart swinging ± 0.171 m and the lean request on its clamp 41 % of the time, against a predicted pair at $+0.50 \pm 4.61i\,\mathrm{s}^{-1}$, or 0.73 Hz. A stiffer angle loop of 40 N/rad and 7 N s/rad leaves $-0.061 \pm 5.76i\,\mathrm{s}^{-1}$, a mode that decays by a factor of e only every 16 s, so the pole stays inside $\pm 2°$ only from 15.12 s after the tap and still rocks $\pm 1.00°$ between 30 and 60 s. A softer cart loop of 4° per meter and 7° per m/s moves the pair to $-0.96 \pm 4.49i\,\mathrm{s}^{-1}$ and ends the rocking, but the tap then carries the cart into the bumper at 3.28 s. The regulator, redesigned for the new rod, is back inside $\pm 2°$ 1.01 s after the tap with 0.461 m of travel.

DID YOU KNOW? A Slower Pole the Loop Cannot Hold

From the simulator: a rod twice as long falls more slowly, reaching the deck at 1.326 s instead of 0.944 s with no force applied. Yet the cascade tuned for the short rod no longer settles, and the pole rocks ±8.3° at 0.70 Hz (Table 7.6).

Case 6 keeps the length and raises the rod mass to 1.0 kg. The unstable root moves only from 5.58 to 6.71 s^{-1}, but $(M+m)g$ rises from 11.3 to 19.6 N/rad and the ideal cascade now needs $K_{p\theta}$ above 29.5 N/rad, so the shipped 30 N/rad leaves a pair at $-0.042 \pm 3.99i\ s^{-1}$ that the lags push to $+0.022 \pm 3.92i\ s^{-1}$. The pole rocks ±9.8° at 0.60 Hz with the cart swinging ±0.178 m from 20 to 60 s. At 40 N/rad it is back inside ±2° 3.28 s after the tap with 0.295 m of travel, and the redesigned regulator takes 0.87 s and 0.194 m. The two cases show the same rocking with different causes, the damping coefficient in case 5 and the stiffness coefficient in case 6, and only the second answers to more angle gain.

7.6.5 The Recoverable Push

Case 7 raises the tap to 150 °/s. The cascade runs out of rail. For 0.77 s of the 0.88 s between the push and the first impact its lean request sits on the −6° clamp, with no drive saturation, and the cart strikes the right bumper at 2.88 s, meets it five more times by 3.04 s, and loses the pole at 3.29 s. With a 15° clamp the cart stops at 0.745 m after a 17.9° peak. The regulator, which has no clamp and weighs the cart position directly, stops it at 0.495 m with a 19.0 N peak and 0.07 s of saturation. The largest tap each configuration recovers without touching a bumper, found by bisection on the tap size, is about 130 °/s for the cascade, 183 °/s with the 15° clamp and 229 °/s for the regulator, and on the 5 °/s steps of the slider 130, 180 and 225 °/s are the largest that pass. The clamp decides the cascade's limit and the rail decides the regulator's, which from 230 °/s strikes a bumper without losing the pole and from 250 °/s loses it as well (beyond 245.2 °/s by bisection).

7.6.6 Swing-Up

Hanging at rest the pole is 0.736 J short of upright. The pump, limited to 4 m/s^2, needs three turning points, and the regulator captures the pole at 2.67 s with the cart never more than 0.496 m from center and a 6.07 N force peak, then returns the cart to center without a release. With a 2 m/s^2 limit the swings stall about 27° short of upright, just outside the capture wedge, and the pole is never caught in 30 s. With 6 m/s^2 it is caught at 2.06 s, but the cart travels 0.743 m and the force reaches 9.18 N. A 5° capture wedge, or an energy gain of 20 m/s^2 per J, never catches it, and the pump keeps swinging the pole.

TRY IT IN THE APP 9 · Swing-Up From Hanging

Load the scenario and press Run, and watch the pole energy gauge and the capture wedge above the pivot. After three turning points the regulator captures the pole at 2.67 s. Press Reset, set Pump acceleration limit to 2 m/s^2 and run again: the swings stall about 27° short of upright, just outside the wedge, and the pole is never caught in 30 s.

7.6.7 The Graded Challenge

The challenge starts the pole 5° from upright and pushes it at 125 °/s at 3 s. Over 30 s the cart must never reach ±0.8 m or touch a bumper, the pole must stay up, and it must be held inside ±2° for at least 20 s. Every change is made before the run, and a changed model parameter or disturbance voids the attempt. The shipped cascade holds the angle for 24.1 s but carries the cart to 0.927 m, 0.127 m past the line, which it crosses at 3.89 s, a pass on the angle and a failure on the hard limit. The suggested regulator, which uses the default weights ($x_s = 0.5$ m, $\theta_s = 20°$, $F_s = 5$ N), settles 1.63 s after the push, holds for 25.4 s, 5.4 s more than required, and keeps the cart within 0.392 m, 0.408 m inside the line, with 0.03 s of saturation. It spends more actuator travel to do so, 138.1 N of

Table 7.7 Regulator and MPC With One Model and One Set of Weights, Runs of 30 s

Case	Law	θ_{pk} (°)	RMS (°)	x_{pk} (m)	t_{lim} (s)	Travel (N)	t_{set} (s)	Plans
5, long pole	LQR	11.9	1.41	0.461	0	167.5	1.01	—
	MPC	11.9	1.41	0.461	0	160.2	1.01	3000
7, 150 °/s	LQR	20.0	2.07	0.495	0.07	129.5	1.85	—
	MPC	17.6	1.98	0.504	0.08	138.3	1.97	3000
8, delay	LQR	fell 2.98 s		1.000	0.66	67.6	—	—
	MPC	fell 3.00 s		1.000	0.68	68.3	—	154
10, 205 °/s	LQR	31.2	3.22	0.809	0.15	142.5	2.04	—
	MPC	28.0	2.97	0.768	0.24	152.7	2.20	3000
C[a]	LQR	16.2	1.72	0.392	0.03	138.1	1.63	—
	MPC	15.1	1.67	0.396	0.03	146.2	1.75	3000

θ_{pk} is the largest angle after the push and x_{pk} the largest cart distance from center, both from the true state at the physics step, t_{lim} the time with the command at ±20 N, Travel the summed change of the command, and RMS and t_{set} the scorecard's angle RMS and recovery inside ±2°. Plans are those of the MPC, none rejected. In case 8 both carts reach the bumper. [a]Challenge, for information.

summed force change over the 30 s run against 81.7 N for the shipped cascade. The cascade with 40 N/rad and 6 N s/rad also passes, at 0.721 m.

7.6.8 Regulator Against Predictive Control

Table 7.7 sets the regulator beside the MPC on case 10 and on the three earlier cases in which a limit or a model error could separate them, the long pole of case 5, the hard push of case 7 and the delay of case 8, with the challenge for information. Each case is loaded, the structure is selected before the run, and every run lasts 30 s. The regulator keeps its shipped weights, which are also the suggested setting of the challenge, and the MPC keeps the settings of case 10 in every case. Both read the same quantized instruments, and the application draws no random numbers, so both meet the same push at the same instant. The lean limit and the prices were chosen on case 10 itself, and its push was chosen as the first step of the 5 °/s push slider past the largest push that the regulator holds inside the line, the step at which the two laws differ in kind.

In case 10 the 205 °/s push at 2 s carries the regulator's cart across the line at 2.46 s to 0.809 m, while the MPC stops it at 0.768 m, as shown in Figure 7.6, and Figure 7.5 sets the two laws side by side on every case of the table. Both laws drive at +20 N from 2.022 s to get the cart under the pole. The first plan that the rail limit shapes comes at 2.060 s, 0.40 s before the regulator's cart reaches the line. The MPC then holds +20 N to 2.18 s against 2.17 s and brakes at −20 N from 2.29 to 2.37 s, where the braking force of the regulator peaks at 16.8 N, and the pole leans back 28.0° instead of 31.2°. Only 64 of the 3000 plans have a limit binding. The cost is a recovery 0.16 s slower and 10.2 N more force travel. The plans are not forecasts. A plan made at 2.17 s put the cart at 0.918 m, 118 mm past the line on its slack, and the cart stopped at 0.768 m, because each plan is replaced 10 ms later by one made from a new measurement, and the model leaves out the drive lag, the slew limit, the delay of the readings and the sine of the lean.

The pushes over which only the MPC keeps the cart inside the line form a narrow band. On the 5 °/s steps of the push slider the largest push held inside the line rises from 200 to 205 °/s, and by bisection from 203.6 to 209.6 °/s. By bisection the largest push kept off the bumpers rises from 229.4 to 230.2 °/s, and the largest push that keeps the pole from 245.2 to 248.6 °/s. At 240 °/s the rail slack of the plan reaches 414 mm, and both carts meet the bumper once and keep the pole. The travel after a push is set mostly by the push and the 20 N drive, because the cart must run under the pole and then lean it back to stop.

Where no limit binds the two laws are one law. In case 5 no limit shapes any plan, and the two commands agree to 3.8×10^{-9} N until the first encoder reading differs at 5.05 s. From there the quantizer dither of Section 7.7 gives the two runs different count crossings and a force travel of 160.2 against 167.5 N, while the peak angle, the RMS, the excursion and the recovery agree to the digits printed. On the rig of case 10, pushes of 60 to 100 °/s reproduce the regulator's excursion to 0.1 mm.

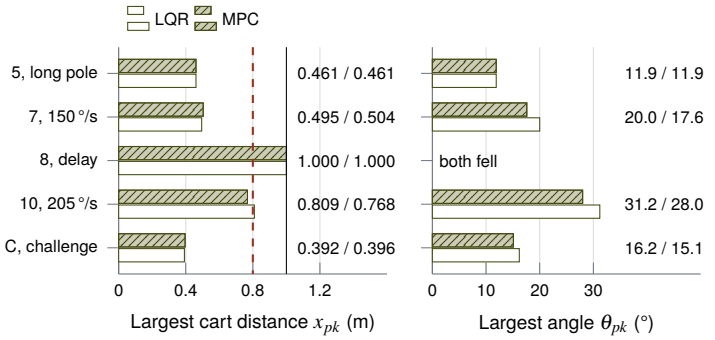


Figure 7.5 The regulator (white) and the predictive controller (hatched) of Table 7.7, with one model and one set of weights: the largest cart distance from center against the 0.8 m line (dashed) and the bumper at 1.0 m (thin solid), and the largest angle after the push. Values read LQR / MPC. Notice where the bars part: the predictive controller keeps the cart of case 10 inside the line, gives up 8.5 mm of travel to its lean limit in case 7, and falls with the regulator in case 8.

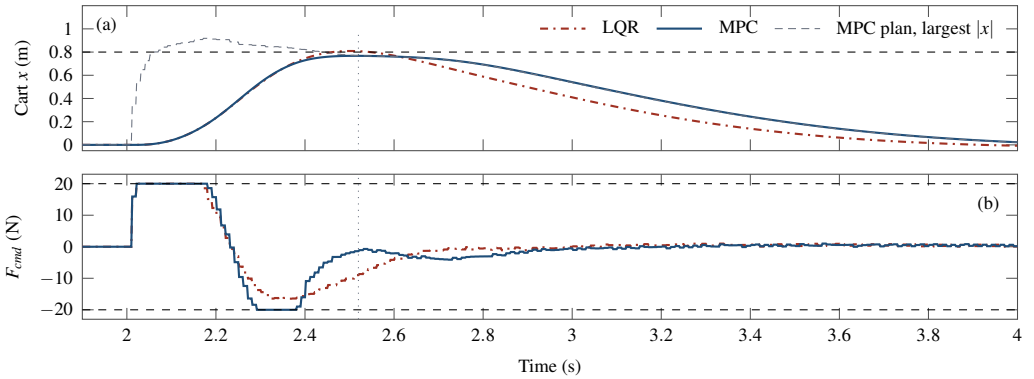


Figure 7.6 Case 10 under the regulator and under the MPC, which share one model and one set of weights, from 1.9 to 4.0 s with the 205 °/s push at 2 s, the regulator dash-dotted and the MPC thick solid. (a) Cart position against the 0.8 m line (thin dashed), with the largest cart position in each plan of the MPC (gray, densely dashed). (b) Force command against the 20 N drive limits (thin dashed). At the dotted line, $t = 2.52$ s, the regulator's cart is at 0.808 m past its peak of 0.809 m and the MPC's at 0.767 m.

The MPC loses case 7 on travel and recovery while it leans less. The cart never nears the line, and only the 15° lean limit binds, on the backward lean that brakes the cart. The MPC leans back 17.6° against 20.0°, with 9.4° forward in both runs, so it brakes the cart less, and the cart runs to 0.504 m against 0.495 m and recovers in 1.97 against 1.85 s. With $\theta_{max} = 45°$ the MPC reproduces the regulator's 0.495 m and 1.85 s. The same mechanism carries the MPC's cart 0.8 to 10.3 mm further than the regulator's for pushes of 120 to 180 °/s in steps of 20 °/s. Neither law survives case 8, whose 60 ms delay and 20 ms samples are in neither model. Both carts cross the line at 2.87 s and strike the bumper, the regulator loses the pole at 2.98 s and the MPC at 3.00 s. On the challenge both pass, with 0.392 and 0.396 m of travel and the pole held for 25.4 and 25.3 s.

The 30 ms drive lag, which neither model contains, is also why the limits cannot be priced higher. At ten times the shipped rail and lean prices the cart reaches 0.813 m with 0.53 s at the drive limit, and at a hundred times the pole falls at 2.68 s. With the drive lag cut to 5 ms the hundredfold prices keep the pole and stop the cart at 0.767 m, but the regulator on that faster drive stops it at 0.740 m, so

Table 7.8 The 150 °/s Push of Case 11 With and Without the Terminal Weight, Runs of 15 s

Law	T_p (s)	Pole		θ_{pk} (°)	x_{pk} (m)	t_{lim} (s)	Plans
Cascade	—	fell	3.29 s	—	1.000	0.30	—
Angle PD	—	fell	2.98 s	—	1.000	0.22	—
LQR	—	held		20.0	0.495	0.07	—
MPC, P	0.3	held		17.4	0.509	0.07	1499
MPC, P	0.5, 0.6, 1	held		17.6	0.504	0.08	1499
MPC, Off	0.3	fell	3.04 s	—	1.000	0.24	303
MPC, Off	0.5	fell	7.54 s	—	1.000	0.43	754
MPC, Off	0.6	held		11.9	0.793	0.00	1499
MPC, Off	1	held		15.9	0.547	0.05	1499

Columns as in Table 7.7, with θ_{pk} for the runs that hold. Every fall follows bumper contact, first at 2.88, 2.76, 2.79 and 6.61 s. No plan is rejected.

a quicker actuator buys more rail than the predictive controller does. Case 10 makes 3000 plans in 30 s, and 48 runs at the corners of the horizon, the two limits of the plan and the sample interval, on cases 1 and 10 and the challenge, reject no plan and need at most 28 iterations of the solver. A plan takes about 90 µs in Node.js and 105 µs in the Chrome browser, the difference between whole runs of case 10 under the MPC and under the regulator divided by the plans, or 1.05 % of one processor core at real time. At $T_s = 2$ ms with $T_p = 2$ s a plan has 1000 steps and takes 0.92 ms, 9.1 % of a core at real time and 18.3 % at twice real time.

7.6.9 Terminal Cost at a Short Horizon

Case 11 is the rig of case 1 with the 150 °/s push of case 7 at 2 s, under the MPC with the settings of case 10 except a horizon of 0.3 s, the shortest the application offers, and no terminal weight. Table 7.8 sets every shipped structure, the MPC with P and the MPC without it on that push, each run for 15 s from the same instant with the same instruments, the terminal weight and the horizon being the only settings varied. The regulator alone is the shipped structure given the same signal, because its cost-to-go is the terminal weight.

As loaded, the plan without P catches the push with at most 14.2 N before the cart reaches the bumper, against the 20 N that the plan with P applies, and it never leans the pole back to brake. The cart reaches 1.70 m/s, the rail limit first shapes a plan at 2.38 s, and from then on paying the rail slack inside 0.3 s costs the plan less than a backward lean, so at 2.50 s its predicted pole ends 11.4° forward. The cart strikes the right bumper at 2.79 s, the impacts throw the pole, and it falls at 3.04 s. Each of its 303 plans was optimal and accepted, because the plan solved the problem that it was given. Restoring P at the same horizon holds the pole with 0.509 m of travel, against 0.495 m for the bare regulator. The difference is the 15° lean limit, which shapes its plans from 2.13 to 2.43 s, and with a 45° limit the plan reproduces the regulator's 0.495 m. The cascade and the angle loop lose the same push at 3.29 and 2.98 s.

Longer horizons without P hold the pole only well past the linear crossing of Section 7.5.6. On case 11's push, 0.45 s falls at 3.65 s, 0.46 s at 5.20 s and 0.5 s at 7.54 s, after the cart creeps into the bumper at 6.61 s, while 0.55 s holds with 0.823 m of travel. Run for 30 s at every horizon of the setting from 0.3 to 2 s in steps of 0.1 s, the shortest horizon from which Off holds the pole at that and every longer horizon is 0.5 s for pushes of 60 and 100 °/s, 0.6 s for 150 and 205 °/s and 0.9 s for 240 °/s. With P every horizon holds every one of these pushes, with one to three bumper contacts at 240 °/s, where the bare regulator also strikes the bumper once. Halving the physics step to 0.25 ms leaves every verdict and every travel unchanged and moves the fall at 0.5 s from 7.54 to 7.57 s.

The hold at 0.3 s is therefore owed to the terminal weight. The constraints are not the cause, because Off keeps them and falls, and the horizon is not the cause, because with P it hardly matters, at 0.509, 0.504 and 0.504 m for 0.3, 0.6 and 1 s. The terminal weight is the regulator's cost-to-go, so the gain of the plan with P over the plan without it is the regulator's. A horizon long enough to do

without *P* buys stability and not the regulator's performance. Off carries the cart 0.793 m at 0.6 s and 0.547 m at 1 s, against 0.504 m with *P*.

The same holds across the whole sequence. Over the ten earlier cases and the challenge, with the challenge counted once at its shipped settings and once at the suggested ones, twelve runs of 30 s for each structure, the pole falls in 11 of 12 runs under the angle loop, 4 under the cascade, 2 under the regulator, 2 under the MPC with *P* and 2 under Off at the shipped 1 s horizon, the two being the manual case and the delay of case 8 for the last three, and in 11 under Off at 0.3 s, which loses every case that any law balances. The swing-up start counts as no balancing case for these laws, because the watchdog holds the drive at zero while the pole hangs. Off at 1 s holds every case that the MPC with *P* holds but carries the cart farther in each, 0.539 m against 0.510 m in case 1, 0.432 against 0.396 m on the challenge, and 0.823 against 0.768 m in case 10, across the 0.8 m line, with a smaller peak lean, and in case 8 it falls at 2.94 s against 3.00 s. Off is thus less safe than the shipped plan wherever the two differ, and the application states this beside the setting.

The terminal weight costs nothing to compute, because it changes the weights of the last step and not the size of the quadratic program. In Node.js, measured while other simulations loaded the computer, a plan of case 11 takes 32 µs as loaded and 59 µs at 0.6 s, and case 10 takes 95 and 96 µs with and without *P* at 1 s. The cost grows with the horizon. Timed in Chrome while the application animates itself under the same load, the whole controller of case 10 as shipped occupies 2.6 to 3.2 % of one processor core at 1× speed and 4.8 to 5.7 % at 2× over two passes. The 2 s horizon without *P* takes 4.8 to 5.2 % and 8.7 to 9.3 %, close to the limit of 10 % of one core at the top speed that the apps are held to, and with *P* at 2 s it takes 4.3 to 5.1 % and 7.5 to 9.8 %, so the cost belongs to the horizon and not to the terminal weight.

7.7 Repeatability of a Quantized Loop

Reset (Section 1.6.1) restarts a case from $t = 0$ with the settings in force and replays the run bit for bit on one JavaScript engine, and yet the encoder quantizers make the fine detail of a run sensitive to arithmetic. Perturbing the case 1 tap by one to twenty parts in 10^{13}, in either direction, moves the cart position at 12 s between 0.4962 and 0.4984 m and the summed force change between 37.2 and 40.0 N. A last-bit difference changes the sample at which the angle crosses a count boundary, and the rate estimate turns a single count into a kick of 0.37 N through $K_{d\theta}$. Linear feedback of quantized measurements on an unstable plant is known to produce pseudorandom rather than convergent trajectories [24], and here the result is a deterministic dither that keeps the angle within 0.14° of upright and the command within ±0.42 N at rest. The screenshot behind Figure 7.4 was captured by stepping the same engine in a browser, and its scorecard, cut from the figure, matches the headless run behind the tables to the last digit, with 0.498 m of travel and a summed force change of 39.85 N, because the browser and the headless run perform the same arithmetic.

7.8 Technology in Practice

What Measures It

Laboratory carts and poles read both outputs with quadrature encoders, one on each cart and one on the pendulum [6]. A rig that must also see a flexing pole adds a rate gyro at its tip, because an angle read at the base alone can match a pole that bends and a pole that does not. Comparing the tip's rate with the angle at the base then reveals the flex [25]. The chapter's encoders round the cart to 0.02 mm and the pole to 4096 counts a turn, carry no noise and miss no count, and the delay line of case 8 stands in for a camera (Section 7.3).

TECHNOLOGY Sensor: Encoders Give Position, Not Speed

On one laboratory rig at a space flight center, every cart and the pendulum carry a high-resolution quadrature encoder. The encoders measure position only, so a state estimator supplied the velocities that full state feedback needs [6]. A second rig read its pole with a rate gyro instead, on ranges from ± 250 to $\pm 2000\,°/s$, where a wider range costs precision. Its angle came from integrating the rate, and the gyro's bias showed at rest as a steady walk to $-10°$ after 25 s, cut to $\pm 1°$ once the bias was subtracted [25]. The chapter's encoders are exact to one count and drift-free, and its rates are differences filtered over $\tau_f = 15\,\text{ms}$.

What Moves It

The motor that moves the cart is a direct-current motor on one rig and servos turning the wheels of a mobile base on another. The first rig commands its direct-current motor with a voltage and converts volts to newtons with a hardware gain identified on the rig [6]. The second commands continuous-rotation servos whose signal drops out past the ends of its range, so a shaping function maps the controller's command onto the usable band on each side of the dead band [25]. The chapter's drive is the idealized end of that chain, a force source with a 30 ms lag, a slew limit of 1000 N/s and a limit of 20 N, and its Assumptions name the backlash and back electromotive force that it lumps into the lag.

TECHNOLOGY Final Element: A Voltage-Driven Motor With a Dead Band

The space-center rig drives its carts with a direct-current motor commanded by a voltage, and a hardware gain turns volts into newtons in the model. The gain found by experiment proved wrong once the feedback gains were tried, and the pendulum stayed up only about ten seconds while the gain was checked [6]. On a wheeled rig a band of commands around neutral gives no motion, the band shifts from time to time, and a routine steps the command until the motor encoders report motion [25]. On a trend, friction of this kind shows as a cart that waits until the pole has fallen far enough, then moves, in a repeating oscillation [26]. The chapter's drive is a lagging force source that lumps all of this.

Where the Controller Runs

On the space-center rig a data acquisition board first ran the control algorithm. It was replaced by a graphical programming environment on a computer with interface cards, and a design script discretized the two-cart model at a sampling period of 0.001 s [6]. The chapter's law runs every 10 ms on a 2 ms clock. The state feedback $F = -K\tilde{x}$, with the lean clamp of the cascade or the plan of the predictive controller around it, is the one block that holds the chapter's structure, and the swing-up adds a switch on the capture region. Figure 7.7 draws the loop on rig hardware.

What Is Modernizing It

The first development is sensing of the pole's flexure. A wheeled rig with a flexible aluminum pole, built as an analogue of a launch vehicle, added the tip gyro, and work continued on sensor blending and direct output adaptive control to cancel the flex dynamics. Its models were not yet accurate enough to move a simulated controller onto the hardware [25]. The second is the safety filter. A model predictive safety certification changes the input of a learned controller only when the cart and pole approach its constraints, and on a simulated cart and pole it kept a learned controller inside constraints that the controller broke alone [27]. The chapter's predictive controller acts in the same way, departing from the regulator only when a limit binds. The third is open benchmarks. Simulated cart and pole environments now randomize inertial properties and initial states and set state and input constraints, and a reinforcement learner trained there used over nine hours of simulated data [27].

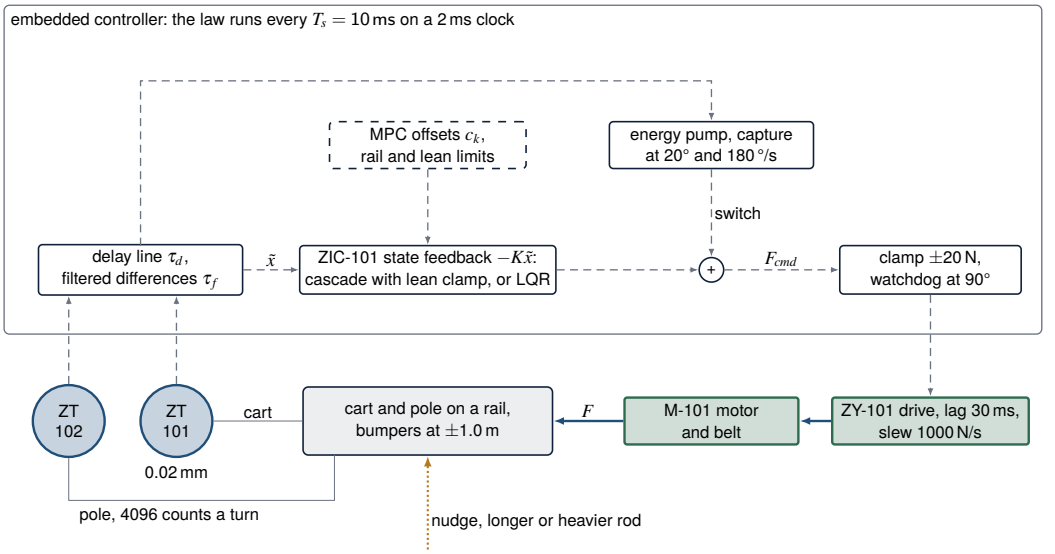


Figure 7.7 The balance loop on rig hardware. The encoders ZT-101 on the cart and ZT-102 on the pole feed a delay line and filtered differences, which form the state $\tilde{x}$. The state feedback ZIC-101 holds the cascade or the regulator, the predictive controller (dashed) adds offsets to the regulator's move, and in the swing-up the energy pump drives until the pole enters the capture region. The clamped command goes to the drive ZY-101 and the motor M-101. Notice that one force serves both outputs, so every block above the drive shares one limit of 20 N.

> **IN PRACTICE** **A Fast Pair for the Pole, a Slow Pair for the Cart**
>
> A textbook design for a cart and pendulum splits the closed loop into fast dynamics that hold the pendulum upright and slower dynamics that place the cart. The fast pair is set from the natural frequency of the pendulum hanging down, with a damping ratio of 0.5. That first design gave the desired step response but asked for a peak force almost three times the force of gravity, so the fast pair was slowed by about a factor of three and the cart was moved more slowly still [26]. The peak force, not the step response, decided the redesign. The chapter's cascade, a fast angle loop inside a slow cart loop, makes the same split.

7.9 Limitations

The rod is rigid, the friction is viscous and the encoders are noise-free, so the small limit cycles that stick-slip friction produces on real carts do not appear [16].

The drive is a lagging force source. A belt or rack drive with backlash, back electromotive force and a voltage limit would add lag and lower the ceiling below the 50.7 N/rad that the present lags allow.

None of the laws has integral action, and no constant force acts on the cart, so the steady offset that a tilted rail would cause is never exposed.

The regulator is designed with a perfect model of the pole and redesigned instantly, which makes cases 5 and 6 more favorable to it than a real redesign would be.

The linear analysis leaves out the quantizers, the clamps and the limits, so its window is necessary and not sufficient for recovery, as the exception in Figure 7.2 shows.

OPEN CHALLENGE **Certifying Where a Learned Controller Is Safe**

Learning-based control is often certified with a region of attraction, the states from which the controller is guaranteed to bring the system back to its equilibrium. A review of safe learning finds that the guarantees rest on assumptions, such as bounded disturbances, hard to verify before a robot runs. It calls for systematic ways to check them with little data [27]. Here the linear window agrees with 39 of 40 nonlinear runs, and the exception shows it necessary for recovery, not sufficient.

The predictive controller plans on the regulator's model, linear at upright and without the drive lag, the slew limit, the measurement delay or the sine of the lean, and it is given the regulator's instant redesign after every change of pole. Its plans are therefore not forecasts, and its limits priced a hundred times higher ask for force changes that the 30 ms drive delivers too late. It has no disturbance estimator, its rail and lean limits are soft and judged every 50 ms of the plan, and its prices were chosen on case 10, the case it wins.

The crossing at 0.46 s is a linear, unconstrained result on the plan's own model at the default sample interval, weights and pole, and it moves when any of them changes. It is a lower estimate of the horizon that a push needs without P and not that horizon, because the bumpers, the drive limit and the lag lie outside the analysis. The application offers no terminal constraint set, the other ingredient of the usual stability proof for constrained plans [18], so the argument for the plan with P covers only states from which no limit is active at the end of its horizon.

7.10 Conclusion

A cart and pole is an unstable plant with one actuator, two outputs and a right-half-plane zero, and the simulator of this chapter puts each of those properties within reach of an experiment. Its unstable root of $5.58\,\mathrm{s}^{-1}$, one e-fold every 179 ms, leaves a hand on the drive that much time for each correction at full speed and ten times as long at the 0.1 speed setting. An angle loop alone needs more than $(M+m)g = 11.3\,\mathrm{N/rad}$, and with rail friction it runs the cart into a bumper 4.11 s after a 2° start while the pole leans no more than 1.53°. The cascade that cures the runaway starts every move the wrong way, 21.3 mm for a 0.5 m step, and its characteristic polynomial explains why its floor is 17.4 N/rad and why one tuning rocks a 1.0 m rod through its damping and a 1.0 kg rod through its stiffness. An exact sampled-data linearization agrees with 39 of 40 nonlinear runs and closes the gain window at about 105 ms of average sensing lag. The regulator stops a 150 °/s push at 0.495 m where the clamped cascade loses the pole, and the energy pump brings the hanging pole to capture in 2.67 s. A predictive controller on the regulator's model and weights equals the regulator until a limit binds, stops the cart of case 10 at 0.768 m where the regulator's crosses the line to 0.809 m, gives up 8.5 mm of travel to its own lean limit in case 7, and falls with the regulator in case 8. Without its terminal weight the same plan is an unstable law below a 0.46 s horizon and loses at 0.3 s the 150 °/s push that it holds with the weight, while a horizon long enough to hold without it carries the cart farther.

The model is rigid, viscous and noise-free. A fuller model would add stick-slip friction, encoder noise, drive backlash and a regulator designed on a mismatched pole. Within those bounds the application is a laboratory for balancing an underactuated machine within hard limits.

KEY POINTS

- A pole on a cart falls on its own, and only moving the cart under it holds it up.
- Rail friction runs the cart of an angle loop alone into a bumper 4.11 s after a 2° start.
- The cascade that cures the runaway starts every move the wrong way, 21.3 mm for a 0.5 m step.
- The energy pump brings the hanging pole to capture in 2.67 s.
- A predictive controller on the regulator's model equals the regulator until a limit binds.

Where else this applies. Volume 3, Chapter 10 meets the same wrong-way start in a rocket.

References

[1] F. Grasser, A. D'Arrigo, S. Colombi, and A. C. Rufer, "JOE: a mobile, inverted pendulum," *IEEE Transactions on Industrial Electronics*, vol. 49, no. 1, pp. 107–114, 2002.

[2] S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa, "The 3D linear inverted pendulum mode: a simple modeling for a biped walking pattern generation," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, 2001, pp. 239–246.

[3] K. H. Lundberg and T. W. Barton, "History of inverted-pendulum systems," *IFAC Proceedings Volumes*, vol. 42, no. 24, pp. 131–135, 2010.

[4] J. L. Cabrera and J. G. Milton, "On-off intermittency in a human balancing task," *Physical Review Letters*, vol. 89, no. 15, p. 158702, 2002.

[5] O. Boubaker, "The inverted pendulum benchmark in nonlinear control theory: A survey," *International Journal of Advanced Robotic Systems*, vol. 10, no. 5, p. 233, 2013.

[6] C. L. Wilson and P. Capo-Lugo, "Robust control algorithm for a two cart system and an inverted pendulum," NASA Marshall Space Flight Center, Huntsville, AL, Internship final report NTRS 20110015819, 2011. [Online]. Available: https://ntrs.nasa.gov/citations/20110015819

[7] 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.

[8] K. J. Åström, "Limitations on control system performance," *European Journal of Control*, vol. 6, no. 1, pp. 2–20, 2000.

[9] G. Stépán, "Delay effects in the human sensory system during balancing," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1891, pp. 1195–1212, 2009.

[10] J. A. Rossiter, B. Pasik-Duncan, S. Dormido, L. Vlacic, B. Jones, and R. Murray, "A survey of good practice in control education," *European Journal of Engineering Education*, vol. 43, no. 6, pp. 801–823, 2018.

[11] B. W. Bequette, "Process control practice and education: Past, present and future," *Computers & Chemical Engineering*, vol. 128, pp. 538–556, 2019.

[12] S. Mori, H. Nishihara, and K. Furuta, "Control of unstable mechanical system Control of pendulum," *International Journal of Control*, vol. 23, no. 5, pp. 673–692, 1976.

[13] K. Furuta, M. Yamakita, and S. Kobayashi, "Swing-up control of inverted pendulum using pseudo-state feedback," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 206, no. 4, pp. 263–269, 1992.

[14] K. J. Åström and K. Furuta, "Swinging up a pendulum by energy control," *Automatica*, vol. 36, no. 2, pp. 287–295, 2000.

[15] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 834–846, 1983.

[16] S. A. Campbell, S. Crawford, and K. Morris, "Friction and the inverted pendulum stabilization problem," *Journal of Dynamic Systems, Measurement, and Control*, vol. 130, no. 5, p. 054502, 2008.

[17] A. E. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. Routledge, 2018.

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

[19] B. Kouvaritakis, J. A. Rossiter, and J. Schuurmans, "Efficient robust predictive control," *IEEE Transactions on Automatic Control*, vol. 45, no. 8, pp. 1545–1549, 2000.

[20] P. O. M. Scokaert and J. B. Rawlings, "Constrained linear quadratic regulation," *IEEE Transactions on Automatic Control*, vol. 43, no. 8, pp. 1163–1169, 1998.

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

[22] J. B. Rawlings and K. R. Muske, "The stability of constrained receding horizon control," *IEEE Transactions on Automatic Control*, vol. 38, no. 10, pp. 1512–1516, 1993.

[23] A. Jadbabaie and J. Hauser, "On the stability of receding horizon control with a general terminal cost," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 674–678, 2005.

[24] D. F. Delchamps, "Stabilizing a linear system with quantized state feedback," *IEEE Transactions on Automatic Control*, vol. 35, no. 8, pp. 916–924, 1990.

[25] N. A. Janmohamed and M. A. Dupuis, "Modeling and control for a flexible inverted pendulum robot," NASA Kennedy Space Center, Kennedy Space Center, FL, Internship final report NTRS 20205009993, 2020. [Online]. Available: https://ntrs.nasa.gov/citations/20205009993

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

[27] 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.

7.11 Exercises

7.1 ○ Concept check · 2 min

In case 4 the angle loop alone holds the pole within 1.53° of upright, yet the cart runs into the right bumper at 4.11 s. Which statement explains the runaway?

(a) Rail friction makes the constant term of the loop's polynomial negative, so one real root is positive at any gain

(b) The angle gain of 30 N/rad is below the floor $(M+m)g$, so the pole is held only loosely and drags the cart

(c) The quantized angle reading adds a small mean force through the derivative term that pushes the cart to one side

(d) The 30 ms drive lag delays every correction, and the delayed corrections add up to a growing cart speed

7.2 ○ Concept check · 1 min

In case 1 the cart target steps from 0 to 0.5 m at 8 s. What does the cart do first under the cascade?

(a) It moves toward the target at once, while the pole leans away from the target to balance the push

(b) It backs about 21 mm away from the target, which leans the pole toward the target before the cart follows

(c) It stays still until the lean request reaches the 6° clamp, and then it moves toward the target

(d) It moves slowly toward the target, because a cart gain of 7° per meter asks for very little lean

7.3 ○ Concept check · 2 min

The MPC of case 10 plans offsets c_k from the regulator's move $-K\tilde{x}$, with the regulator's Riccati solution P as its terminal weight. When no limit binds anywhere in the plan, what force does it apply?

(a) The regulator's move, but only once the horizon is longer than the crossing at 0.46 s

(b) A force smaller than the regulator's, because the weight W on the offsets is larger than R

(c) The regulator's move at every horizon, because the cost is $\tilde{x}_0^T P \tilde{x}_0$ plus $W\sum c_k^2$

(d) A force that already brakes for the 0.8 m line, because the plan looks one second ahead

7.4 ▭ Theory and calculation · 6 min

The default rig has a cart of $M = 1.0\,\mathrm{kg}$ and a uniform rod of $L = 0.5\,\mathrm{m}$ and $m_p = 0.15\,\mathrm{kg}$ with no tip mass, and $g = 9.81\,\mathrm{m/s^2}$. Neglect both frictions.

(a) Find the first moment $m\ell$, the inertia J about the pivot, the length $\ell_{eq} = J/m\ell$ and $\Delta_0 = (M+m)J - (m\ell)^2$. *Checkpoint:* 0.0375 kg m, 0.0125 kg m^2, 0.333 m and 0.012 97 kg^2 m^2.

(b) Find the unstable root λ_0 from Equation (7.7), its e-folding time, and the force per radian $(M+m)g$ that holds a tilted pole still. *Checkpoint:* 5.71 s^{-1}, 175 ms and 11.3 N/rad.

(c) Find the wrong-way zero $z = \sqrt{g/\ell_{eq}}$, how far it lies below λ_0, and its value for the 1.0 m rod of case 5. *Checkpoint:* 5.42 rad/s, 5 % below, and 3.84 rad/s.

7.5 Theory and calculation · 6 min

The controller sees the pole through an encoder of 4096 counts per turn, and it forms the rate as the filtered difference of Equation (7.6) with $\tau_f = 15\,\text{ms}$.

(a) At the default $T_s = 10\,\text{ms}$, find the filter coefficient a and the size of one count in degrees. *Checkpoint:* 0.487 and 0.0879°.

(b) One count changes between two samples. Find the step this causes in the rate estimate $\hat{\omega}$, and in the force through $K_{d\theta} = 5\,\text{N s/rad}$ and through $K_{p\theta} = 30\,\text{N/rad}$. *Checkpoint:* 0.075 rad/s, 0.37 N and 0.046 N.

(c) In case 8 ($T_s = 20\,\text{ms}$, $\tau_d = 60\,\text{ms}$), find the average sensing lag as a fraction of the 179 ms e-folding time, and compare it with the lag at which the window of Figure 7.2 closes. *Checkpoint:* 70 ms, 0.39, against about 105 ms, 0.59.

7.6 Theory and calculation · 7 min

Inside its lean clamp the cascade is four-state feedback with $k_x = K_{p\theta}K_{px}$ and $k_v = K_{p\theta}K_{dx}$, and with the frictionless model it closes into the quartic of Equation (7.10). Use the defaults $K_{p\theta} = 30\,\text{N/rad}$, $K_{d\theta} = 5\,\text{N s/rad}$, $K_{px} = 7^\circ$ per meter and $K_{dx} = 10^\circ$ per m/s, with the values of Δ_0, $m\ell$ and ℓ_{eq} for the default rod.

(a) Find k_x, k_v and the five coefficients. *Checkpoint:* 3.67 N/m, 5.24 N s/m, and 0.346, 3.25, 17.5, 51.4, 36.0.

(b) Find the s^3 coefficient for the 1.0 m rod of case 5 ($\ell_{eq} = 0.667\,\text{m}$) and the s^2 coefficient for the 1.0 kg rod of case 6 ($M + m = 2.0\,\text{kg}$). *Checkpoint:* 1.51 and 9.16.

(c) Find the s^3 coefficient with $K_{p\theta} = 75\,\text{N/rad}$ (case 3), first with $K_{d\theta} = 5$ and then with 8 N s/rad. Explain why more angle gain cannot restore this coefficient. *Checkpoint:* 0.64 and 3.64.

7.7 App challenge · 8 min

Load scenario 4, The angle loop alone, and change no setting. Use the default seed (the app draws no random numbers).

(a) Press Run and let the run go until the app pauses itself, about 0.6 s after the pole falls. Read the Pole and Rail hits tiles. *Checkpoint:* Fell at 4.54 s, with 5 rail hits.

(b) Export the CSV file and read the cart speed, column `v_mps`, at 1 and 4 s. Find the average growth rate of the speed between them and compare it with the estimate $b_c g/(K_{p\theta} - (M+m)g)$. *Checkpoint:* 0.128 and 0.520 m/s, a rate of 0.47 s^{-1} within 0.01 s^{-1}, against 0.52 s^{-1}.

(c) Press Reset, set Controller structure to Cascade before pressing Run, and run to 10 s. Read the Cart travel tile. *Checkpoint:* 0.046 m, within 0.002 m.

7.8 App challenge · 8 min

Load scenario 7, A hard push, which taps the pole at 150 °/s at 2 s. Use the default seed and run each part until the app pauses itself after a fall or until 8 s.

(a) Run the scenario as shipped. Read the Pole and Rail hits tiles. *Checkpoint:* Fell at 3.29 s, with 6 rail hits.

(b) Press Reset, switch to Advanced, raise Lean limit on the cart card from 6 to 15° and run. Read the Cart travel tile. *Checkpoint:* 0.745 m, within 0.005 m.

(c) Press Reset, set Controller structure to LQR and run. Read the Cart travel and Force saturated tiles. *Checkpoint:* 0.495 m within 0.005 m, and 0.07 s.

State which limit decides the push each configuration survives.

7.9 App challenge · 8 min

Load scenario 11, Terminal cost: Off against LQR at a short horizon. The MPC plans 0.3 s ahead with no terminal weight, and the pole takes a 150 °/s push at 2 s. Use the default seed.

(a) Run as loaded until the app pauses itself. Read the Pole tile. *Checkpoint:* Fell at 3.04 s.

(b) Press Reset, switch to Advanced, set Terminal cost to LQR on the MPC card and run to 8 s. Read the Cart travel tile. *Checkpoint:* 0.509 m, within 0.005 m.

(c) Press Reset, set Terminal cost back to Off, set Prediction horizon to 0.5 s and run to 8 s. Then press Reset, set the horizon to 0.6 s and run to 8 s. Read the Pole and Cart travel tiles each time. *Checkpoint:* 0.5 s: fell at 7.54 s. 0.6 s: up, with 0.793 m, within 0.005 m.

Compare the shortest horizon that holds the pole with the linear crossing of Section 7.5.6.

7.10 Application area · 4 min

A team plans to balance the default pole with a camera in place of the angle encoder. The camera delivers 30 frames per second, each image arrives 50 ms late, and the law runs once per frame.

(a) Estimate the average sensing lag $\tau_d + T_s/2$ and its fraction of the 179 ms e-folding time. *Checkpoint:* About 67 ms, or 0.37.

(b) Compare the result with case 8 and say whether the shipped gain of 30 N/rad can be kept. Name two effects of a real camera that the chapter's sensing model leaves out, and say which way each moves the conclusion. *Checkpoint:* Close to case 8's 70 ms, so the gain must come down.

7.11 Application area · 3 min

The bumpers return the cart with restitution $e = 0.4$, and through the pivot the impact changes the pole rate by $\Delta\omega = -(m\ell\cos\theta/J)\,\Delta v$ with $\Delta v = -(1+e)v$, as in Equation (7.3). For the default rod $m\ell/J = 3\,\mathrm{m}^{-1}$.

(a) Find the change in pole rate for a cart that reaches the bumper at 0.55 m/s with the pole near upright. *Checkpoint:* 132 °/s.

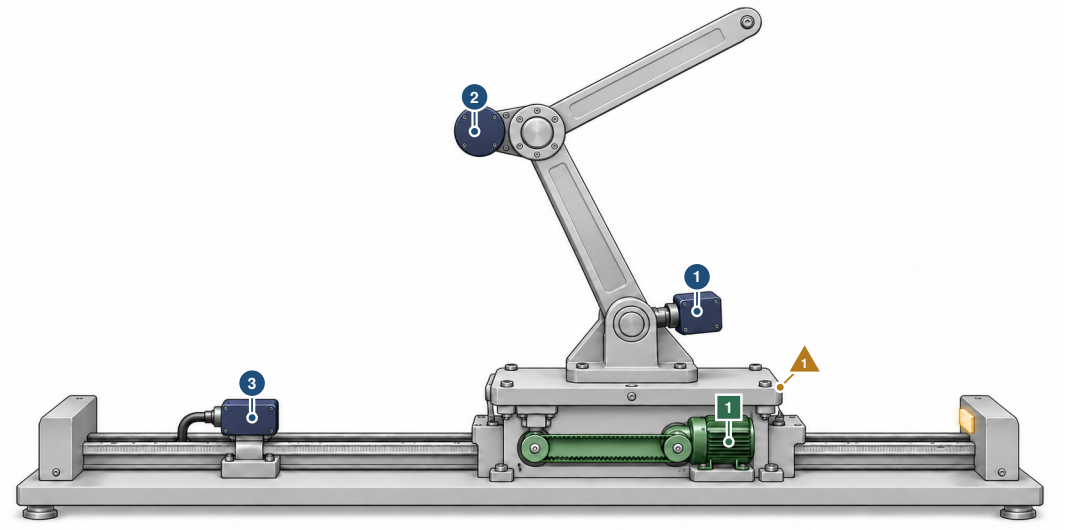
(b) Find the largest arrival speed that keeps the change below the 60 °/s tap of case 1, and the change at 0.55 m/s for a bumper with no rebound ($e = 0$). *Checkpoint:* 0.25 m/s and 95 °/s.

7.12 Review the agent · 5 min

Case 8 replaces the encoder angle with a camera-like reading every 20 ms that arrives 60 ms late, and the shipped cascade loses the pole at 6.84 s. An AI assistant proposes a fix: "The delay does not change the plant, whose unstable root stays at $5.58\,\mathrm{s}^{-1}$. The loop fails because a stale measurement makes it react too weakly. Raise the angle gain from 30 to 40 N/rad so that it pushes harder for the same reading, and keep the angle damping at 5 N s/rad." Find the flaw and say how you would check it.

CHAPTER 8

Double Inverted Pendulum Simulator for Model-Based Balance and Computed Swing-Up



The controller moves 1 motor and belt that push the cart

The controller reads 1 lower link's angle at its pivot 2 angle between the links at the elbow 3 where the cart is on the rail

What disturbs it 1 a push on the cart

Notice that one motor moving one cart must hold both hinged links upright while the cart stays between the bumpers at the ends of the rail. Illustration, not to scale.

IN PLAIN TERMS **Two Rods Stacked on One Moving Cart**

The main idea. A motor pushes a cart along a short rail. Two rods stand on the cart, the second balanced on the tip of the first, each on a free hinge. The controller keeps both rods upright by moving only the cart.

The challenge. The top rod folds back faster than a person can react. Pushing the cart toward the side each rod leans never works. The controller needs a model of the whole machine and a good guess of how fast each rod turns.

Why it matters. Standing robots, cranes with swinging loads and people keeping their balance face the same problem.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history in Section 8.2, the fact on the optimizer in Section 8.5, the try-its and the fact on sampling in Section 8.6, and the key points
Study	65 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 8.6), Technology in Practice (Section 8.7), Limitations (Section 8.8) and the open challenge

The control question Two rods stand on a cart that one motor pushes, and the faster of their two unstable modes is too quick for a hand to hold. What must a stabilizing law know and do that single loops cannot, and what fails when each of those requirements is taken away?

Learning objectives

- Show from the low-order coefficients of the closed-loop polynomial why no pair of angle loops that push toward both leans can hold the pendulum.
- Explain how differenced encoder counts make the force chatter, and why a Kalman filter on the same readings does not.
- Relate the longest stable sample interval to the drive lag the design leaves out, and compare it with the Nyquist interval.
- Describe how time-varying feedback along a stored swing-up absorbs a model error that the stored force alone cannot.
- Compare a constrained predictive controller with the regulator it is built from, and say where knowing the limits helps and where a linear plan hurts.

Prerequisites Sections 1.4.4 and 1.4.5 (state space, LQR), 1.4.6 (observer, Kalman filter), 1.4.7 (RHP poles and zeros, lag limits) and 1.6.3 (predictive control). Chapter 7 for the single pendulum on the same cart.

The app https://apmonitor.com/apps/double-inverted-pendulum/

See also Volume 9, Chapter 4 for an offset-free predictive controller whose Kalman estimator carries a disturbance, and Volume 9, Chapter 2 for predictive control of a slow plant behind a dead time. Volume 5, Chapter 13 for an extended Kalman filter that estimates where an underactuated vehicle is when nothing measures its position.

8.1 Introduction

A double inverted pendulum on a cart was under computer control by 1978 [1], and in 1993, when its stabilization was already a familiar demonstration of state-space design, it was also swung up from hanging [2]. What makes the machine a hard test is that one motor must hold two unstable modes. The simulator of this chapter keeps the cart, rail, drive and encoders of the single-pendulum application of Chapter 7 and adds a 0.3 m, 0.10 kg rod on a free pivot at the tip of the 0.5 m pole. Linearized at upright, the rig has unstable roots at $+11.07\,\mathrm{s}^{-1}$ and $+4.49\,\mathrm{s}^{-1}$. The fast mode folds the upper link back against the lower one, $\theta_2 = -3.86\,\theta_1$, and multiplies by e every 90 ms. Released 0.5° from upright with the drive idle, the upper link passes 2° after 0.205 s, and the angle watchdog trips the drive at 0.615 s.

Three habits learned on single loops fail on this plant, and the simulator makes each failure measurable. The first is to give each link a loop of its own. Proportional-derivative loops that push the cart toward each link's lean, beside a loop that pulls the cart to its target, are never stable here, and two coefficients of the closed-loop polynomial say why: the two angle gains may sum to no more than $(M+m_1+m_2)g = 12.26\,\mathrm{N/rad}$, while the lower link's gain alone must exceed 58.6 N/rad. The second habit is to size the sample interval by the Nyquist frequency of the fastest motion. A regulator redesigned for 20 ms samples loses a 3.9 Hz oscillation, about a sixth of that sampler's 25 Hz Nyquist frequency, because its model leaves out the drive's 30 ms lag, and with that lag no interval longer than 12.3 ms is stable. The third is to trust a trajectory computed offline. Played without feedback, the stored swing-up reaches the balance regulator only on the exact rig it was computed for, while feedback along it absorbs an error in the upper link's mass from −27 % to +27 %.

These difficulties belong to a class of machines in which several modes share one actuator, so that no mode can be corrected without disturbing the others. People standing on a support shorter than their feet change from an ankle strategy, which moves the body about the ankle joints, to a hip strategy, whose motion is focused about the hips [3]. A model of that balance built from linearized mechanics, full-state feedback and state estimation reasonably matched experimental data once its objective took advantage of the fast modes that the body's inertia and geometry dictate [4]. A crane

whose payload swings as a double pendulum has two oscillation frequencies to suppress through the commands to its motors [5], where the single-pendulum crane of Chapter 12 has one. Maneuvers in which a humanoid robot gets up from the ground or recovers from a large disturbance acrobatically have been synthesized by online trajectory optimization [6]. Each of these needs commands that coordinate several modes at once, works from a state that is partly unmeasured, or plans its large motions instead of regulating them.

However, the experiments cited here document designs that succeeded on each rig, such as a digital servo that held one on an inclined rail [7], swing-ups by energy and by a limit cycle [8], and trajectories followed with feedforward and feedback [9, 10], and a failed alternative, such as a loop pair that cannot hold or a stored force played without feedback, is a fall that a physical rig is built to prevent. This chapter identifies, on a simulator described equation by equation, which requirements of full-state design this plant enforces and what goes wrong when each one is removed. The regulator, the Kalman filter, iterative LQR, time-varying tracking and constrained model predictive control are all standard methods from the literature.

Section 8.2 traces the history of the benchmark. Section 8.3 gives the nonlinear model, whose impacts pass through the mass matrix, and the elbow encoder that reads the angle between the links. Section 8.4 derives a closed form for the low-order coefficients of the closed loop under any single-input law, which excludes every loop pair that pushes toward both leans. Section 8.5 describes the controllers as coded: three estimators, among them a Kalman filter that predicts with the nonlinear model, the loop pair, the regulator, a stored swing-up that a repository script regenerates byte for byte, and a predictive controller built on the regulator's own model and told only what the regulator is not, the force the drive can deliver and where the rail ends (Section 8.5.6). Section 8.6 runs the ten cases and the challenge, each of which takes one requirement away or, in the tenth, adds one. It maps the longest stable sample interval against the drive lag that the design leaves out, checked against 11 nonlinear runs, and tests a linear rule that predicts the tilts the force limit can catch. Its comparison of the regulator with the predictive controller (Section 8.6.10) separates what planning within the force limit buys on a noisy estimate from what a linear plan costs near that limit. Section 8.7 describes the encoders, drives and real-time controllers of laboratory rigs, and Sections 8.8 and 8.9 state the limitations and the conclusions, and the exercises follow in Section 8.10.

8.2 A Benchmark Held, Then Swung Up

Computer control of a double inverted pendulum was reported in 1978 [1]. Two years later a digital controller designed from linear servo theory kept a double inverted pendulum on a cart at a given reference point while the cart stood on an inclined rail, removing the effect of the incline on the cart position, and its authors gave theoretical analyses of the pendulum and of the loop [7]. By 1993 the stabilization had become a standard demonstration of the state-space approach, and swinging the pendulum from its stable hanging state to the upright one, a transfer between equilibria, was posed as the next problem and shown to work in experiments [2].

Energy came next. A swing-up of the double pendulum built on an energy principle was followed by a more robust transfer through a limit cycle, and both were demonstrated on hardware [8]. Passivity and energy shaping were later used to drive a double pendulum on a cart from any initial position into a region of attraction around upright, where a balance controller designed on the linearized model takes over, a scheme shown in simulation [11]. The application keeps that division into a swing-up and a balance law joined by a handover, but it computes its swing-up instead of shaping energy.

Trajectories offered the other route. A reference trajectory obtained by optimizing a spline-interpolated initial guess was tracked with a gain-scheduled linear quadratic controller designed for that reference, and it swung a self-erecting double pendulum up in simulation and on a rig [9]. Feedforward and feedback together swung up a double pendulum on a cart with experimental validation [10], and a swing-up with experimental validation was later reported for a triple pendulum on a cart [12].

The optimizer used here improves a whole control sequence at once. A second-order method that successively improves such a sequence for a nonlinear discrete-time system, converging in one step when the system is linear and the cost quadratic, was derived in 1966 with an argument motivated by dynamic programming [13]. Iterative LQR, which repeats a linear quadratic approximation about the current trajectory, was introduced for locally optimal feedback control and computed energy-optimal reaching movements of a musculoskeletal arm with 10 states and 6 controls, converging substantially faster than three earlier methods [14]. Online trajectory optimization later synthesized getting-up and recovery maneuvers for humanoid robots at seven times slower than real time and solved the acrobot problem in real time [6]. A stored trajectory is only as useful as the feedback around it. Trajectories stabilized by LQR, with stability regions computed by sums-of-squares programming, have been assembled into a tree whose region of attraction probabilistically covers the controllable part of the state space [15]. The handover test of Section 8.5 is a far simpler and unverified relative of such a region.

> ⧖ HISTORY **The Funnel Picture of Feedback, 1985 to 2010**
>
> A robotics paper of 1985 popularized the funnel as a picture of a feedback policy: a law that collapses a large set of starting states into a smaller set of final ones. In 1999 feedback motion planning was drawn as a sequence of such funnels, each a locally valid policy, that carries a broad set of starting states to a goal. The weak point then was the mouth of each funnel, hard to compute and often estimated by trial and error. A Lyapunov function joins the two pictures, because a function that falls along every motion inside a ball proves that the ball drains to the goal. The cost-to-go of a regulator supplies such a candidate automatically, which is why a method of 2010 built its funnels around trajectories stabilized by LQR [15]. The handover test of this chapter is a level set of the same kind of cost-to-go.

8.3 The Cart and Its Two Links

The rig is the single-pendulum application's cart on a 2 m rail, with the same bumpers, drive and encoders, carrying a second uniform rod on a free pivot at the tip of the first, as shown in Figure 8.1. It is the upright cart of Section 1.4.1 with a second link on top and a force in place of a commanded acceleration. The cart position x counts from the middle of the rail, and both link angles are absolute, taken from upright and positive toward $+x$, so θ_2 is not the angle between the links. The eight states are x, $\dot{x}$, θ_1, $\dot{\theta}_1$, θ_2, $\dot{\theta}_2$, the force F that the drive has developed, and a latch for a lower link lying on its deck stop. Table 8.1 sorts the signals by role.

8.3.1 The Equations for Three Coordinates

The kinetic energy of the cart and both rods is $\frac{1}{2}\dot{q}^{\mathsf{T}}M(q)\dot{q}$ with $q=(x,\theta_1,\theta_2)$, and Lagrange's equations with viscous friction take the form of Equation (8.1).

$$M(q)\,\ddot{q}=\begin{bmatrix} F-b_c\dot{x}+h_2 s_1\dot{\theta}_1^2+h_3 s_2\dot{\theta}_2^2 \\ h_7 s_1-h_5 s_{12}\dot{\theta}_2^2-b_p\dot{\theta}_1+b_{p2}\dot{\theta}_{21} \\ h_8 s_2+h_5 s_{12}\dot{\theta}_1^2-b_{p2}\dot{\theta}_{21} \end{bmatrix} \tag{8.1}$$

where $s_i=\sin\theta_i$, $c_i=\cos\theta_i$, $s_{12}=\sin(\theta_1-\theta_2)$, $c_{12}=\cos(\theta_1-\theta_2)$ and $\dot{\theta}_{21}=\dot{\theta}_2-\dot{\theta}_1$, the elbow damping acting on the relative rate. The mass matrix and its constants are given in Equation (8.2).

$$\begin{gathered} M(q)=\begin{bmatrix} h_1 & h_2 c_1 & h_3 c_2 \\ h_2 c_1 & h_4 & h_5 c_{12} \\ h_3 c_2 & h_5 c_{12} & h_6 \end{bmatrix} \\ h_1=M+m_1+m_2,\quad h_2=l_1\left(\tfrac{m_1}{2}+m_2\right),\quad h_3=\tfrac{1}{2}m_2 l_2 \\ h_4=l_1^2\left(\tfrac{m_1}{3}+m_2\right),\quad h_5=\tfrac{1}{2}m_2 l_1 l_2,\quad h_6=\tfrac{1}{3}m_2 l_2^2 \\ h_7=g\,h_2,\qquad h_8=g\,h_3 \end{gathered} \tag{8.2}$$

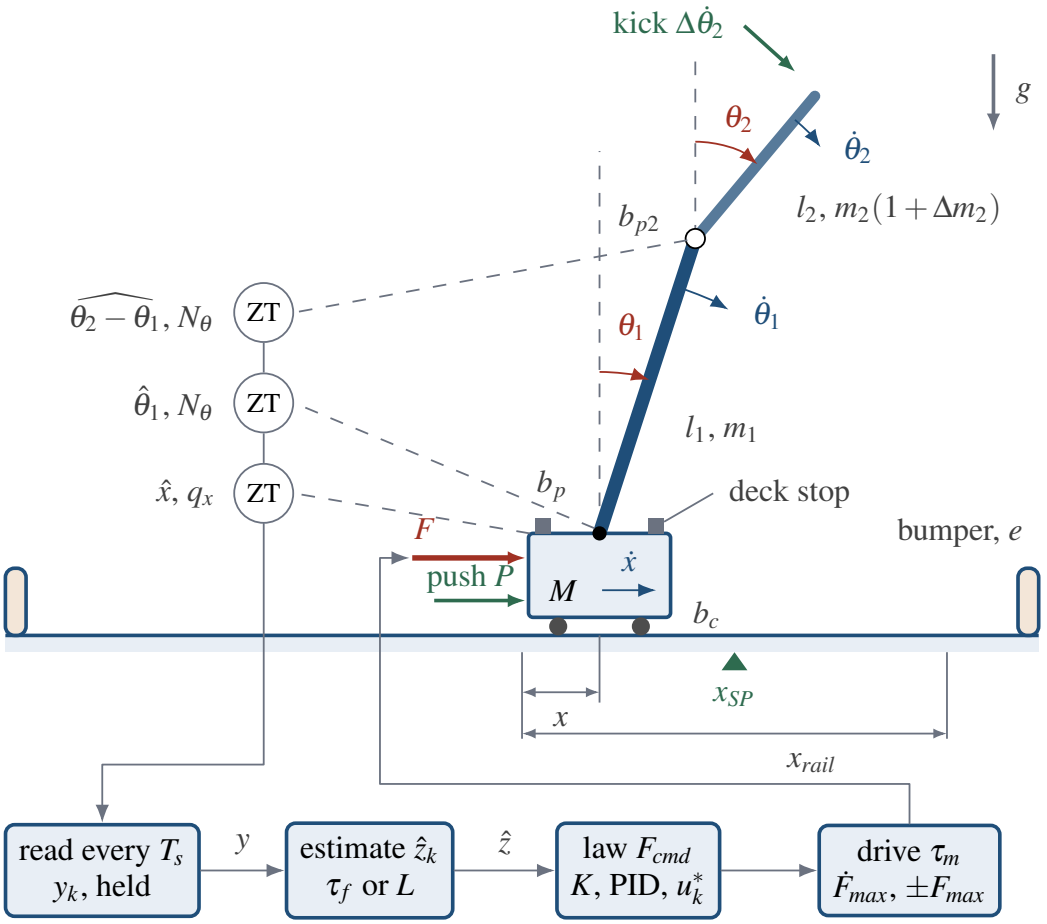


Figure 8.1 The cart and two links on a rail with bumpers, and the chain from the three encoders to the drive, not to scale. The cart of mass M on viscous friction b_c is pushed by the developed force F. The lower rod (l_1, m_1) turns on a pivot with damping b_p and the upper rod (l_2, m_2) on a free elbow with damping b_{p2}, and both angles are measured from upright, positive toward $+x$. Deck stops catch a lower link that passes 90°, and the cart center meets the bumpers at $\pm x_{rail}$ with restitution e. The encoders ZT read the cart to q_x, and the lower link and the angle between the links to N_θ counts per turn. Every T_s the readings y become an estimate $\hat z$, a law computes F_{cmd}, and the drive follows it with lag τ_m, slew limit $\dot F_{max}$ and limit $\pm F_{max}$. A push P strikes the cart, a kick $\Delta\dot\theta_2$ the upper link, and Δm_2 changes the upper link's mass.

Its top row balances the forces on the cart in newtons, and the lower two rows balance moments in newton meters about the cart pivot and the elbow. For the default rods $h_1 = 1.25\,\mathrm{kg}$, $h_2 = 0.0875\,\mathrm{kg\,m}$, $h_3 = 0.015\,\mathrm{kg\,m}$, $h_4 = 0.0375\,\mathrm{kg\,m^2}$, $h_5 = 0.0075\,\mathrm{kg\,m^2}$ and $h_6 = 0.003\,\mathrm{kg\,m^2}$. Because $M(q)$ is the matrix of a kinetic energy that is positive for every nonzero velocity, it is positive definite in every configuration, and the application solves the 3×3 system from its cofactors at each evaluation of the right-hand side. Its verification suite checks the definiteness over a 5° grid of both angles. Without force or friction the total energy $\frac{1}{2}\dot q^{\mathsf T} M \dot q + h_7 c_1 + h_8 c_2$ holds to one part in 10^6 of the swing-up energy $2(h_7 + h_8) = 2.011\,\mathrm{J}$ over 10 s, from three starts.

Table 8.1 Signals of the Rig by Role

Symbol	Role	Nominal	Range
θ_2, θ_1	CVs, link angles	0°	band ±2°
x	CV, cart position	0 m	SP ±0.6, band ±0.02
F_{cmd}	MV, drive command	0 N	±20
x, $\dot{x}$	states, cart	0	$\lvert x\rvert \le 1.0$ m
θ_i, $\dot{\theta}_i$	states, links	0	full circle
F	state, developed force	0 N	±20
—	state, deck latch	0	0 or ±1
$\hat{x}$, $\hat{\theta}_1$, $\hat{\phi}$	measurements	—	q_x, $2\pi/N_\theta$ steps
$\hat{z}$	estimate in charge	—	3 estimators
P	DV, push on the cart	0.5 N s	0 to 3
$\Delta\dot{\theta}_2$	DV, kick on link 2	30 °/s	0 to 300
Δm_2	DV, mass of link 2	0 %	−30 to +60
Δb_c	DV, rail friction	0 %	−100 to +300
$\theta_i(0)$	DV, release tilts	0°	±20°
—	DV, lost sample	—	event
λ_f, λ_s	outputs, roots	11.07, 4.49 s^{-1}	—
E	output, link energy	0 J	−2.011 J hanging

8.3.2 Drive, Rail, Deck Stop and Disturbances

The drive is a force source, and Equation (8.3) gives the lag and the rate bound through which its developed force follows the command.

$$\dot{F} = \mathrm{sat}\Big(\frac{F_{cmd} - F}{\tau_m},\ \dot{F}_{max}\Big), \qquad |F_{cmd}| \le F_{max} \tag{8.3}$$

where $\mathrm{sat}(v,c) = \max(-c, \min(c,v))$, $\tau_m = 30$ ms, $\dot{F}_{max} = 1000$ N/s and $F_{max} = 20$ N. Every impact is an impulse applied between integration steps through the inverse mass matrix, as shown in Equation (8.4).

$$\Delta\dot{q} = \frac{\Delta\dot{q}_j}{e_j^{\mathsf{T}} M(q)^{-1} e_j} M(q)^{-1} e_j \tag{8.4}$$

where e_j selects the coordinate struck and $\Delta\dot{q}_j$ is the change imposed on it. At a bumper, j is the cart and $\Delta\dot{x} = -(1+e)\dot{x}$ when the cart moves into the stop at more than $v_{rest} = 0.02$ m/s, with $e = 0.4$, and below that speed the restitution is zero. The same impulse reaches both links through the off-diagonal terms of $M(q)^{-1}$, which the verification suite checks against the closed form, and with $e = 1$ it conserves kinetic energy exactly. A lower link that passes 90° lands on a deck stop, $j = \theta_1$ and $\Delta\dot{\theta}_1 = -\dot{\theta}_1$, and from then on it moves with the cart while the upper link swings on from its tip, so the dynamics reduce to the two coordinates x and θ_2. The swing-up and the balance with a hanging lower link have no deck stop.

Three disturbances act on the state directly and one on the measurements. A push is an impulse P on the cart, $\Delta\dot{q} = P M(q)^{-1} e_1$. A kick adds $\Delta\dot{\theta}_2$ to the upper link's rate and leaves the cart and the lower link as they were. A release sets both links to chosen tilts at rest, and a dropout withholds one encoder sample. The mass change Δm_2 and the friction change Δb_c alter the rig that the physics runs, and only a present-links design hears of Δm_2. The energy of the links relative to upright rest, $E = \frac{1}{2}(h_4\dot{\theta}_1^2 + h_6\dot{\theta}_2^2) + h_5 c_{12}\dot{\theta}_1\dot{\theta}_2 + h_7(c_1 - 1) + h_8(c_2 - 1)$, is −2.011 J hanging at rest and zero balanced.

8.3.3 What the Controller Reads

The controller never reads the state unless the estimate is set to perfect, which the application labels a teaching comparison. When the law executes, the three encoders are read at that instant, as shown in

Table 8.2 Plant Constants. Class a Is Adjustable in the Model Dialog, s Is a Disturbance Slider, h Is Hidden and Fixed.

Symbol	Meaning	Value	Range	
M	cart mass	1.0 kg	0.4 to 3	a
b_c	cart friction	1.0 N s/m	0 to 8	a
l_1, m_1	lower rod	0.5 m, 0.15 kg	0.3 to 0.8, 0.05 to 0.5	s
l_2, m_2	upper rod	0.3 m, 0.10 kg	0.15 to 0.8, 0.02 to 0.4	s
b_p	pivot damping	0.002 N m s/rad	0 to 0.05	a
b_{p2}	elbow damping	0.0005 N m s/rad	0 to 0.01	a
g	gravity	9.81 m/s^2	1.6 to 24.8	a
F_{max}	drive force limit	20 N	5 to 40	a
τ_m	drive lag	30 ms	5 to 100	a
$\dot{F}_{max}$	drive slew limit	1000 N/s	100 to 5000	a
x_{rail}	bumper position	1.0 m	fixed	h
e	restitution	0.4	0 to 1	a
v_{rest}	resting-contact speed	0.02 m/s	fixed	h
q_x	cart encoder step	0.02 mm	0.001 to 2	a
N_θ	joint encoder counts	4096/turn	256 to 65536	a
λ_f, λ_s	unstable roots	11.07, 4.49 s^{-1}	derived	
z_1, z_2	zeros of x/F	10.95, 4.20 s^{-1}	derived	

Equation (8.5).

$$\hat{x} = q_x \,\mathrm{rnd}\Big(\frac{x}{q_x}\Big), \qquad \hat{\theta}_1 = q_\theta \,\mathrm{rnd}\Big(\frac{\theta_1}{q_\theta}\Big)$$
$$\hat{\phi} = q_\theta \,\mathrm{rnd}\Big(\frac{\theta_2 - \theta_1}{q_\theta}\Big), \qquad \hat{\theta}_2 = \hat{\theta}_1 + \hat{\phi} \tag{8.5}$$

where rnd rounds to the nearest integer, $q_x = 0.02\,\mathrm{mm}$ and $q_\theta = 2\pi/N_\theta$ with $N_\theta = 4096$, one count being 0.0879°. The elbow encoder reads the angle $\hat{\phi}$ between the links, so a count of either joint moves the controller's $\hat{\theta}_2$. The readings carry no noise and no transport delay, and nothing in the application is random, so a reset, which restarts from $t = 0$ with the settings in force, replays a run exactly, and the seed of Section 1.6.1 changes nothing in this app.

8.3.4 Constants and Integration

Table 8.2 lists the plant constants with their classes. The engine of Section 1.6.1 advances the eight states by the fourth-order Runge-Kutta method on a 1 ms step. The controller clock ticks on that same step and runs the law once per T_s, the log keeps 200 samples per simulated second, and a run can be played at speeds from 0.1 to 2 seconds of simulated time per second of real time. The bumper and deck-stop impulses follow each integration step, before the outputs are computed and logged. Halving the step moves the peak upper-link angle and the peak cart travel of the case 2 catch by less than 0.5 %, and the application passes all 52 of its verification checks.

8.3.5 Assumptions

Rigid, uniform rods. The mass matrix of Equation (8.2) holds for two rigid rods of uniform section. A long and light upper link would also bend, and the model has no way to show it. The elbow damping is a quarter of the cart pivot's because the upper rod's inertia about the elbow is only $h_6 = 0.003\,\mathrm{kg\,m^2}$, and the pivot's value placed at the elbow would by itself move the fast root from 11.37 to 10.42 s^{-1}.

Viscous friction. The cart and both pivots resist motion in proportion to speed. Stick-slip on a real rail makes a balanced cart hunt about its target in a small limit cycle, which nothing here can produce.

A lumped drive. One lag and one slew limit stand for the amplifier, the motor and the belt together. Backlash and a voltage limit would lengthen that lag, which Section 8.6 charges against the sampling margin.

Ideal impacts. A bumper contact takes zero time and has a single restitution coefficient, a lower link landing on the deck loses all of its rate, and the cart never lifts from the rail.

Encoders that only round. Each encoder errs by rounding to its count and in no other way, without noise, lost counts or delay. A camera, or noise on the counts, would spend part of the same lag margin.

Instantaneous disturbances on a level rail. Releases, kicks and pushes act in zero time. The rail is horizontal, which leaves the cart no steady force to resist, and gravity, the masses and the frictions change only when a disturbance changes them.

One stored trajectory. The swing-up was computed for the default rig, the default drive and 5 ms samples, and on any other rig it is a trajectory for a different machine.

8.4 What the Linearization Allows

Sections 1.4.4 and 1.4.7 set out the tools used here on a single pendulum. Three models are kept apart. The nonlinear Equations (8.1) to (8.5) are the simulation model. This section's linearizations are used for analysis and nowhere else, while the controllers of Section 8.5 carry design models of their own, built from the default links unless the reader chooses the present ones.

8.4.1 Two Modes and Two Wrong-Way Zeros

About upright rest, where $q = 0$, small motions obey $M_0\ddot{q} = Gq - D\dot{q} + e_1F$ with $M_0 = M(0)$, $G = \mathrm{diag}(0, h_7, h_8)$ and D the friction matrix. Its unstable roots are +11.37 s^{-1} and +4.57 s^{-1} without friction and +11.07 s^{-1} and +4.49 s^{-1} with it, one e-fold every 90 and 223 ms. The fast mode folds the links, $\theta_2 = -3.86\,\theta_1$, while the slow one leans them together, $\theta_2 = 1.50\,\theta_1$, and as $m_2 \to 0$ without friction one root becomes the single pendulum's 5.71 s^{-1}. The fast mode grows by a factor of 9.15 in 200 ms, 1.057 in one 5 ms sample and 1.248 in one of 20 ms.

Writing the linearized dynamics as $P(s)\,q(s) = e_1F(s)$ with $P(s) = M_0s^2 + Ds - G$, the transfers from force to the three coordinates share the denominator $\det P(s)$, and their numerators are the cofactors of the first row of $P(s)$, as shown in Equation (8.6).

$$\frac{x}{F} = \frac{C_{11}(s)}{\det P(s)}, \qquad \frac{\theta_1}{F} = \frac{C_{12}(s)}{\det P(s)}, \qquad \frac{\theta_2}{F} = \frac{C_{13}(s)}{\det P(s)} \tag{8.6}$$

The cofactor C_{11} is the characteristic polynomial of the two links with the cart held still, so the links' own unstable roots, +10.95 s^{-1} and +4.20 s^{-1}, are right-half-plane zeros of the force-to-cart transfer. Friction moves their stable partners to –4.24 s^{-1} and –11.52 s^{-1}, off the mirror image. One such zero makes a cart move start the wrong way, as Section 1.4.7 shows for the single pendulum. Two make a quick move of the cart start the right way, reverse, and reverse again before heading out, as cases 2 and 9 measure.

8.4.2 Why No Loop Pair Holds It

With the force applied without lag or sampling, any law $F = -Kz$ on the six mechanical states closes the loop on the characteristic polynomial of Equation (8.7), because the force enters only the first row of $P(s)$, so the closed-loop determinant expands along that row.

$$\Delta(s) = \det P(s) + (K_x + K_{\dot{x}}s)\,C_{11}(s) + (K_{\theta_1} + K_{\dot{\theta}_1}s)\,C_{12}(s) + (K_{\theta_2} + K_{\dot{\theta}_2}s)\,C_{13}(s) \tag{8.7}$$

A pair of angle loops that push the cart toward each link's lean, beside a cart loop that pulls it toward its target, is the law $F = K_{p1}\theta_1 + K_{d1}\dot{\theta}_1 + K_{p2}\theta_2 + K_{d2}\dot{\theta}_2 - K_{px}x - K_{dx}\dot{x}$, with every gain positive in the intuitive tuning. Without friction, C_{12} and C_{13} share their s^2 term gh_2h_3, because a slow acceleration

of the cart tilts both links alike, the pendulum leaning as one body in the accelerating frame. The two angle loops enter the lowest coefficients only through their sums, as shown in Equation (8.8).

$$\begin{aligned} a_2 &= g h_2 h_3 \left[(M + m_1 + m_2) g - K_{p1} - K_{p2} \right] - g \gamma K_{px} \\ a_3 &= -g h_2 h_3 \left(K_{d1} + K_{d2} \right) - g \gamma K_{dx} \\ a_4 &= \alpha K_{p1} - \beta K_{p2} + \delta K_{px} - a_4^0 \end{aligned} \tag{8.8}$$

where a_k multiplies s^k in $\Delta(s)$, $\gamma = h_3 h_4 + h_2 h_6$, $\alpha = h_2 h_6 - h_3 h_5$, $\beta = h_2 h_5 - h_3 h_4$, $\delta = h_4 h_6 - h_5^2$, all positive for uniform rods, and $a_4^0 = g[h_1 \gamma - h_2 h_3 (h_2 + h_3)]$ is the deficit of the open loop. A stable polynomial needs every coefficient positive. At the default rig $a_2 > 0$ caps the summed angle gain, $K_{p1} + K_{p2} < 12.26\,\mathrm{N/rad} - (0.629\,\mathrm{m}) K_{px}$, and $a_4 > 0$ demands $K_{p1} - 0.625 K_{p2} > 58.6\,\mathrm{N/rad} - (0.375\,\mathrm{m}) K_{px}$. Both hold only if $K_{p2} < -28.5\,\mathrm{N/rad} - (0.156\,\mathrm{m}) K_{px}$, a loop that pushes against the upper link's lean. The conflict is not peculiar to the default rig. With $K_{px} = 0$ the threshold exceeds the cap by $a_4^0/\alpha - (M + m_1 + m_2) g = g h_3 (h_1 h_4 - h_2^2 + h_1 h_5 - h_2 h_3)/\alpha$, and both differences in the bracket are positive for any uniform rods, while a cart gain only widens the gap because $\delta/\alpha < \gamma/(h_2 h_3)$. Over 20 000 random rigs the threshold was never less than 1.08 times the cap. The third coefficient asks independently for $K_{d1} + K_{d2} < -(0.629\,\mathrm{m}) K_{dx}$, so on balance the derivative action must also push against the links' rates. With the default friction the cap becomes 12.20 N/rad and the threshold 58.4 N/rad, and the largest a_3 that nonnegative gains reach while a_2 stays positive is -0.0081, still below zero.

The contrast with a single pole is the sign pattern of the open loop. With one unstable root λ, $\det P(s)$ is proportional to $s^2 (s^2 - \lambda^2)$, its s^2 coefficient is negative, and a loop that pushes toward the lean repairs it, which is the familiar floor of $(M + m_1) g$ on the angle gain that Chapter 7 derives for the single pole. With two, the polynomial is $s^2 (s^2 - \lambda_f^2)(s^2 - \lambda_s^2)$, the s^2 coefficient is already positive and the deficit sits at s^4, so pushing toward both leans only spends the s^2 coefficient. The shipped pair, 40 N/rad and 4 N s/rad on each link, sums to 80 N/rad. With its cart loop removed, on the frictionless model, it holds the lean-together motion as a single-pole loop would and leaves one real root at $+15.25\,\mathrm{s}^{-1}$ whose shape, $\theta_2 = -1.58\,\theta_1$, folds the links faster than the open loop's fast mode, and the default friction adds a second real root at $+0.14\,\mathrm{s}^{-1}$. The regulator of Section 8.5 has $K_{p1} = 153.6\,\mathrm{N/rad}$, $K_{p2} = -315.7\,\mathrm{N/rad}$ and $K_{d1} + K_{d2} = -52.7\,\mathrm{N\,s/rad}$ in these terms, and on the frictionless model its coefficients from a_0 to a_6 are 3.79, 4.00, 2.00, 0.422, 0.0455, 0.00233 and 5.86×10^{-5}, all positive. A search of 257 600 pairs that push toward both leans on the lag-free linear model finds none stable, the best leaving a root at $+9.12\,\mathrm{s}^{-1}$, and the application's own verification grid of 3125 pairs agrees.

🕮 **Pause.** Try Exercise 8.1 before reading on.

8.5 The Digital Controller

8.5.1 Sampling, Command and Watchdog

The law executes when a whole sample interval T_s has passed on the simulation clock, 5 ms by default and adjustable from 1 to 40 ms, and its force is held until the next execution. At each execution the encoders are read, all three estimators are updated, the law computes a force toward $+x$, and the command is clipped at $\pm F_{max}$, saturation being flagged whenever the unclipped command exceeds that limit. A watchdog trips the drive once either angle reading lies within one count of 90° from its target, and in the swing-up only after the handover. The settings are collected in Table 8.3.

8.5.2 Three Estimates of the State

The law needs $z = (x, \dot{x}, \theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2, F)$ and has three position readings, so the three estimators process the same samples in parallel and the trends show them together. The perfect estimate is the true state, sampled. The differenced estimate takes the positions as read, forms each rate as a filtered

Table 8.3 Controller Settings

Symbol	Meaning	Default	Range
T_s	sample interval	5 ms	1 to 40
τ_f	rate filter, differenced	5 ms	0 to 30
σ_F	Kalman force noise	2 N	0.1 to 20
σ_τ	Kalman joint torque noise	2 mN m	0 to 50
r	measurement variance, $\times q^2/12$	1	0.1 to 100
x_s	LQR cart excursion	0.3 m	0.05 to 1
θ_{1s}, θ_{2s}	LQR angle excursions	10°	1 to 45
F_s	LQR force scale	10 N	1 to 40
$\dot{x}_{ref}$	cart reference rate	0.25 m/s	0.05 to 2
K_{p1}, K_{p2}	link loop gains	40	−200 to 400
K_{d1}, K_{d2}	link loop damping	4	−20 to 40
K_{px}, K_{dx}	cart loop	5, 5	0 to 50
θ_h	handover ellipsoid	5°	1 to 20
t_0	earliest swing-up start	0.5 s	0.2 to 5
F_{push}	manual push	2 N	0.5 to 20
T_d	MPC decision period	10 ms	fixed
T_p	MPC horizon	1 s	fixed
δ	MPC rail margin	0.05 m	fixed

Loop gains in N/rad, N s/rad, N/m and N s/m.

backward difference, Equation (8.9), and replays the drive's lag on the commands sent to estimate F.

$$\hat{v}_k = \hat{v}_{k-1} + \eta\left(\frac{\hat{p}_k - \hat{p}_{k-1}}{T_s} - \hat{v}_{k-1}\right), \qquad \eta = 1 - \mathrm{e}^{-T_s/\tau_f} \tag{8.9}$$

where $\hat{p}$ is a reading and $\hat{v}$ its rate, and $\eta = 0.632$ at the defaults. One count differenced over 5 ms is a rate step of 17.6 °/s. Backward differences of encoder data respond well to velocity transients, while least-squares estimators filter imperfect measurements better [16], and the trade between chatter and lag appears in case 3.

The Kalman filter, whose filtering problem is the dual of the regulator problem [17], has the form of the observer of Section 1.4.6. It predicts the readings from the model and corrects the prediction by how far each reading misses. Its prediction runs the design model's full nonlinear equations over one sample, in Runge-Kutta steps of about 1 ms, with the command that was held, and its correction uses a steady-state gain, as shown in Equation (8.10).

$$\bar{z}_k = f(\hat{z}_{k-1}, u_{k-1}), \qquad \hat{z}_k = \bar{z}_k + L\,(y_k - H\bar{z}_k) \tag{8.10}$$

where $y = (\hat{x}, \hat{\theta}_1, \hat{\phi})$ and H maps z to $(x, \theta_1, \theta_2 - \theta_1)$. The gain L comes from the dual Riccati iteration of the shared module `shared/src/framework/core/riccati.js` on the seven-state model that includes the drive lag, sampled with a hold at T_s. Its process noise is a white force of $\sigma_F = 2$ N at the cart and white torques of $\sigma_\tau = 2$ mN m at each joint, entering through the columns of M_0^{-1}, and its measurement noise is each encoder's quantization variance $q^2/12$. The iteration converges in 232 sweeps, and the slowest error mode of the filter decays at $-7.6 \pm 6.3i\ \mathrm{s}^{-1}$. A lost sample leaves the readings as they were, so the differenced rates see a zero step and then a double one, while the Kalman filter only predicts.

8.5.3 The Loop Pair

The loop pair sums three copies of the shared velocity-form PID of Section 1.6.2, one on each estimated angle with the action reversed so that the loop pushes toward its link's lean, and one on the cart that pulls it toward the cart reference. Each derivative term differences the measurement after

smoothing it with a lag of τ_f, and each emitted output is clipped at $\pm F_{max}$. The integral is disabled in all three loops, so nothing is back-calculated. In manual mode the angle loops track their own proportional and derivative terms while the cart loop absorbs the rest of the operator's force in its integral state, so switching back to automatic causes no bump, and with no integral gain that state then stays as a fixed bias.

8.5.4 The Regulator

The regulator applies $F = -K(\hat{z} - z^*)$ with $z^* = (x_{ref}, 0, \theta_1^*, 0, \theta_2^*, 0, 0)$ holding the cart reference and the target angles. Its design model is the linearization at the target for the default links, without the drive lag (six states, F taken as F_{cmd}) unless the lag is modeled as a seventh state, discretized with a zero-order hold at T_s through one matrix exponential evaluated by scaling and squaring. Each weight is one over the square of the excursion the designer will accept in that quantity, the rule of Section 1.4.5, as shown in Equation (8.11).

$$Q = \operatorname{diag}\Big(\frac{1}{x_s^2}, 0, \frac{1}{\theta_{1s}^2}, 0, \frac{1}{\theta_{2s}^2}, 0\Big), \qquad R = \frac{1}{F_s^2} \tag{8.11}$$

The fixed-point Riccati iteration of the same shared module runs on the cost-to-go matrix S from $S = Q$ until the relative change falls below 10^{-11}, 1384 sweeps at the defaults, and gives $K = [30.0, 31.7, -153.6, 18.0, 315.7, 34.7]$ on $(x, \dot{x}, \theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)$, with a zero appended on F. The force goes toward the lower link's lean and against the upper link's, at about twice the gain. The gain is recomputed whenever the sample interval, a weight, a plant constant or the design model changes. On the rig with its drive lag, the closed loop sampled at 5 ms has poles at $-1.56 \pm 1.23i$, $-1.94 \pm 27.48i$, $-5.62 \pm 4.73i$ and $-13.31\ \mathrm{s}^{-1}$, a largest magnitude in the z-plane of 0.992, which the Modes view draws with the eigenvalue routine of `shared/src/framework/core/eig.js`. The cart reference walks to the target at $\dot{x}_{ref}$, and the law carries no velocity reference, so a ramp is chased rather than tracked.

The advanced controls offer two other balance targets, and each has a single unstable root. Chosen in case 3 on the Kalman filter, either target starts the upper link 2° from it, and a 30 °/s kick on the upper link was fired at each whole second from 3 to 8 s, once that start had been caught. With the upper link hanging below an upright lower one the root is $+5.68\ \mathrm{s}^{-1}$, and the regulator designed there holds every kick within 2.65° on the lower link and 38.8 mm of cart travel. With the lower link hanging and the upper one standing on its tip the root is $+8.84\ \mathrm{s}^{-1}$, with $\theta_2 = 6.55\,\theta_1$, and the angle gains grow to 980.0 and 1129.1 N/rad in magnitude. The 2° start alone swings the links 21.7° and 16.4°. Five of the six kicks swing them up to 67.0° and 54.1° from the target and the cart up to 0.635 m, with both links back within 0.75° 10 to 12 s later, and the kick at 7 s tips them past 90°.

8.5.5 The Stored Swing-Up

The swing-up has four phases. While the links swing far from hanging, an energy sink asks for the cart acceleration $k_E(h_2 c_1 \dot{\theta}_1 + h_3 c_2 \dot{\theta}_2)$ with $k_E = 12$, which in the cart's frame drains the links' energy at the rate $k_E(h_2 c_1 \dot{\theta}_1 + h_3 c_2 \dot{\theta}_2)^2$, plus a centering spring at 1.2 rad/s with damping ratio 0.9. Inside 15° and 1 rad/s of hanging, and until the links leave 30° or 3 rad/s, an LQR about the hanging equilibrium settles them and parks the cart, both laws capped at 8 N. At the first sample after t_0 at which both links are within 2° and 0.1 rad/s of hanging at rest, and the cart within 1 cm and 0.02 m/s of its start, the law begins to follow the stored trajectory with time-varying feedback, as shown in Equation (8.12).

$$u_k = u_k^* - K_k(\hat{z}_k - z_k^*), \qquad (\hat{z} - z_{up})^{\mathsf{T}} S (\hat{z} - z_{up}) \le S_{55}\,\theta_h^2 \tag{8.12}$$

where z_k^* and u_k^* are the stored states and forces, K_k the stored gains, z_{up} is upright at rest, and the second relation is the handover test, a level set of the balance regulator's cost-to-go matrix S that reaches θ_h on the upper link's angle alone. From the first sample inside it the balance regulator takes

over, with its cart reference starting where the handover found the cart. A trajectory that ends outside the set trips the drive rather than fling the cart after a lost pendulum.

The trajectory comes from iterative LQR run offline by the repository script `double-inverted-pendulum/tools/swingup.mjs` on the application's own simulation model. One step of its dynamics, $z_{k+1} = f(z_k, u_k)$, is one 5 ms sample of Equations (8.1) to (8.3), integrated by the same Runge-Kutta steps of 1 ms, so replaying the stored force on the default rig reproduces the stored states to better than 10^{-6}. Over $N = 500$ samples, 2.5 s, from hanging at rest to upright at rest at the rail center, it minimizes Equation (8.13).

$$J = \sum_{k=0}^{N-1} \Big[\tfrac{1}{2} w_u u_k^2 + \tfrac{1}{2} q_w (\dot\theta_{1,k}^2 + \dot\theta_{2,k}^2) + \tfrac{1}{2} w_x \big(|x_k| - x_{max}\big)_+^2 + \tfrac{1}{2} w_s \big(|u_k - F_k| - \tau_m \dot F_{max}\big)_+^2 \Big] + \tfrac{1}{2} \|z_N - z_{up}\|_{Q_f}^2 \quad (8.13)$$

subject to $|u_k| \le 17\,\mathrm{N}$, which leaves the feedback a margin of the drive's 20 N. Here $(\cdot)_+$ is the positive part, $w_u = 10^{-3}$ per N^2, $q_w = 2 \times 10^{-3}$ per $(\mathrm{rad/s})^2$, $w_x = 2 \times 10^4$ per m^2 beyond $x_{max} = 0.6\,\mathrm{m}$, $w_s = 10$ per N^2 beyond the 30 N that the drive can slew within its lag, and $Q_f = \mathrm{diag}(2000, 200, 50000, 2000, 50000, 2000, 1)$. Each iteration takes finite-difference Jacobians of f along the nominal, runs a backward pass with a regularization that grows when a step fails and shrinks when it succeeds, clamps a step that would leave the force bound and gives it no feedback, and accepts the forward pass by a backtracking line search, whose form the script's comments credit to [6]. From a fixed guess, a 6 N sine at 0.8 Hz ramped in over 0.5 s, it stops after 446 iterations at a cost of 18.497, when an accepted step improves the cost by less than one part in 10^9, and the check mode regenerates the stored file byte for byte in about 6 s. The nominal force never exceeds 15.21 N, the cart reaches 0.600 m from center at 1.20 s, and both links first swing further from upright, to 216.7° and 230.5°, before they come up and are within 90° of upright from 1.51 s.

> **+ DID YOU KNOW? The First Race of the Swing-Up Optimizer**
>
> When iterative LQR was introduced in 2004, one of its three test problems was swinging up a simple pendulum. It converged in 0.26 s of computing and 5 iterations, against 1.61 s and 20 iterations for differential dynamic programming in the same environment [14]. The two links on a cart of this chapter took 446 iterations.

The gains K_k come from a final backward Riccati pass along the nominal with the balance weights, starting from the cost-to-go of the seven-state balance design, so they end on that design's gain, $[30.0, 31.3, -248.0, 14.0, 404.1, 43.4, 1.18]$. At the start they read $[30.1, 24.7, 113.2, 5.4, -21.1, 2.5, 0.7]$, and both angle gains change sign 1.43 s into the trajectory, as the nominal links pass 115.9 and 98.9° from upright, near horizontal. Because the shipped balance design leaves the drive lag out, the gain steps at the handover.

8.5.6 Model Predictive Control on the Regulator's Model

The fourth balance structure, the last value of the controller choice and the one that case 10 selects, is a constrained model predictive controller (MPC). It works in the receding-horizon manner of Section 1.6.3, whose stability and optimality under constraints are reviewed in [18]. The regulator asks for whatever force its gain gives and the drive clips the excess, and none of the laws above knows where the rail ends. The MPC is told both limits and nothing else that the regulator lacks. It acts where the regulator acts, at a balance target from a start near it, and the swing-up and its handover stay as described.

The prediction model is the regulator's design model. With $\tilde z_k = \hat z_k - z^*$ the estimate measured from the target and u_k the force command, that model is $\tilde z_{k+1} = \Phi \tilde z_k + \Gamma u_k$, the six-state linearization, or seven with the drive lag modeled, held over T_s. The MPC rebuilds Φ, Γ, K and S with the regulator's own design routines from the same design constants and weights Q and R, so its matrices equal

the regulator's bit for bit, and a change of sample interval, weight, plant constant or design model redesigns both laws on the same sample, on which the MPC also plans afresh. With the cart mass raised to 1.3 kg in the middle of case 10, the two gains change on the same law sample and agree bit for bit on all 599 law samples of the run. Both controllers are therefore given the present plant, as stated for the regulator above, while a disturbance that the design does not hear, such as Δb_c, is model error to both. Nothing is identified. The MPC reads the estimate selected for the regulator, the target, the force held since the last sample, the limit F_{max} as the drive clamp reads it and the fixed bumper position x_{rail}. The plant of Equations (8.1) to (8.3), the encoders and estimators of Equations (8.5) to (8.10), and this linear model remain three separate things.

The decision variables are departures from the regulator's force, a pre-stabilized parametrization in the manner of [19], and each plan predicts the regulator's own closed loop, as shown in Equation (8.14).

$$u_k = -K\tilde{z}_k + v_k, \qquad \tilde{z}_{k+1} = (\Phi - \Gamma K)\tilde{z}_k + \Gamma v_k \tag{8.14}$$

where v_k is the departure at step k, counted in samples of T_s from the present estimate $\tilde{z}_0$. The prediction thus starts from a stable loop instead of the open-loop roots at $+11.07\,\mathrm{s}^{-1}$ and $+4.49\,\mathrm{s}^{-1}$. Because K and S satisfy the Riccati equation of the regulator, completing the square in each stage cost and summing makes the terms in S telescope, and the regulator's infinite-horizon cost of any plan whose departures stop after N steps reduces to Equation (8.15).

$$\sum_{k=0}^{\infty} \left(\tilde{z}_k^{\mathsf{T}} Q \tilde{z}_k + R u_k^2\right) = \tilde{z}_0^{\mathsf{T}} S \tilde{z}_0 + W \sum_{k=0}^{N-1} v_k^2 \tag{8.15}$$

where $W = R + \Gamma^{\mathsf{T}} S \Gamma$, $0.012\,31\,\mathrm{N}^{-2}$ at the design of case 10 against $R = 0.01\,\mathrm{N}^{-2}$. Every step after the horizon is the regulator's own, so the cost-to-go S acts as the terminal weight and the sum of squared departures is the only part a plan can change. With no limit in reach the best departures vanish whatever the horizon or the blocking, and the MPC is the regulator. Within its limits the plan is a move-blocked, finite-horizon form of the constrained linear quadratic regulator [20].

The horizon holds $N = 200$ samples, 1 s at 5 ms and shorter at faster sampling, and the departures are held over 13 blocks, four of 2 samples, two of 4, and then 6, 8, 12, 18, 30, 50 and 60 samples. Each plan solves the quadratic program of Equation (8.16).

$$\begin{aligned} &\min_{v,\ \varepsilon \geq 0} \ W \sum_{k=0}^{N-1} v_k^2 + \rho\, \varepsilon^2 \\ &|u_k| \leq F_{max},\ k \in \mathscr{K}_F, \qquad |x_k| \leq x_{rail} - \delta + \varepsilon,\ k \in \mathscr{K}_x \end{aligned} \tag{8.16}$$

where u_k and the predicted cart position $x_k = x_{ref} + \tilde{z}_{1,k}$ follow Equation (8.14), $\mathscr{K}_F$ holds every sample of the first 0.1 s and every fourth sample after it, 65 of the 200, $\mathscr{K}_x$ every fourth sample, $\delta = 0.05\,\mathrm{m}$ is the margin inside each bumper and ε is one slack for the whole horizon. The force rows are hard. The rail band is soft, priced at $\rho = 10 N Q_{11} = 10 \times 200 \times 11.1\,\mathrm{m}^{-2} = 22\,222\,\mathrm{m}^{-2}$, ten times the regulator's own cart weight summed over the horizon. That rule replaced, once and openly, a first choice of a 0.1 m margin priced at 10^5 per m^2 and 10^3 per meter, which lost a kick that the regulator holds at a 40 N limit. The program has 14 variables and 230 inequality rows, 130 for the force and 100 for the rail. The dual active-set solver shared across the series handles it, with a ridge of 4.4×10^{-5} added to its diagonal. Between the judged samples a planned force is free, and in 98 of 3473 accepted plans from eight closed-loop runs one passes F_{max} by up to 17.5 N. None is applied, because the samples a plan applies before the next decision are always judged.

A plan is made every $T_d = 10\,\mathrm{ms}$, once in $\mathrm{round}(T_d/T_s)$ executions of the law and at least once per execution, which gives 9 ms at 3 ms samples and 7 ms at 7 ms, and the priming call at $t = 0$ makes the first plan. Between decisions the regulator keeps correcting from the fresh estimate with the planned departure for that sample added, $u_j = -K\tilde{z}_j + v_j$, so the departure is held and the force is

not. Holding the first planned force alone for 10 ms drops the pendulum at 1.79 s after a 30 °/s kick and at 1.82 s after a 0.4 N s push, with the cart target at +0.6 m, the Kalman filter and the drive lag left out of the design, while the regulator and the held departure both catch these disturbances, and with the lag modeled all three hold. The drive receives $F_{cmd} = \mathrm{sat}(u_j, F_{max})$, the same clamp as under the regulator, so the correction between decisions may still ask past the limit, as it did on 154 of 6989 balancing samples in the eight runs, never on a decision sample with the plan fully in force.

The MPC uses the estimate in charge and adds no disturbance estimator. Offset-free predictive control usually augments the model with an integrating disturbance, as the boiler of Volume 9, Chapter 4 does with a fuel bias in its Kalman estimator, but no law here integrates, the rail is level, and upright rest at the target with zero force is an equilibrium of the plant and of the model for any masses, while an estimator would break Equation (8.15) and the identity with the regulator. A plan is rejected if the solver does not report an optimum, if a departure, the slack or the cost is not finite, if the first force or any judged force lies outside $\pm F_{max}$, or if the plan has the wrong length. The regulator then takes the drive with the departure ramped out over 50 ms, counted in whole samples, the record is marked, and after five accepted plans in a row the departure is ramped back in. A state from which no plan meets every force row is rejected in the same way, and the verification suite forces each path of the rejection.

An independent condensation of the explicit objective, with the stage costs and S as the terminal weight, reproduces the application's plans on 3473 accepted decisions of the eight runs to 2.0×10^{-8} N with the same ridge. In the 3158 of them with no limit in reach every departure is exactly zero. A decision takes 0.255 ms on average and 0.334 ms at the 99th percentile in Node.js, most of it spent building the prediction. On the application's frame loop in the Chrome browser the whole controller uses 3.2 to 5.4 % of one core at real time and 6.0 to 7.4 % at twice real time over two sets of runs on a shared, loaded host, against a budget of a tenth of a core and 0.22 to 0.40 % for the regulator.

8.5.7 Manual Operation and the Models Kept Apart

Manual operation hands the operator a drive-force slider of ± 20 N in 0.5 N steps and pushes of F_{push} that last 0.05 s from a button, or while an arrow key is held down. While a law is in charge the slider tracks its command, so taking manual control starts from the force already applied. The estimators run on the commands actually sent in either mode, which leaves the regulator ready the moment the law resumes. Equations (8.1) to (8.5), run with the true links, form the simulation model. The design models are the regulator's linearization, the Kalman filter's nonlinear prediction with the drive lag, and the swing-up's simulation model of the default rig. The MPC predicts with the regulator's linearization and has no model of its own. None hears about Δb_c, and only a present-links design hears about Δm_2. The analysis models are the linearizations of Section 8.4 and the sampled closed loop of the true rig under the design's gain that the Modes view plots.

A balance target chosen before Run, or carried through a reset, starts the links at the case's tilt from that target, and the swing-up starts hanging at rest. Either way the case and event markers in the trends are kept.

8.6 The Cases, One Requirement at a Time

Table 8.4 lists the ten cases and the challenge, and Table 8.5 collects the outcomes. Every value was measured by running the application's engine headlessly on the named scenario, with the one change stated made before the run, which gives the same run as that change carried through a reset, and reading the trajectories and the scorecard the application computes. Each case starts from the reference settings of Table 8.3, with the regulator on the Kalman filter at 5 ms unless its row says otherwise.

8.6.1 No Hand and No Loop Pair

In case 1 the pendulum stands exactly at upright for 1 s, and then the upper link is let go 0.5° from upright with the drive at zero. It passes 2° after 0.205 s and 10° after 0.359 s, the watchdog trips the

Table 8.4 Scenario Sequence

#	Change from the reference	Requirement exposed
1	manual, loop pair in reserve, θ_2 released 0.5° at 1 s	a model, not a hand
2	perfect states, $\theta_2(0) = 2°$, target 0.5 m at 3 s, 2 m/s	signs from the model
3	differenced rates, $\theta_2(0) = 2°$	estimated rates
4	$\theta_2(0) = 1°$, T_s 5 to 20 ms at 3 s	margin for unmodeled lag
5	perfect states, released at 6° and 2.5°	a force budget
6	$l_2 = 0.6$ m, design for 0.3 m	a current model
7	swing-up from hanging	a computed trajectory
8	swing-up, $\Delta m_2 = 10\,\%$, no feedback	feedback along it
9	target 0 to 0.5 m at 1 s	reference shaping
10	MPC, lag modeled, target 0.6 m, 60 °/s kick at 1 s	the worth of known limits
C	swing-up, differenced, target 0.5 m	all at once

Table 8.5 Outcomes Read From the Running Engine

Case and condition	Result
1, no force after the release	θ_2 past 2° at 0.205 s, tripped at 0.615 s
1, loop pair / reversed pair / LQR	falls 0.530 s after / within 0.50° / within 0.638°
2, catch of the 2° start	cart left 27.0 mm, θ_1 to 4.17°, 11.82 N, in 0.1° from 2.46 s
2, target step at 3 s	right 5.9 mm, left to −51.3 mm, settled in 2.00 s
3, differenced / Kalman, 10 s	chatter 7.17 / 0.43 N, rate error 5.76 / 0.53 °/s
3, τ_f 0 / 10 / 15 ms	chatter 12.89 N / falls at 1.038 s / at 0.855 s
4, switch to 20 ms	falls at 6.47 s, pair $+1.64 \pm 24.28i$ s^{-1}
4, set after the switch, 12 / 13 / 15 ms	holds / falls at 14.96 s / at 7.95 s
4, lag modeled, 20 / 30 / 40 ms	holds at all three
5, released at (6°, 2.5°) / (6, −2.5°)	caught on 10.1 N, 117 mm / falls 0.576 s after
5, band at 10 / 20 / 40 N	8.4 / 26.1 / 52.4 % of the grid
6, design for 0.3 / 0.6 m	falls at 2.83 s / within 0.33°
6, l_2 0.4 / 0.5 m, old design	holds 10 s / falls at 9.01 s
7, handover at θ_h 2 / 5 / 10°	2.165 / 2.030 / 1.900 s into the trajectory
7, tracking error, Kalman / differenced	0.142° / 2.18°
8, feedforward alone / with feedback	misses, 225.6 times the level / hands over
8, rescued mass changes	−27 to +27 %, present-links design +20 to +60 %
9, $\dot{x}_{ref}$ 0.25 / 1 / 2 m/s	dip 7.1 / 27.6 / 50.8 mm, settled 3.23 / 2.14 / 2.01 s
10, LQR / MPC	both 0.114 m from the bumper, at the limit 25 / 15 ms
C, shipped / suggested	5.93 / 4.59 s, chatter 6.55 / 0.59 N

drive 0.615 s after the release, and the lower link lands on its deck stop at 2.097 s. At the 0.1 speed setting the same 0.615 s lasts 6.15 s of wall time, which gives a hand on the slider time to see the fold but no force to stop it. Handed to the loop pair in automatic operation, the release falls 0.530 s later on a peak command of 6.7 N, as Section 8.4.2 predicts. A pair with its upper loop reversed, 80 N/rad and 3 N s/rad on the lower link, −80 N/rad and −10 N s/rad on the upper and 1 N/m and 1 N s/m on the cart, satisfies all three conditions and holds the release within 0.50°, and the regulator on the Kalman filter holds it within 0.638° with 14.8 mm of cart travel.

> **TRY IT IN THE APP** **1 · Balance It by Hand**
>
> Load the scenario, leave it in Manual and press Run. Watch the upper link after it is let go at 1 s: it folds against the lower link, passes 2° after 0.205 s, and the watchdog trips the drive 0.615 s after the release. At the 0.1 speed setting the same fall lasts 6.15 s.

8.6.2 Signs From the Model

Case 2 runs the regulator on perfect states from a 2° tilt of the upper link. The cart's first move is to the left, 27.0 mm in 0.136 s, which swings the lower link 4.17° toward the upper link's lean and brings the elbow back under it. The peak force is 11.82 N, and both links are inside 0.1° from 2.46 s. At 3 s the target moves to 0.5 m with a reference walking at 2 m/s. The cart goes right 5.9 mm in 0.161 s, back to −51.3 mm at 0.460 s, and only then out to the target, inside 2 cm and 2° 2.00 s after the step, with the lower link leaning up to 6.51°. Two reversals where a single pole makes one are the two right-half-plane zeros of Equation (8.6) at work.

8.6.3 Rates From Three Encoders

Case 3 repeats the catch on the encoder readings. With differenced rates through the 5 ms filter, one count of the elbow encoder becomes a rate step of 0.307 rad/s, which the gain of 34.7 N s/rad on $\dot{\theta}_2$ turns into 10.6 N, 6.7 N after the first filter step. Over a 10 s run the command then changes by 7.17 N RMS from one sample to the next, the upper link wobbles ±0.64° at 4.7 Hz between 3 and 10 s with the force swinging ±9.10 N, and the estimate of $\dot{\theta}_2$ is wrong by 5.76 °/s RMS. On the same readings the Kalman filter gives 0.43 N of chatter, ±0.11° and 0.53 °/s. Filtering the differences harder is no escape. With a 10 ms filter the pendulum falls at 1.038 s and with 15 ms at 0.855 s, while with no filter the chatter reaches 12.89 N. A longer filter delays the rates the law uses, the same kind of lag that limits the sample interval in case 4.

Retuning either estimator moves it along a trade between chatter and lag, and the two trades do not overlap. Differenced through filters of 0, 2 and 5 ms, the chatter is 12.89, 11.23 and 7.17 N and the rate error 9.88, 8.66 and 5.76 °/s, and an 8 ms filter drops the pendulum at 1.190 s. The Kalman filter with joint torque noise of 0.5, 2, 5 and 10 mN m gives 0.35, 0.43, 0.64 and 1.09 N of chatter and 0.32, 0.53, 0.94 and 1.44 °/s of rate error, so the worst of those settings still improves on the best differenced one by a factor of 6.6 in chatter and 4.0 in rate error. A single lost sample at 4 s hardly shows in either. Between 3.9 and 4.2 s the largest change of the command from one sample to the next is 0.73 N on the Kalman filter, against 0.93 N in the same window without the loss, and 15.80 N on differenced rates, whose own chatter there reaches 15.92 N without the loss.

8.6.4 The Sampling Margin

Case 4 balances on the Kalman filter after a 1° start and at 3 s lengthens the sample interval from 5 to 20 ms, redesigning the regulator and the filter for the new interval at once. The pendulum falls at 6.47 s, 3.47 s after the switch. On the rig, the lightly damped pair at $-1.94 \pm 27.48i\,\mathrm{s}^{-1}$ crosses to $+1.64 \pm 24.28i\,\mathrm{s}^{-1}$, an oscillation of 3.9 Hz against the 25 Hz Nyquist frequency of the sampler, so the sampler still resolves the motion that escapes. With the interval set again just after the switch, 10 and 12 ms hold through 15 s, while 13 ms falls at 14.96 s and 15 ms at 7.95 s, and with the drive lag modeled the same loop holds after a switch to 20, 30 or 40 ms.

> **DID YOU KNOW?** **Fast Enough to See the Motion, Too Slow to Hold It**
>
> From the simulator: switched from 5 to 20 ms samples, the regulator loses the pendulum at 6.47 s on a growing pair at $+1.64 \pm 24.28i\,\mathrm{s}^{-1}$, a motion the sampler still resolves. With the drive lag in the design, the same loop holds at 20, 30 and 40 ms (Table 8.5).

The design is not blind to the hold, since it is redone for each interval on the zero-order-hold model, and the redesign helps, because the 5 ms gains kept at 20 ms reach a largest magnitude of

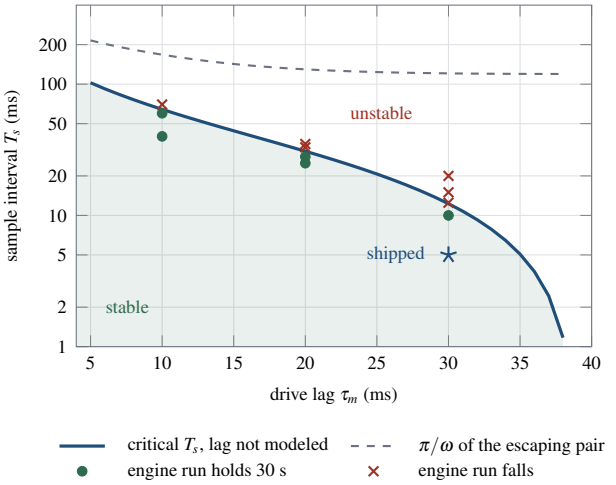


Figure 8.2 Longest stable sample interval of the regulator designed without the drive lag, against the drive lag, with the regulator redesigned at every interval. Below the solid curve the sampled loop on the rig is stable. The dashed curve is the Nyquist interval π/ω of the pair that escapes at the critical interval. The markers are case 4 runs on the nonlinear engine with the lag set before the run, the marked interval taking effect at the scripted switch at 3 s, and each run continued to 30 s, a filled circle for a run that holds and a cross for one that falls. The star marks the shipped setting. With the lag modeled, every interval up to 200 ms is stable.

1.053 and new ones 1.033. What it leaves out is the drive's lag, and the lag it can absorb shrinks as the interval grows. Section 1.4.7 bounds the lag that a loop around an unstable pole can carry on the single pendulum, and the map here adds the sample interval to that bound. Bisecting the largest pole magnitude of the sampled loop on the rig gives the critical interval as a function of the drive lag, as shown in Figure 8.2. At the default 30 ms lag it is 12.28 ms, the pair that escapes there oscillates at 26.0 rad/s, and its Nyquist interval would be 120.8 ms, ten times longer. The tolerable lag falls from 38.1 ms at 1 ms samples to 35.1 ms at 5 ms. Over drive lags from 15 to 35 ms the lag plus half the critical interval stays between 35.2 and 37.5 ms, so the hold spends the margin roughly as half a sample of extra lag would, although the design accounts for it, and the relation fails at short lags, where the sum reaches 56.3 ms at 5 ms of lag. The nonlinear engine agrees at all 11 points tested, run to 30 s because the growth near the boundary is slow. With a 10 ms lag the loop holds at 40 and 60 ms and falls at 70 ms, with 20 ms it holds at 25 and 28 ms and falls at 33 and 35 ms, and with 30 ms it holds at 10 ms and falls at 12.5, 15 and 20 ms. The two falls closest to the boundary, at 33 ms with a 20 ms lag and 12.5 ms with a 30 ms lag, come only at 10.35 and 10.72 s. With the lag in the design model, every interval from 1 to 200 ms is stable at every drive lag from 5 to 40 ms, on a grid of 1 ms in both.

🕮 **Pause.** Try Exercise 8.2 before reading on.

8.6.5 The Catchable Band

Case 5 releases the links at rest at 6° and 2.5° with perfect states. The regulator catches them on a 10.1 N peak with 117 mm of cart travel, and both links settle 1.62 s later, a catch that Figure 8.3 shows on the trends, while a release at 6° and −2.5°, fired at a sample instant once that catch has settled, falls 0.576 s after letting go, and 0.575 to 0.596 s after it over the five 1 ms phases of the sample. The Modes view bakes the full nonlinear catch on perfect states over a 41×41 grid of releases spanning $\pm 16°$ on the lower link and $\pm 8°$ on the upper, and a cell counts as caught when, over 2.5 s, neither link passes 90°, the cart stays off the bumpers and both links end within 2° of upright. At 20 N the

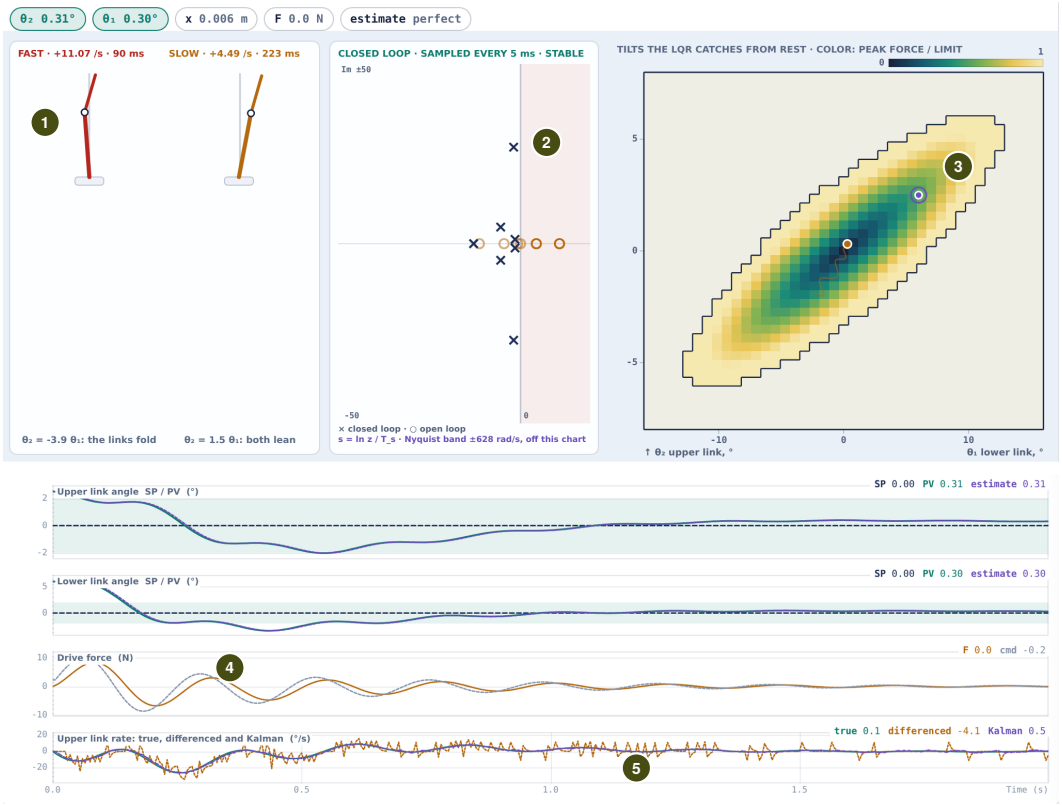


Figure 8.3 The application in case 5, The tilts it can catch, captured at $t = 2.00\,\text{s}$ on the Modes view, after a release at rest at 6° on the lower link and 2.5° on the upper with the regulator on perfect states. The figure is a composite of the badge row and the Modes view above the four trends, and the header, the control panel and the Learn panel with its scorecard are cut. (1) The left panel names the two unstable modes, the fast one at $+11.07\,\text{s}^{-1}$ that folds the links ($\theta_2 = -3.9\,\theta_1$) and the slow one at $+4.49\,\text{s}^{-1}$ that leans them together ($\theta_2 = 1.5\,\theta_1$), and animates each shape, here caught with the fold of the fast mode visible. (2) The middle panel maps the loop sampled every 5 ms into the s-plane, crosses for the closed loop and circles for the open-loop roots, with the lightly damped pair that the text places near ± 27.5 rad/s and the 628 rad/s Nyquist band far off the chart. (3) The right panel is the band of tilts caught from rest at the 20 N limit, colored by the peak force over the limit, with the release as a violet ring and the path of the live state back to the center. (4) Over the 2 s of trends the drive force rings down from its peak of 10.1 N (Table 8.5), while the upper link swings through −2.0° and the lower through −3.4° before settling. (5) The differenced estimate of the upper link's rate, plotted with the true and Kalman rates although perfect states are in charge, jumps each time a count arrives.

caught cells form a thin band through upright along $\theta_2 \approx 0.41\,\theta_1$ that covers 26.1 % of the grid and reaches ±6.4° on the lower link alone but only ±3.2° on the upper. At 40 N it covers 52.4 % and catches the second release on a 39.2 N peak, and at 10 N it shrinks to 8.4 %. That catch is marginal. Fired 1 ms after a sample, so that the drive holds the force computed before the release for 4 ms, the same release at 40 N falls 1.00 to 1.13 s later at six settled instants, while fired 2 to 4 ms after a sample it is caught with the command at the limit. About one release in five is lost, and the case's briefing tells the learner to reset and fire again.

The shape of the band follows from the linear loop. Released by one degree along a direction

in the (θ_1, θ_2) plane, the sampled loop with the drive lag asks for a peak force of 3.28 N along θ_1 alone, 7.34 N along θ_2 alone, and as little as 1.50 N along $\theta_2 = 0.43\,\theta_1$, so at 20 N its unsaturated reach is 6.09°, 2.72° and 13.4°. The farthest cell the nonlinear runs catch, at (−12.8°, −5.6°), lies 14.0° from upright along that direction. The band does not follow the direction that leaves the fast mode unexcited, which is $\theta_2 = 1.51\,\theta_1$. Applied to the grid, the rule that the linear peak force stay within the limit marks 331 cells at 20 N, 19.7 %, and the nonlinear runs catch every one of them plus 108 more that they recover through saturation. At 10 N the rule marks 83 cells, all caught. At 40 N it marks 961 and the nonlinear runs lose 80 of them, 72 of which are still lost with the slew limit removed, so the nonlinear terms of the largest tilts decide those cells. A linear rule sized by the force limit is a safe estimate of what a saturating drive can catch at the default limit, and an optimistic one at twice that limit.

8.6.6 A Model That Is Out of Date

Case 6 doubles the upper link to 0.6 m at the same 0.10 kg while the regulator and the filter keep the design for 0.3 m. The roots fall from 11.07 and 4.49 to 9.06 and 3.92 s^{-1}, an easier plant by its growth rates, and yet the old gains leave a closed-loop pair at $+4.46 \pm 20.68i\ \mathrm{s}^{-1}$, a 3.3 Hz oscillation that grows from the 0.5° start until the pendulum falls at 2.83 s. Redesigned for the present links, with 347.0 N/rad instead of 315.7 N/rad on the upper angle, it holds within 0.33° after the first second. At 0.4 m the old design is already unstable by a factor of 1.009 per sample but grows slowly enough to hold through a 10 s run, and at 0.5 m it falls at 9.01 s.

8.6.7 Swinging Up From Hanging

Case 7 starts hanging. The links are still from the start, so the trajectory begins at 0.50 s, and the handover comes at 2.53 s, 2.03 s into it, with the links at 13.9° and 18.0° but moving home at −85 °/s and −82 °/s, a state the cost-to-go counts as close. Both links are inside 2° and the cart within 2 cm of center from 4.15 s, the cart never passes 0.600 m, and the command peaks at 15.22 N. Played without its feedback on this rig, the stored force alone hands over at the same 2.03 s into the trajectory. With the Kalman filter in charge the links stay within 0.142° of the stored angles at every sample, against 2.18° on differenced rates. A 2° ellipsoid moves the handover to 2.165 s into the trajectory, at 4.1° and 8.7°, and a 10° one to 1.900 s, at 26.0° and 30.4°. On differenced rates the 2° ellipsoid is never entered and the swing-up misses.

> **TRY IT IN THE APP** **7 · Swing-Up Along the Stored Trajectory**
>
> Load the scenario and press Run on the Swing-up view. The links hang still until 0.50 s, then follow the stored trajectory, and the regulator takes over at 2.53 s with the links still 13.9° and 18.0° from upright but moving home. Press Reset, set Handover ellipsoid to 2° and run again: the handover moves to 2.165 s into the trajectory.

8.6.8 Feedback Along the Trajectory

Case 8 makes the upper link 10 % heavier, a payload the trajectory and the designs never hear of, and plays the stored force alone. The upper link goes over the top, and when the trajectory ends at 3.00 s it is at 286.5° with a cost-to-go 225.6 times the handover level, so the drive is tripped. Played alone the stored force hands over only on the exact rig, since of the mass changes from −5 % to +5 % in steps of one percent only zero succeeds. With the time-varying feedback the same +10 % run hands over 2.03 s into the trajectory on a 15.58 N peak, and the feedback rescues every change from −27 % to +27 %. Lighter than that the swing-up misses the ellipsoid. Heavier, it hands over but the balance design for the default links cannot hold the heavier link, and with the design for the present links every change tried from +20 % to +60 % is held. Balancing alone, in case 6 with the upper link set back to 0.3 m and its 0.5° start, the default design holds an upper link 25 % heavier for 10 s but drops one 30 % heavier at 7.46 s, although its sampled loop on that rig still has a largest magnitude of 0.999.

Table 8.6 Regulator, MPC, Governor and MPC Without the Rail Band, Runs of 10 s

	LQR				MPC					Governor	No rail band
Case and change	Fate	Hits	x_m (m)	t_{lim} (ms)	Fate	Hits	x_m (m)	t_{lim} (ms)	Plans	Fate	Fate
10 as shipped	held	0	0.114	25	held	0	0.114	15	1000/0	held	held
10, perfect states	held	0	0.168	90	held	0	0.164	20	1000/0	held	held
10, differenced rates	1.95	1	0	790	held	1	0	135	1000/1	1.95	held
10, 65 °/s	held	0	0.088	135	held	0	0.084	50	1000/0	held	held
10, 65 °/s, perfect states	held	0	0.114	255	2.27	12	0	600	228/1	held	2.27
10, 70 °/s	1.97	2	0	725	held	1	0	185	1000/1	1.97	held
10, 40 N, 80 °/s	held	0	0.028	0	held	0	0.039	0	1000/0	held	held
10, 40 N, 85 °/s	held	1	0	35	1.81	2	0	290	182/1	held	held
10, 40 N, 90 °/s	held	2	0	100	1.83	5	0	420	183/3	held	held
9, target 0.6 m, differenced, 0.4 N s push	held	0	0.363	510	1.94	1	0	405	194/1	held	1.94
3, down-up target	held	0	0.839	1029	1.82	0	0.658	1140	182/4	held	1.74

Fate is held, or the time in seconds at which the watchdog trips. Hits counts bumper contacts, x_m is the least distance of the true cart from a bumper, t_{lim} the time with the command at the drive limit, and Plans the MPC's accepted plans and rejections, the count stopping at a trip. The kicks act on the upper link at 1 s and the push on the cart at 1 s. The 40 N rows keep the drive lag modeled. Governor is the regulator with the reference governor, and No rail band the MPC without its rail rows.

🕮 **Pause.** Try Exercise 8.3 before reading on.

8.6.9 Moving the Cart

Case 9 moves the target from 0 to 0.5 m at 1 s while balancing on the Kalman filter, with the reference walking at 0.25 m/s. The cart edges right 0.83 mm, backs away 7.1 mm and settles inside 2 cm and 2° 3.23 s after the step, the links within 1.43°. At 1 m/s the dip is 27.6 mm and the settling 2.14 s with the lower link at 5.32°, and at 2 m/s the dip is 50.8 mm and the settling 2.01 s at 6.51°. Each doubling of the rate buys less time and costs a deeper dip.

8.6.10 The Regulator Against the Predictive Controller

Case 10 balances on the Kalman filter with the cart target at +0.6 m, 0.4 m from the right bumper, and the drive lag modeled in both designs, and at 1 s it kicks the upper link at 60 °/s toward $+x$, enough to hold the command at the drive limit. Table 8.6 compares the regulator and the MPC on that case, on its variants with the other estimators, larger kicks and a 40 N drive, and on the two other cases where the two laws part, and Figure 8.5 draws each run's time at the drive limit and its fate. Every run lasts 10 s from the named case, the stated change being applied before the start. The regulator keeps its shipped weights, and the MPC runs with the settings of case 10 throughout, with nothing retuned for a row. Both read the same encoders through the same estimator, and nothing in the application is random. Two further laws isolate the rail. The regulator with a reference governor [21] moves its cart reference at each sample to the value nearest the target for which its own unconstrained prediction keeps $|x| \leq 0.95$ m over 1 s, which gives the regulator the same rail knowledge by the classical route. The MPC with its rail rows removed is the counterfactual for the rail band. The loop pair falls in every row in which it was run, between 1.36 and 1.74 s.

The case as shipped is a tie. Both laws bring the cart to 0.114 m from the bumper, the MPC at 1.605 s and the regulator at 1.610 s, with the links at 13.8° and 6.3°, and they settle 2.18 s and 2.21 s after the start. The two carts never differ by more than 1.87 mm between 0.8 and 2.8 s, as shown in Figure 8.4. The difference lies in the force. At 1.042 s the regulator asks for −22.36 N and the drive delivers −20 N, while the MPC plans −20.00 N with a departure of +2.36 N, its largest. The first plan that the force limit shapes comes at 1.031 s and predicts the cart out to 0.72 m. The command sits at the limit for 15 ms under the MPC against 25 ms under the regulator, and no plan of the 1000 is rejected.

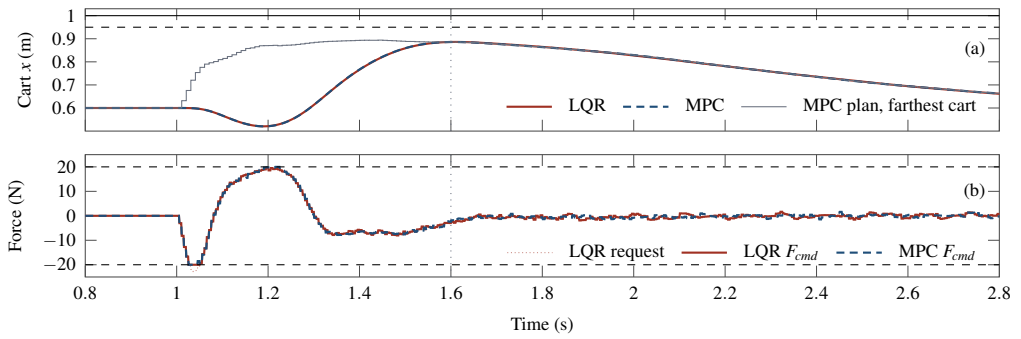


Figure 8.4 Case 10 run by the regulator and by the MPC on a common model with common weights, over a 2 s window from 0.8 to 2.8 s with the 60 °/s kick at 1 s, on every 1 ms step, the regulator as a thick solid line and the MPC as a thick dashed one. (a) Cart position against the bumper (thin horizontal line) and the MPC's soft limit at 0.95 m (thin dashed), with the farthest cart position predicted by the latest accepted plan as a thin stepped line. (b) The commanded force against the 20 N drive limit (thin dashed), with the regulator's unclipped request as a thin dotted line, which peaks at −22.78 N at 1.036 s. At the dotted line, $t = 1.60$ s, both carts are at 0.886 m, within 10 ms of their closest approach to the bumper.

Knowing the limit wins where the force is noisy or large. On differenced rates the regulator's chattering command sits at the limit for 790 ms and the pendulum falls at 1.95 s, while the MPC holds with one touch of the bumper, 135 ms at the limit and one rejected plan. With a 70 °/s kick on the Kalman filter the regulator falls at 1.97 s after two bumper contacts and the MPC holds after one. The same knowledge also costs, because the plan is linear. With perfect states and a 65 °/s kick the regulator holds with the lower link at 16.9°, and the MPC keeps the force at the limit longer than links leaning past 20° can take, reaches 21.2° on the lower link and falls at 2.27 s. The two runs part by 1° on the upper link at 1.523 s. In case 9 with the target at 0.6 m, differenced rates and a 0.4 N s push the MPC falls at 1.94 s where the regulator holds with 0.363 m to spare, the runs parting at 1.347 s with both links still under 6°. On the down-up target of case 3 the regulator holds within 21.8° and the MPC falls at 1.82 s after four rejected plans. At a 40 N limit the regulator survives the 85 and 90 °/s kicks with one and two slow contacts with the bumper, and the MPC falls at 1.81 and 1.83 s.

The rail band causes the two 40 N losses. Without its rail rows the MPC holds both kicks with one contact each, and it holds every case of the table that it holds with them. Avoiding a slow contact that the regulator survives costs a lean that the linear plan misjudges, and the band's only gain in the table is at 80 °/s, 0.039 m of rail left against 0.028 m. The governor changes the regulator's runs only at 80 and 85 °/s on the 40 N drive, leaving 0.035 m at 80 °/s, and it never ends a run worse than the regulator. The losses where the band plays no part, at 65 °/s on perfect states, after the differenced push and on the down-up target, also occur without it. A comparison built on a push that drives the regulator into the rail end is not available at the shipped limit. With the target at +0.6 m and the drive lag left out of the design, the regulator catches pushes up to 0.45 N s and kicks up to 30 °/s with at least 0.27 m of rail left, and every larger push or kick first holds the drive at 20 N and loses the links, the bumper being met only with a link 26.9° or more from upright or after the watchdog trip. The rail comes within reach of a kick that the drive can survive only at the 40 N limit.

Where no limit is in reach the laws do not part. The case 9 target steps at 0.25 and 2 m/s and from −0.6 m to +0.6 m at 2 m/s end alike on every estimator, the last on differenced rates falling under both at 3.27 and 3.21 s. On the ten existing scenarios the MPC is nowhere worse than the regulator. Among 49 changes a learner can make it does worse on seven, the five losses above, the down-up target with a kick, and a 24.8 m/s^2 gravity under which both fall, the MPC after four bumper contacts and the regulator after one, and better on three, the two wins and angle scales of 1°. The watchdog,

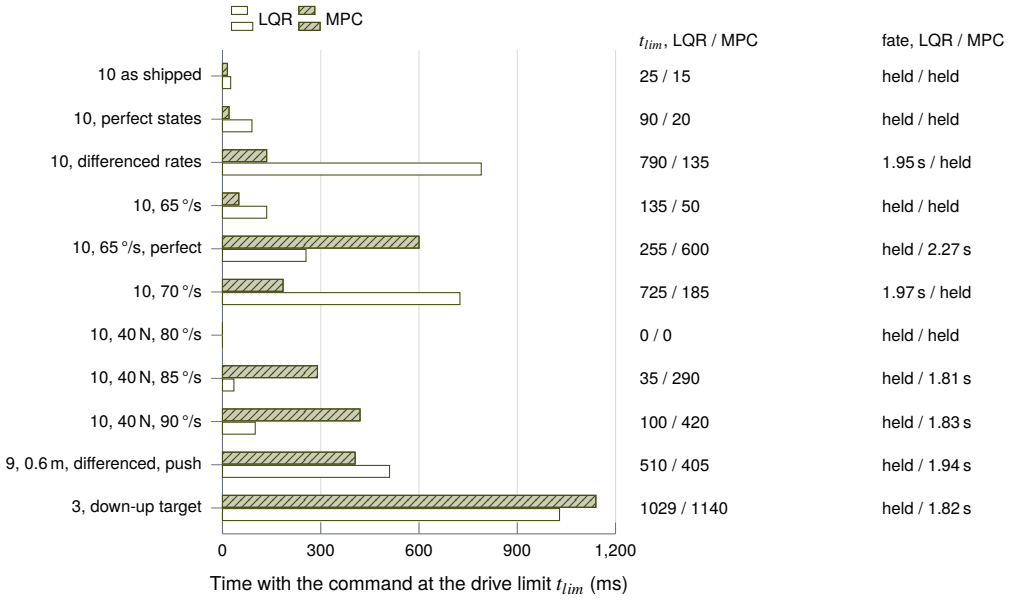


Figure 8.5 The regulator (white) and the predictive controller (hatched) of Table 8.6 on one model with one set of weights: the time each run spends with the command at the drive limit, and its fate, held or the time of the watchdog trip. Notice that the predictive controller holds both runs that the regulator loses, with less time at the limit, while in four of the five runs that it loses itself it stays at the limit longer than the regulator does.

the bumpers, the deck stop and the drive clamp act under every structure.

8.6.11 The Challenge

The challenge starts hanging with the target at +0.5 m and asks for both links inside $\pm 2°$ and the cart inside ± 2 cm of the target for good by 5.0 s of a 10 s run, with no bumper contact and a force chatter of at most 2.0 N RMS. As shipped, on differenced rates with the nominal weights, it settles at 5.93 s with 6.55 N of chatter. Each part of the target fails a different change. The Kalman filter alone quiets the force to 0.40 N but settles at 6.27 s, a 1 m/s reference alone settles at 4.72 s with 6.69 N of chatter, and a stiffer cart weight of $x_s = 0.15$ m on differenced rates drives the cart into a bumper. The suggested configuration, the Kalman filter with $x_s = 0.15$ m and a 1 m/s reference, settles at 4.59 s with 0.59 N of chatter and the cart never past 0.600 m. The shipped run is therefore 0.93 s late with 3.3 times the allowed chatter, and the suggested one meets the target with margins of 0.41 s and 1.41 N, whether its settings are made before the run or carried through a reset. As in other apps (Section 1.6.4), a change to a model parameter or a disturbance voids the attempt.

8.7 Technology in Practice

What Measures It

Laboratory pendulums on carts read their arm angles with encoders or with a motion-capture camera, and send the data over wires or a wireless link [22]. An encoder on a joint above the cart rides on a moving arm, so its signals must cross every joint below it. One open multi-link rig passes them through a slip ring, which keeps the latency low and needs no computing to recover the angle, and so suits high sampling rates [22]. The chapter's three encoders, on the cart, the lower pivot and the elbow, only round to their counts, with no noise, no lost count and no delay (Section 8.3).

TECHNOLOGY Sensor: Encoders That Ride on a Swinging Link

On an open multi-link rig each arm carries an optical encoder of 10 000 counts per revolution, and the cart's linear stage has a magnetic encoder with a resolution of 1 µm. An encoder above the first joint turns with the arm below it, so its signals leave through a slip ring, with the wires run through a hollow shaft. The slip ring adds no latency, unlike camera or wireless links, and needs no battery on the arm. Its brush contact adds friction at the joint, kept small with gold contacts that also cut electrical noise, and its five or eight channels leave little room to add a gyro [22]. The chapter's elbow encoder reads the angle between the links to 4096 counts a turn, with no friction beyond the elbow damping b_{p2}.

What Moves It

The same rig drives its cart with a linear motor whose stage carries its own encoder, and the servo drive closes a speed loop around it, so the controller's output is a commanded velocity rather than a force [22]. The chapter's drive takes a force command instead and develops it through one lag of 30 ms, a slew limit of 1000 N/s and a limit of 20 N, the lumped stand-in for the amplifier, the motor and the belt that its Assumptions name.

TECHNOLOGY Final Element: A Linear Motor in Place of a Belt

Carts on laboratory rigs are driven by a rotary motor, a servo motor with a belt, or a linear motor, and the choice affects how robust a control law proves. A linear motor avoids the backlash of a belt-driven cart, which makes a multi-link pendulum harder to control. On one open rig a servo drive in velocity mode reads the controller's analog voltage as a speed, compares it with the speed from the stage encoder and computes the motor current itself, over a stroke of 2 m [22]. The chapter's drive is a force source behind a 30 ms lag, and its Assumptions name the backlash that a real belt would add.

Where the Controller Runs

The open rig runs its law on a real-time target computer, with the design done on a separate host. Reading and sending its signals takes about 31.5 µs and a large block-diagram model about 56 µs, and in practice the rig samples at about 5 kHz while a time-varying LQR and a Kalman filter stabilize the double or triple pendulum. Its builders ask for at least 1 kHz for that stabilization [22]. The chapter's law runs every 5 ms, 200 Hz, on a 1 ms integration step, and Figure 8.2 shows why its unmodeled drive lag, not the sampler, sets the slowest safe rate. The state feedback $F = -K(\hat{z} - z^*)$ is the one block that holds the chapter's structure: the stored swing-up schedules its gain along the trajectory, and the predictive controller adds departures to its force. Figure 8.6 draws the loop on rig hardware.

What Is Modernizing It

The first development is the direct drive. A linear motor removes the belt and its backlash, and its stage encoder resolves 1 µm over the stroke [22]. The second is open hardware. The rig's drawings, tutorial code and data sets of the arms' angles and speeds, with and without control, are published so that system identification and machine learning methods can be tested on the same machine [22]. The third is the remote experiment. A user's controller would be checked first on a digital twin of the pendulum, sent back with a report if it broke a limit on position, speed or acceleration, and run on the hardware only after passing, with fail-safe limits that act outside the user's code [22].

IN PRACTICE Before the Cart Moves: Switches, Absorbers and the Power

On a documented open rig, limit switches at the ends of the travel tell the real-time controller to stop the motor. A shock absorber at the side of the rail takes up the cart's kinetic energy if the controller fails, and a protective panel guards people from parts thrown off the arm. Before a run the operator checks every wiring connection, and a run that does not go as planned is stopped at once with all power cut [22]. The chapter's watchdog and bumpers play the parts of the switch and the absorber.

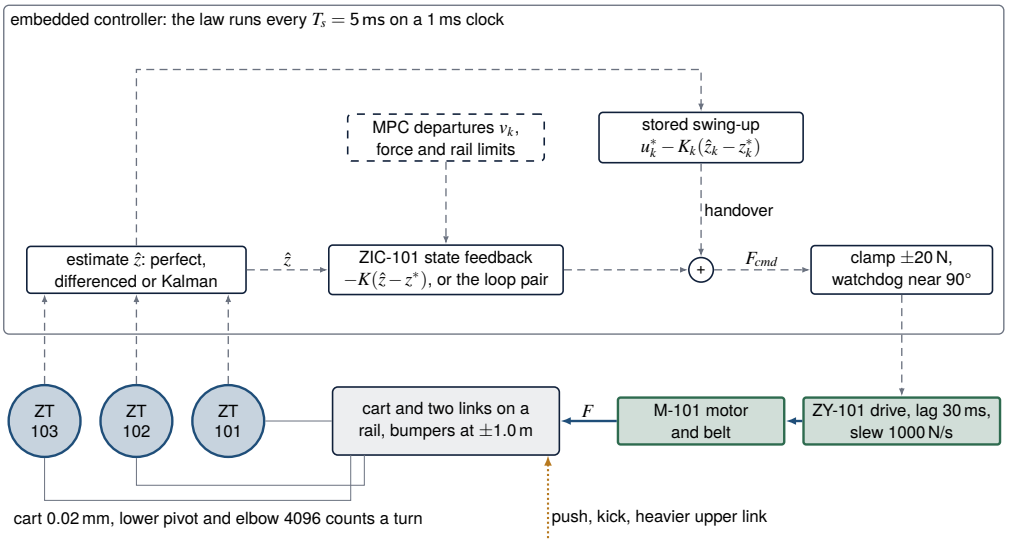


Figure 8.6 The balance loop on rig hardware. The encoders ZT-101 on the cart, ZT-102 at the lower pivot and ZT-103 at the elbow, which reads the angle between the links, feed three estimators that run in parallel. The state feedback ZIC-101 holds the regulator or the loop pair, the predictive controller (dashed) adds departures to the regulator's force, and in the swing-up the stored force with its time-varying feedback drives until the handover. The clamped command goes to the drive ZY-101 and the motor M-101. Notice that one force serves the cart and both links, so every block above the drive shares one limit of 20 N.

8.8 Limitations

Stick-slip hunting, the bending of a long upper link and sensor noise are all absent, because both rods are rigid, every friction is viscous and each encoder only rounds. A single lag and slew limit stand for the whole drive, and a real belt or rack would add backlash and voltage limits that shrink the margin of Figure 8.2.

The design models are perfect apart from the changes a scenario makes, and the present-links option redesigns instantly with the true links, which favors the regulator in cases 6 and 8 over any real redesign.

The analysis of Section 8.4.2 is a set of necessary conditions on the continuous, lag-free model, and the margin of Figure 8.2 is a linear stability boundary that says nothing about the size of disturbance a stable loop recovers. The handover set is a level set of a linear cost-to-go, not a verified region of attraction.

> ⚠ **OPEN CHALLENGE** **Proving a Handover Region for More Links**
>
> Verified feedback motion planning checks each funnel with sums-of-squares programs. Trajectory optimization and time-varying LQR scale well with the number of states, but this verification does not. It has succeeded on systems of up to five dimensions, with examples of six in the literature, and at high dimension the programs grow into very large semidefinite programs. Sampling can replace the proof at much lower cost, at the price of the guarantee [15]. The double pendulum on a cart already has six mechanical states, and the chapter's handover set is not a verified region of attraction.

The stored swing-up exists for one rig, one drive and one sample interval, and the application offers no re-solve in the browser.

The MPC plans on the regulator's linearization at the target and knows the drive limit and the rail, but not the nonlinear terms, the slew limit, the bumpers, the deck stop or the encoder quantization.

Near the edge of what the drive can do its plan can be confidently wrong, and it falls in five measured cases where the clipped regulator holds. Its force limit is judged on 65 of the 200 planned samples and its rail limit is soft, on one slack for the whole horizon. It has no disturbance estimator, which the level rail and the absence of a steady force on the cart make unnecessary here but a real rig with cogging or a tilted rail would need, and it is redesigned instantly from the present constants, as the regulator is.

8.9 Conclusion

A double inverted pendulum on a cart has two unstable modes and one motor, and the simulator of this chapter turns each requirement that follows into an experiment. The fast mode multiplies by e every 90 ms, so with nothing pushing, a pendulum released 0.5° from upright trips the drive 0.615 s later.

Loops that push toward both leans cannot hold it, because the angle gains may sum to no more than 12.26 N/rad while the lower link alone needs more than 58.6 N/rad, and the regulator that does hold it pushes against the upper link's lean at 315.7 N/rad. Differenced counts make its force chatter by 7.17 N where a Kalman filter on the same readings leaves 0.43 N. Its sample interval is limited by the lag it does not model, 12.3 ms at the drive's 30 ms, ten times shorter than the Nyquist interval of the motion that escapes.

A linear peak-force rule predicts the catchable band at 20 N without a false positive, and feedback along the stored swing-up absorbs upper-link mass errors from −27 % to +27 % where the stored force alone needs the exact rig. A constrained MPC on the regulator's own model ties the regulator in case 10, with 0.114 m of rail left under both, holds the differenced and 70 °/s kicks that the clipped regulator loses, and falls in five cases that the regulator holds, two of them because of its rail band.

The rods of the model are rigid and its encoders noiseless. A fuller model would add encoder noise and delay, backlash in the drive, a redesign from identified rather than true links, and a swing-up re-solved in the browser for the rig at hand. The simulator gives one plant on which every requirement of full-state design can be taken away and its failure measured.

KEY POINTS

- Two rods on one cart have two ways to fall, and the faster grows by e every 90 ms.
- Angle loops that push toward both leans cannot hold it: their gains may sum to no more than 12.26 N/rad, while the lower link alone needs more than 58.6 N/rad.
- Differenced counts make the force chatter by 7.17 N, a Kalman filter on the same readings by 0.43 N.
- The sample interval is limited by the lag the design leaves out, 12.3 ms at the drive's 30 ms.
- Feedback along the stored swing-up absorbs upper-link mass errors from −27 % to +27 %.

Where else this applies. The same cart and drive hold a single rod in Chapter 7.

References

[1] K. Furuta, T. Okutani, and H. Sone, "Computer control of a double inverted pendulum," *Computers & Electrical Engineering*, vol. 5, no. 1, pp. 67–84, 1978.

[2] M. Yamakita, K. Nonaka, and K. Furuta, "Swing up control of a double pendulum," in *1993 American Control Conference*. IEEE, 1993, pp. 2229–2233.

[3] F. B. Horak and L. M. Nashner, "Central programming of postural movements: adaptation to altered support-surface configurations," *Journal of Neurophysiology*, vol. 55, no. 6, pp. 1369–1381, 1986.

[4] A. D. Kuo, "An optimal control model for analyzing human postural balance," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 87–101, 1995.

[5] W. Singhose, D. Kim, and M. Kenison, "Input shaping control of double-pendulum bridge crane oscillations," *Journal of Dynamic Systems, Measurement, and Control*, vol. 130, no. 3, p. 034504, 2008.

[6] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4906–4913.

[7] K. Furuta, H. Kajiwara, and K. Kosuge, "Digital control of a double inverted pendulum on an inclined rail," *International Journal of Control*, vol. 32, no. 5, pp. 907–924, 1980.

[8] M. Yamakita, M. Iwashiro, Y. Sugahara, and K. Furuta, "Robust swing up control of double pendulum," in *Proceedings of 1995 American Control Conference – ACC'95*, vol. 1, 1995, pp. 290–295.

[9] J. Rubí, Á. Rubio, and A. Avello, "Swing-up control problem for a self-erecting double inverted pendulum," *IEE Proceedings – Control Theory and Applications*, vol. 149, no. 2, pp. 169–175, 2002.

[10] K. Graichen, M. Treuer, and M. Zeitz, "Swing-up of the double pendulum on a cart by feedforward and feedback control with experimental validation," *Automatica*, vol. 43, no. 1, pp. 63–71, 2007.

[11] W. Zhong and H. Röck, "Energy and passivity based control of the double inverted pendulum on a cart," in *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)*. IEEE, 2001, pp. 896–901.

[12] T. Glück, A. Eder, and A. Kugi, "Swing-up control of a triple pendulum on a cart with experimental validation," *Automatica*, vol. 49, no. 3, pp. 801–808, 2013.

[13] D. Mayne, "A second-order gradient method for determining optimal trajectories of non-linear discrete-time systems," *International Journal of Control*, vol. 3, no. 1, pp. 85–95, 1966.

[14] W. Li and E. Todorov, "Iterative linear quadratic regulator design for nonlinear biological movement systems," in *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, vol. 1. SciTePress, 2004, pp. 222–229.

[15] R. Tedrake, I. R. Manchester, M. Tobenkin, and J. W. Roberts, "LQR-trees: Feedback motion planning via sums-of-squares verification," *The International Journal of Robotics Research*, vol. 29, no. 8, pp. 1038–1052, 2010.

[16] R. H. Brown, S. C. Schneider, and M. G. Mulligan, "Analysis of algorithms for velocity estimation from discrete position versus time data," *IEEE Transactions on Industrial Electronics*, vol. 39, no. 1, pp. 11–19, 1992.

[17] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

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

[19] B. Kouvaritakis, J. A. Rossiter, and J. Schuurmans, "Efficient robust predictive control," *IEEE Transactions on Automatic Control*, vol. 45, no. 8, pp. 1545–1549, 2000.

[20] P. O. M. Scokaert and J. B. Rawlings, "Constrained linear quadratic regulation," *IEEE Transactions on Automatic Control*, vol. 43, no. 8, pp. 1163–1169, 1998.

[21] E. Garone, S. Di Cairano, and I. Kolmanovsky, "Reference and command governors for systems with constraints: A survey on theory and applications," *Automatica*, vol. 75, pp. 306–328, 2017.

[22] K. Kaheman, U. Fasel, J. J. Bramburger, B. Strom, J. N. Kutz, and S. L. Brunton, "The experimental multi-arm pendulum on a cart: A benchmark system for chaos, learning, and control," *HardwareX*, vol. 15, p. e00465, 2023.

8.10 Exercises

8.1 ○ Concept check · 2 min

The loop pair of case 1 puts 40 N/rad and 4 N s/rad on each link angle, each loop pushing the cart toward its own link's lean, beside a cart loop that pulls the cart toward its target. Why does no retuning with all gains positive hold the pendulum?

(a) Derivative action on both links cannot be made large enough to damp a mode that grows by e every 90 ms

(b) The 5 ms sample interval is too long for the fast mode, so every sampled pair of loops goes unstable

(c) The 20 N drive limit saturates before any tuning with positive gains can catch the fast mode

(d) The angle gains enter the s^2 coefficient only through their sum, which must stay below the gain the s^4 coefficient needs

8.2 Concept check · 2 min

In case 4 the sample interval goes from 5 to 20 ms at 3 s, and the regulator and the Kalman filter are redesigned for 20 ms on the zero-order-hold model at once. The pendulum still falls. Which explanation does the chapter support?

(a) The design leaves out the 30 ms drive lag, and the lag a design can absorb shrinks as the interval grows

(b) The 3.9 Hz oscillation that escapes lies above the Nyquist frequency of a 20 ms sampler

(c) The redesign ignores the zero-order hold, whose half sample of lag the new gains do not expect

(d) The redesigned Kalman filter converges too slowly at 20 ms to supply the rates the law needs

8.3 Concept check · 2 min

Case 8 plays the stored swing-up force alone, with no feedback along the trajectory, on a rig whose upper link is 10 % heavier, a payload of 10 g. Which statement explains why it misses the handover?

(a) The heavier link needs more force than the 17 N bound that the trajectory leaves below the drive limit

(b) The trajectory runs through unstable states, so a small model error grows unless feedback corrects it

(c) The swing-up reaches the top, but the balance design for the default links cannot hold the heavier link

(d) The trajectory starts before the links hang still, and that initial error rather than the mass decides

8.4 Theory and calculation · 7 min

The default rig has a cart of $M = 1.0\,\mathrm{kg}$, a lower rod with $l_1 = 0.5\,\mathrm{m}$ and $m_1 = 0.15\,\mathrm{kg}$, an upper rod with $l_2 = 0.3\,\mathrm{m}$ and $m_2 = 0.10\,\mathrm{kg}$, and $g = 9.81\,\mathrm{m/s^2}$.

(a) Evaluate h_1 to h_6 of Equation (8.2) and the energy $2(h_7 + h_8)$ that a swing-up from hanging at rest to upright at rest must supply. *Checkpoint:* $h_1 = 1.25\,\mathrm{kg}$, $h_2 = 0.0875\,\mathrm{kg\,m}$, $h_3 = 0.015\,\mathrm{kg\,m}$, $h_4 = 0.0375\,\mathrm{kg\,m^2}$, $h_5 = 0.0075\,\mathrm{kg\,m^2}$, $h_6 = 0.003\,\mathrm{kg\,m^2}$, and 2.011 J.

(b) With $K_{px} = 0$, find the cap on $K_{p1} + K_{p2}$ that $a_2 > 0$ sets and the threshold a_4^0/α that $a_4 > 0$ sets on the lower link's gain, with $\gamma = h_3 h_4 + h_2 h_6$, $\alpha = h_2 h_6 - h_3 h_5$ and $a_4^0 = g[h_1\gamma - h_2 h_3(h_2 + h_3)]$. *Checkpoint:* 12.26 N/rad and 58.6 N/rad.

(c) Check both conditions for the shipped pair, $K_{p1} = K_{p2} = 40\,\mathrm{N/rad}$, using $\beta/\alpha = 0.625$. *Checkpoint:* Both fail: the sum of 80 exceeds the cap by 67.7 N/rad, and $40 - 0.625 \times 40 = 15\,\mathrm{N/rad}$ is below 58.6 N/rad.

8.5 Theory and calculation · 6 min

The fast unstable root of the default rig is $\lambda_f = 11.07\,\mathrm{s^{-1}}$.

(a) Find the time in which the fast mode grows by a factor e, and the factor by which it grows over one sample of 5 ms, one of 20 ms, and 200 ms. *Checkpoint:* 90.3 ms, then 1.057, 1.248 and 9.15.

(b) After the switch to 20 ms in case 4 the lightly damped pair crosses to $+1.64 \pm 24.28i\,\mathrm{s^{-1}}$. Find its frequency in hertz and compare it with the Nyquist frequency of the 20 ms sampler. *Checkpoint:* 3.86 Hz against 25 Hz, 6.5 times lower.

(c) At the default 30 ms drive lag the critical interval is 12.28 ms, and the pair that escapes there oscillates at 26.0 rad/s. Find the Nyquist interval π/ω of that pair and the drive lag plus half the critical interval. *Checkpoint:* 120.8 ms, 9.8 times the critical interval, and 36.1 ms.

8.6 Theory and calculation · 5 min

The joint encoders count $N_\theta = 4096$ per turn, the law runs every $T_s = 5\,\text{ms}$, the differenced rate passes the filter of Equation (8.9) with $\tau_f = 5\,\text{ms}$, and the regulator's gain on $\dot{\theta}_2$ is 34.7 N s/rad.

(a) Find one count in degrees and the rate step that one count makes when it is differenced over one sample. *Checkpoint:* 0.0879°, and 17.6 °/s or 0.307 rad/s.

(b) Find η, and the step in the commanded force that one count of the elbow encoder causes after the first filter step and in full. *Checkpoint:* $\eta = 0.632$, with 6.7 N and 10.6 N.

(c) Find η for $\tau_f = 15\,\text{ms}$, and say why case 3 shows that a longer filter is no cure for the chatter. *Checkpoint:* $\eta = 0.283$. The pendulum falls 0.855 s into the run.

8.7 App challenge · 7 min

Load scenario 1, Balance it by hand, at speed 1 with the default seed. The upper link is let go 0.5° from upright at 1 s with the drive at zero.

(a) Press Run in Manual and touch nothing until the app pauses itself, about 1 s after the fall. Read the Pendulum tile, and find in the exported file the first row at which the upper link angle, column `th2_deg`, is beyond 2°. *Checkpoint:* Fell at 1.61 s, and the row at 1.205 s.

(b) Press Reset, switch to Auto so that the PID pair takes the drive, and run until the app pauses itself. Read the Pendulum tile. *Checkpoint:* Fell at 1.53 s.

(c) Press Reset, set Controller structure to LQR, keep Auto, and run to 5 s. Read the Cart travel tile, and find the largest $|$`th2_deg`$|$ after 1 s in the exported file. *Checkpoint:* 0.015 m and 0.64°, within 0.01°.

8.8 App challenge · 8 min

Load scenario 3, Encoders, differences and a Kalman filter. The upper link starts 2° from upright, and the regulator reads rates differenced from the encoder counts through the 5 ms rate filter. Use the default seed.

(a) Run as shipped to 10 s. Read the Force chatter and Rate error tiles. *Checkpoint:* 7.17 N and 5.8 °/s.

(b) Press Reset, set State estimate to Kalman, and run to 10 s again. *Checkpoint:* 0.43 N and 0.5 °/s.

(c) Press the reload button, switch to Advanced, set Rate filter to 15 ms with State estimate on Differenced, and run until the app pauses itself. Read the Pendulum tile. *Checkpoint:* Fell at 0.86 s.

Accept a chatter within 0.02 N. Explain the result of (c).

8.9 App challenge · 9 min

Load the Challenge, swing up, balance and move. Use the default seed and run each case for the full 10 s. Press the reload button before each part and make the part's settings before Run. After each run read the challenge result in the Learn panel, the settling instant and the force chatter.

(a) Run as shipped, on differenced rates with the nominal weights. *Checkpoint:* Settled at 5.93 s, chatter 6.55 N, not met.

(b) Set State estimate to Kalman and change nothing else. *Checkpoint:* 6.27 s, 0.40 N, not met.

(c) Set State estimate to Kalman and Cart reference rate to 1 m/s. *Checkpoint:* 4.81 s, 0.41 N, met.

(d) Set State estimate to Kalman and Cart excursion scale to 0.15 m, with the reference rate left at 0.25 m/s. *Checkpoint:* 6.00 s, 0.54 N, not met.

Say which part of the target each change addresses.

8.10 Application area · 4 min

A laboratory builds this rig with a belt drive. A step test shows that the belt's backlash and the amplifier's voltage limit stretch the effective drive lag from 30 to 35 ms. The regulator is designed without the lag and samples every 5 ms.

(a) Using Figure 8.2 and the results of case 4, say whether 5 ms samples are still safe. *Checkpoint:* No margin is left: the tolerable lag at 5 ms is 35.1 ms.

(b) Name two changes that restore a margin, and what to check after either one. *Checkpoint:* Model the lag in the design, or sample faster.

8.11 Application area · 3 min

A balancing robot built on this design lifts payloads that make its upper link up to 40 % heavier. It swings up along the one stored trajectory with the feedback on and balances with the design for the default links. Using the results of case 8, say which payloads it handles and what must be added for the rest.

8.12 Review the agent · 5 min

An AI assistant reviews case 10 and writes: "The MPC uses the regulator's own model, weights and Riccati solution, so with no limit in reach its force equals the regulator's, and near a limit it plans within the limit instead of being clipped. A controller that does everything the LQR does and also respects the limits cannot do worse. Make the MPC the default balance law on this rig." Find the flaw in the argument and say how to check it in the app.

CHAPTER 9

A Riderless Benchmark Bicycle That Rides Itself Only Between Two Speeds



The controller moves 1 steering motor that turns the front fork

The controller reads 1 rate gyro and accelerometer that sense the lean 2 encoder that reads the steer angle

What disturbs it 1 a side gust at the center of mass

Notice that the motor can only turn the front wheel: the frame stays up only if each steer brings the wheels back under its lean. Illustration, not to scale.

IN PLAIN TERMS A Bicycle That Steers Itself Upright

The main idea. A bicycle stays up because its front wheel steers under its fall. It is like a hand that moves under a stick balanced on a fingertip. With nobody on it, a bicycle does this by itself, but only inside a narrow band of speeds.

The challenge. Below that band it wobbles and falls. Above it the bicycle leans slowly into a spiral, so a motor on the handlebars must take over. Settings that hold it at one speed fail at another, and at a walking pace no controller tried can hold it up.

Why it matters. Aircraft, engines and process plants also change with speed or load, and their controllers must change with them.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the history segments and the fact in Section 9.2, the try-its in Sections 9.4 and 9.6, the fact on trail in Section 9.6, 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 9.6), Technology in Practice (Section 9.7), Limitations (Section 9.8) and the open challenge

The control question A riderless bicycle falls below 4.29 m/s, rides itself up to 6.02 m/s and slowly tips over above it. What does a speed-dependent plant ask of a lean loop, a schedule, a regulator, a rider and a path loop, and at what speed does no controller help?

Learning objectives

- Read the weave, capsize and castering modes and the self-stable window from the eigenvalues of the benchmark bicycle against speed.
- Compute the steer's grip on the lean, $gK_{0\varphi\delta} + v^2K_{2\varphi\delta}$, and explain why it rules out balance at walking pace.
- Explain why a balance loop must steer into the fall, and why a left turn starts with right steer.
- Compare a fixed lean PD, its two-gain schedule and a scheduled LQR across speed, counting the controller's sample hold.
- Order a path loop, a lean loop and a delayed rider by speed, and test each claim on the simulator.

Prerequisites Sections 1.4.3 (gain schedule), 1.4.4 (eigenvalues), 1.4.5 (LQR, Bryson's rule), 1.4.7 (RHP zero, delay) and 1.6.1 (engine). Chapter 7 for a cascade that commands an angle.

The app https://apmonitor.com/apps/bicycle-balance/

See also Volume 20, Chapter 2 for a path-following loop on a kinematic bicycle whose behavior changes with speed. Volume 3, Chapter 10 for an outer loop that commands an angle through a right-half-plane zero, and Volume 3, Chapter 9 for the bandwidth such a zero allows.

9.1 Introduction

A bicycle released at 3 m/s with its frame leaning 2° to the right is on the ground 2.00 s later. Released the same way at 5 m/s, it steers itself upright and the lean is below 0.5° by 5.5 s. At 8 m/s it leans slowly into a tightening spiral and passes 40° after 20.2 s. These three runs come from the simulator of this chapter, whose bicycle is the benchmark of Meijaard, Papadopoulos, Ruina and Schwab, stable with no rider and no control only between 4.292 and 6.024 m/s [1]. Nothing in the frame changes from one run to the next. Only the speed does.

Riding a bicycle teaches two beliefs that this simulator is designed to unsettle. The first is that stability is a property of a machine. The frame that rides itself at 5 m/s has a mode growing at 3.53 s^{-1} at 1 m/s, which doubles the lean every 0.20 s, and a lean loop tuned at 2 m/s is unstable on the same frame above 3.45 m/s. The second is that steering left turns a bicycle left. Asked for a 5° lean to the left at 5 m/s, the regulator in the simulator first steers 4.3° to the right, because a bicycle turns only once it leans and leans only once its wheels have run out from under it. A student reasoning from intuition may also get the sign of the balance loop backward, and with the wrong sign a lean loop drops the bicycle in 0.61 s, against 1.66 s with nobody steering at all.

The bicycle belongs to a broad class, plants whose stability and gain are set by an operating variable that the controller does not choose. The relation between a bicycle's design and its behavior depends heavily on speed [2], the stability of motorcycles has been compared across models that differ in their tire assumptions [3], and aircraft and engines are operated across their envelopes on gain schedules, one linear design for each operating condition [4]. A process unit whose throughput changes moves its own poles in the same way. Two further features of the bicycle are general. The response from steer to heading starts the wrong way, a right-half-plane zero that bounds what any loop wrapped around it can achieve [5, 6], and the authority of the steer over the lean depends on the speed, so that the controllability of the lateral motion is a question of its own [7].

The equations, the modes and the rider models of the bicycle are presented on paper [2, 8, 9], and an instrumented bicycle of the kind that tests them [10–12] falls over when a sign is wrong. This chapter shows what the self-stable window of the benchmark bicycle means for a lean loop, a schedule, a regulator, a rider and a path loop, on a browser simulator in which the eigenvalue map against speed, the running bicycle and the controller that holds it share one model. The equations, the benchmark values, the proportional-derivative (PD) and linear quadratic regulator (LQR) laws and

gain scheduling are all established. Section 9.2 follows the problem from the safety bicycle to the benchmark. Section 9.3 gives the model: matrices derived from 23 design parameters that equal the published ones to 1.4×10^{-14}, a nine-state model with a handlebar stop, a fall and a course, and a measurement boundary between the bicycle and its controller, so that the simulator can be rebuilt from this chapter. Section 9.4 reads the window in speed. Section 9.5 gives the controllers as coded and shows that the controller's 10 ms hold moves the stability edge of a lean PD tuned at 2 m/s from 5.59 m/s, where an analysis that applies the law continuously puts it, to 3.45 m/s, and that a two-gain schedule of the same loop leaves a band from 2.81 to 4.29 m/s in which the closed loop grows at up to 0.157 s^{-1}, while a scheduled regulator stays stable from 0.5 to 10 m/s. Section 9.6 reports the nine cases and the graded slalom as measured. Each is scored by a rule that credits completion, and no tuning can win case 9. Section 9.7 describes the sensors, steering motors and controllers of real test bicycles, Section 9.8 states the limitations, and the exercises of Section 9.10 work through the cases.

9.2 Self-Steering Bicycles, From the Safety Bicycle to the Benchmark

Balancing by steering is as old as the two-wheeler, and the history segment below follows it from the running machine to the first equations of a self-stable bicycle [1]. The model that carries Whipple's name, four rigid bodies on knife-edge wheels, is the one simulated here.

> ⧖ HISTORY **Balancing by Steering, 1818 to 1899**
>
> The review that accompanies the benchmark traces the idea. In 1818 Karl von Drais showed that a rider on two wheels in line could balance by steering the front wheel, and the pedal velocipede of the 1860s, with its vertical steering axis, was balanced by the rider's steering as well. A patent of 1866 stated the rule plainly: a carriage that leans to the right is righted by turning the wheel to the right. William Rankine's account of 1869 seems to be the first description of countersteering. By the 1890s the safety bicycle had a tilted steer axis and a bent fork, like a modern frame. In 1897 Emmanuel Carvallo and then, more generally, Francis Whipple in 1899 showed with rigid-body equations that some safety bicycles balance themselves in the right range of speed [1].

What makes a bicycle steer itself upright was argued for most of the following century. In 1970 Jones built bicycles from which selected features had been removed, among them one designed with computer calculations to be unstable [13], and for the bicycles he tried both the spin of the front wheel and a positive trail were needed [1]. Motorcycle models of the following year compared three assumptions about the tires, the stability characteristics each predicts and the effect of design changes [3]. By 2007 about 30 rigid-body models general enough to allow self-stability had been published, several of them in disagreement with the equations the benchmark then confirmed [1].

The benchmark of 2007 settled the Whipple model. Its authors derived the linearized equations by hand in two ways, checked them against two nonlinear simulation programs, and published one fully specified bicycle with its four matrices and its eigenvalues to fourteen digits so that any later formulation could be checked against them [1]. A companion paper wrote the nonlinear equations of the same bicycle and found three families of hands-free circular motions, two of them starting at the capsize bifurcation [14]. Measurements on a riderless bicycle whose masses and geometry had been measured found eigenvalues in good agreement with the formulas [1, 11], and in 2011 a bicycle built without gyroscopic or caster effects proved self-stable all the same, which points to the distribution of mass [12]. Case 3 of the simulator lets a student watch the benchmark's window move with the trail and the steer axis tilt.

DID YOU KNOW? Measuring the Window by Running Alongside

To test the model of an uncontrolled bicycle, a riderless bicycle with side wheels was launched 76 times in a university sports hall by a person running beside it. Most runs, 56, fell between 3.5 and 5 m/s, where the weave turns stable. Below 2 m/s the bicycle was so unstable that the time left to measure was very short [11].

Control came to the problem from two directions. Robotics asked for a bicycle that balances itself without a rider, on reduced models in which the steer is an input [1, 15], and found the balance problem nonminimum phase [6]. Control education took up bicycles adapted for teaching and research [10], and the same magazine surveyed bicycles, motorcycles and their models for the control community [8]. Rider control has been studied with optimal preview, which finds a rider's steering powerful compared with the bicycle's own self-steering, especially at low speed [2]. The stability and controllability of a bicycle that carries a passive rider model have been analyzed [7], and two reviews of 2013 collect the dynamics and the rider models [9, 16]. The rider model of case 8 has passive arms on the bars for a measured reason, given in Section 9.5.4.

HISTORY A Steering Motor Takes the Rider's Place, 1975 to 1986

Attempts to balance single-track vehicles automatically began once predictive motorcycle models had been refined in the 1970s. A thesis of 1975 tried to stabilize a small unmanned motorbike with a controlled pendulum that mimicked a rider's lean. It failed in practice, but it recognized that feedback of the roll angle and its rate was essential. In the early 1980s a robotic bicycle was balanced on a treadmill with steering control and a laterally moving mass. A study of 1986 then balanced a motorcycle with a steering motor alone. Roll angle feedback stabilized the capsize, roll rate feedback the weave and steer rate feedback the wobble, and the gains had to change with speed [17]. The scheduled laws of Section 9.5 make the same change.

9.3 The Bicycle as Modeled

The simulated machine is the benchmark bicycle with a rigid dummy in the rider's place, a steer motor on the head tube, an inertial sensor on the frame, a steer encoder, a wheel-speed encoder, a lateral position sensor referenced to the lane and a tilt switch, as shown in Figure 9.1. Lean φ and steer δ are positive to the right, the heading ψ is positive clockwise seen from above, and (x, y) is the rear contact point with x along the road and y to the rider's right, the SAE J670e convention of the benchmark [1]. Every dimension and mass is the benchmark's, published for exactly this kind of reproduction. The motor, the instruments, the limits and every threshold are teaching values, and the application says so in its model dialog and on its Instructions page. Table 9.1 sorts the signals by role.

9.3.1 Lean and Steer

The Whipple bicycle is four rigid, laterally symmetric bodies hinged together, a rear wheel R, a rear frame B that includes the rider's body, a front frame H and a front wheel F, on knife-edge wheels that roll without slip on level ground. Linearized about upright, straight running at forward speed v, its lean and steer obey Equation (9.1) [1].

$$\mathbf{M}\ddot{\mathbf{q}} + v\,\mathbf{C}_1\dot{\mathbf{q}} + \left(g\,\mathbf{K}_0 + v^2\mathbf{K}_2\right)\mathbf{q} = \begin{bmatrix} T_\varphi \\ T_\delta \end{bmatrix}, \qquad \mathbf{q} = \begin{bmatrix} \varphi \\ \delta \end{bmatrix} \tag{9.1}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}_1$ collects the gyroscopic and inertial terms that grow with v, $\mathbf{K}_0$ the gravity terms and $\mathbf{K}_2$ the terms that grow with v^2. Every entry is a formula in the design parameters of Table 9.2, which the simulator evaluates at every stage of every step (memoized, since the parameters change only when a student changes them). The bicycle's totals about the rear contact come first, as

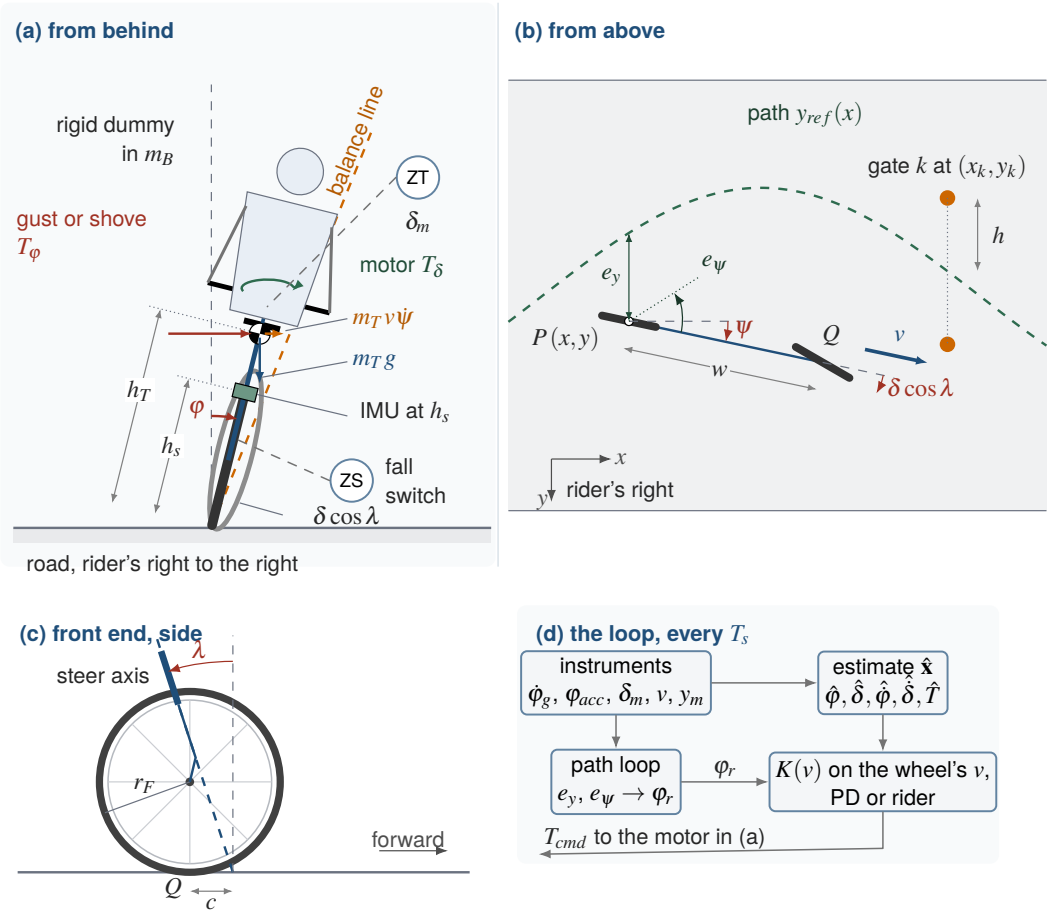


Figure 9.1 The riderless benchmark bicycle as the Bike view draws it, and the loop that holds it, not to scale. (a) From behind, the frame leans φ about the contact line with the dummy in the rear body, and the center of mass at h_T carries the weight $m_T g$ and the centripetal force $m_T v\dot{\psi}$, whose resultant sets the balance line. A gust F at h_T or a shove at the saddle is the lean torque T_φ, the steer motor applies T_δ through the lag τ_a within $\pm T_{max}$, the front wheel turns out at the ground steer angle $\delta\cos\lambda$, the IMU at h_s carries the gyro and the accelerometer, the encoder ZT reads δ_m, and the tilt switch ZS cuts the motor at φ_{fall}. (b) From above, the rear contact P at (x,y) heads at ψ with speed v, the front contact Q is the wheelbase w ahead, and the path loop sees the lateral error e_y and the heading error e_ψ against a path through gates of half-width h. (c) The front end from the side, with the trail c, the steer axis tilt λ and the wheel radius r_F, as the Map view draws it. (d) Every T_s the controller estimates the state from the instruments and applies a law scheduled on the wheel's speed.

Table 9.1 Signals of the Bicycle by Role

Symbol	Role	Nominal	Range
φ	CV and state, lean, right positive	per case	FALLEN at $\pm 40°$
φ_r	SP, lean setpoint or the path loop's output	0	$\pm 15°$
δ	state, steer, right positive	0	stop at $\pm 45°$
$\dot{\varphi}$, $\dot{\delta}$	states, lean and steer rates	0	—
T_δ	state, steer motor torque	0	± 10 N m
T_{cmd}	MV, torque command	per law	$\pm T_{max}$
v	state, forward speed	per case	SP 0 to 10 m/s
ψ, x, y	states, heading and rear contact	0	—
$\dot{\varphi}_g$	measured, rate gyro	$\dot{\varphi}$	noise 0.1 °/s
φ_{acc}	measured, accelerometer tilt	φ	noise 0.5°
δ_m	measured, steer encoder	δ	2^{14} counts a turn
y_m, ψ_m	measured, position and heading	y, ψ	0.02 m, 0.2°
T_φ	DV, gust, side wind, shove, cross slope	0	per event
J_p	DV, pothole, a steer impulse	—	0.5 N m s
m_p	DV, passenger on the rear rack	0	0 to 40 kg
b_g	DV, gyro bias	0	± 5 °/s

shown in Equation (9.2).

$$\begin{aligned}
m_T &= m_R + m_B + m_H + m_F \\
x_T &= (x_B m_B + x_H m_H + w m_F)/m_T \\
z_T &= (-r_R m_R + z_B m_B + z_H m_H - r_F m_F)/m_T \\
I_{Txx} &= I_{Rxx} + I_{Bxx} + I_{Hxx} + I_{Fxx} \\
&\quad + m_R r_R^2 + m_B z_B^2 + m_H z_H^2 + m_F r_F^2 \\
I_{Txz} &= I_{Bxz} + I_{Hxz} - m_B x_B z_B - m_H x_H z_H + m_F w r_F \\
I_{Tzz} &= I_{Rxx} + I_{Bzz} + I_{Hzz} + I_{Fxx} \\
&\quad + m_B x_B^2 + m_H x_H^2 + m_F w^2
\end{aligned} \tag{9.2}$$

where the wheels are axisymmetric, so $I_{Rzz} = I_{Rxx}$ and $I_{Fzz} = I_{Fxx}$, and heights z are negative because the z axis points down. The front assembly A, the front frame with its wheel, has its own totals about its center of mass, as shown in Equation (9.3).

$$\begin{aligned}
m_A &= m_H + m_F, \quad x_A = (x_H m_H + w m_F)/m_A \\
z_A &= (z_H m_H - r_F m_F)/m_A \\
I_{Axx} &= I_{Hxx} + I_{Fxx} + m_H (z_H - z_A)^2 + m_F (r_F + z_A)^2 \\
I_{Axz} &= I_{Hxz} - m_H (x_H - x_A)(z_H - z_A) \\
&\quad + m_F (w - x_A)(r_F + z_A) \\
I_{Azz} &= I_{Hzz} + I_{Fxx} + m_H (x_H - x_A)^2 + m_F (w - x_A)^2 \\
u_A &= (x_A - w - c)\cos\lambda - z_A \sin\lambda \\
I_{A\lambda\lambda} &= m_A u_A^2 + I_{Axx} \sin^2\lambda \\
&\quad + 2 I_{Axz} \sin\lambda \cos\lambda + I_{Azz} \cos^2\lambda \\
I_{A\lambda x} &= -m_A u_A z_A + I_{Axx} \sin\lambda + I_{Axz} \cos\lambda \\
I_{A\lambda z} &= m_A u_A x_A + I_{Axz} \sin\lambda + I_{Azz} \cos\lambda
\end{aligned} \tag{9.3}$$

where u_A is the distance of the assembly's center of mass ahead of the steer axis, $I_{A\lambda\lambda}$ its moment of inertia about that axis, and $I_{A\lambda x}$ and $I_{A\lambda z}$ products about crossed axes. With the trail ratio $\mu =$

Table 9.2 Model Parameters. B Is the Benchmark's Published Value, A a Teaching Value Adjustable in the Model Dialog, H a Hidden Teaching Value

Symbol	Meaning	Value	Range	
w	wheelbase	1.02 m	0.6–1.6	B
c	trail	0.08 m	−0.1–0.2	B
λ	steer axis tilt	$\pi/10$ (18°)	0–40°	B
g	gravity	9.81 m/s^2	1.6–24.8	B
r_R, m_R	rear wheel	0.3, 2		B
I_{Rxx}, I_{Ryy}	its inertias	0.0603, 0.12		B
x_B, z_B, m_B	rear frame, dummy	0.3, −0.9, 85		B
I_{Bxx}, I_{Bxz}, I_{Bzz}	its inertias	9.2, 2.4, 2.8		B
x_H, z_H, m_H	front frame	0.9, −0.7, 4		B
I_{Hxx}, I_{Hxz}, I_{Hzz}	its inertias	0.05892, −0.00756, 0.00708		B
r_F, m_F	front wheel	0.35, 3		B
I_{Fxx}, I_{Fyy}	its inertias	0.1405, 0.28		B
T_{max}	steer motor limit	10 N m	1–30	A
τ_a	steer motor lag	20 ms	2–200	A
τ_v	drive speed loop	0.5 s	0.1–5	A
σ_g	gyro noise	0.1 °/s	0–3	A
σ_a	accelerometer noise	0.5°	0–5	A
h_s	sensor height	0.6 m	0.1–1.2	A
N_δ	encoder resolution	14 bit	8–18	A
σ_y	position noise	0.02 m	0–0.5	A
φ_{fall}, δ_{lock}	fall angle, stop	40°, 45°	fixed	H

Lengths in m, masses in kg and inertias in kg m^2. Every benchmark value is adjustable in the model dialog. The frames' inertias about the lateral axis, $I_{Byy} = 11\,\mathrm{kg\,m^2}$ and $I_{Hyy} = 0.06\,\mathrm{kg\,m^2}$, complete the benchmark's 25 parameters but do not enter Equations (9.2) to (9.4), so the simulator omits them.

$(c/w)\cos\lambda$, the gyrostatic coefficients $S_R = I_{Ryy}/r_R$, $S_F = I_{Fyy}/r_F$ and $S_T = S_R + S_F$, and the static moment $S_A = m_A u_A + \mu m_T x_T$, the entries of the matrices are Equation (9.4).

$$
\begin{aligned}
M_{\varphi\varphi} &= I_{Txx}, \quad M_{\varphi\delta} = M_{\delta\varphi} = I_{A\lambda x} + \mu I_{Txz} \\
M_{\delta\delta} &= I_{A\lambda\lambda} + 2\mu I_{A\lambda z} + \mu^2 I_{Tzz} \\
K_{0\varphi\varphi} &= m_T z_T, \quad K_{0\varphi\delta} = K_{0\delta\varphi} = -S_A \\
K_{0\delta\delta} &= -S_A \sin\lambda \\
K_{2\varphi\delta} &= (S_T - m_T z_T)\cos\lambda/w, \quad K_{2\varphi\varphi} = K_{2\delta\varphi} = 0 \\
K_{2\delta\delta} &= (S_A + S_F \sin\lambda)\cos\lambda/w \\
C_{1\varphi\delta} &= \mu S_T + S_F \cos\lambda + I_{Txz}\cos\lambda/w - \mu m_T z_T \\
C_{1\delta\varphi} &= -(\mu S_T + S_F \cos\lambda), \quad C_{1\varphi\varphi} = 0 \\
C_{1\delta\delta} &= I_{A\lambda z}\cos\lambda/w + \mu\left(S_A + I_{Tzz}\cos\lambda/w\right)
\end{aligned}
\tag{9.4}
$$

These are the formulas of the benchmark's appendix [1]. For the benchmark bicycle they give $m_T = 94\,\mathrm{kg}$, $x_T = 0.342\,\mathrm{m}$, $z_T = -0.861\,\mathrm{m}$, $\mu = 0.0746$ and $S_A = 2.600\,\mathrm{kg\,m}$. A passenger of mass m_p on the rear rack, at $x = -0.05\,\mathrm{m}$ and $z = -1.0\,\mathrm{m}$, is folded into the rear frame by the parallel-axis theorem before the derivation, so it changes every matrix the physics integrates, while the controller keeps the matrices of the bicycle it was told about.

Table 9.3 Eigenvalues of the Benchmark Bicycle From the Simulator's Own Derivation (1/s), and the Largest Difference From the Published Table in Each Row

v (m/s)	Re weave	Im weave	capsize	castering	max $\lvert\Delta\rvert$
0	real: 5.530944, 3.131643	−3.131643	−5.530944	5.3×10^{-15}	
1	3.526962	0.807740	−3.134231	−7.110080	1.4×10^{-14}
2	2.682345	1.680663	−3.071586	−8.673880	1.4×10^{-14}
3	1.706756	2.315824	−2.633661	−10.351015	2.5×10^{-14}
4	0.413253	3.079108	−1.429444	−12.158614	4.3×10^{-14}
5	−0.775342	4.464868	−0.322866	−14.078390	1.6×10^{-14}
6	−1.526445	5.876731	−0.004067	−16.085371	1.1×10^{-14}
7	−2.138756	7.195259	0.102682	−18.157885	6.1×10^{-13}
8	−2.693487	8.460380	0.143279	−20.279409	5.3×10^{-14}
9	−3.216754	9.693774	0.157902	−22.437886	8.3×10^{-14}
10	−3.720168	10.906811	0.161053	−24.624596	7.5×10^{-14}

9.3.2 The Benchmark Reproduced

With the parameters of Table 9.2, the sixteen derived entries equal the matrices printed in the benchmark [1], Equation (9.5), to 1.4×10^{-14}, the largest difference being in the entry -80.95.

$$\begin{aligned}
\mathbf{M} &= \begin{bmatrix} 80.81722 & 2.31941332208709 \\ 2.31941332208709 & 0.29784188199686 \end{bmatrix} \\
\mathbf{K}_0 &= \begin{bmatrix} -80.95 & -2.59951685249872 \\ -2.59951685249872 & -0.80329488458618 \end{bmatrix} \\
\mathbf{K}_2 &= \begin{bmatrix} 0 & 76.59734589573222 \\ 0 & 2.65431523794604 \end{bmatrix} \\
\mathbf{C}_1 &= \begin{bmatrix} 0 & 33.86641391492494 \\ -0.85035641456978 & 1.68540397397560 \end{bmatrix}
\end{aligned} \tag{9.5}$$

The eigenvalues are taken from the state matrix $\mathbf{A}(v) = [\mathbf{0},\ \mathbf{I};\ -\mathbf{M}^{-1}(g\mathbf{K}_0 + v^2\mathbf{K}_2),\ -v\mathbf{M}^{-1}\mathbf{C}_1]$ by the balanced Hessenberg QR routine of the shared framework (Section 1.6), `shared/src/framework/core/eig.js`. Against the 44 numbers of the benchmark's eigenvalue table from 0 to 10 m/s, the real and imaginary parts of the weave, the capsize and castering roots and the roots at rest, they agree to 6.2×10^{-13}, and 43 of them to 8.4×10^{-14}, as listed in Table 9.3. The exception is the castering root at 7 m/s, -18.15788466125201 against the printed -18.15788466125262. Newton's method on the characteristic quartic of the derived matrices returns the simulator's value to the last bit, so the difference lies in the thirteenth decimal of the printed value rather than in the derivation. Bisection on the largest real part finds the weave speed $v_w = 4.292\,382\,536\,341\,10\,\mathrm{m/s}$ and the capsize speed $v_c = 6.024\,262\,015\,388\,36\,\mathrm{m/s}$, against the printed 4.29238253634111 and 6.02426201538837, differences of 5.3×10^{-15} and 8.9×10^{-15}. The weave frequency at v_w, $3.435\,033\,848\,661\,44\,\mathrm{s^{-1}}$, and the speed $v_d = 0.684\,283\,078\,892\,\mathrm{m/s}$ at which the weave pair forms both agree to 3.2×10^{-15}. The published values were transcribed from the benchmark's preprint (final revision of 9 June 2007), because the publisher's site refused automated requests.

9.3.3 Path, Drive and Steer Motor

The rear contact moves along its heading, and the heading turns with the steer by the benchmark's kinematics [1], as shown in Equation (9.6).

$$\dot{\psi} = \frac{(v\delta + c\dot{\delta})\cos\lambda}{w}, \qquad \dot{x} = v\cos\psi, \qquad \dot{y} = v\sin\psi \tag{9.6}$$

A first-order speed loop on the rear wheel, limited to the drive's acceleration, holds the forward speed, and the steer motor's torque lags its command, as shown in Equation (9.7).

$$\begin{aligned} \dot{v} &= \mathrm{sat}_{\pm a_{max}}\left(\frac{v_{SP}-v}{\tau_v}\right) \\ \tau_a \dot{T}_\delta &= T_{cmd} - T_\delta, \qquad |T_{cmd}| \le T_{max} \end{aligned} \tag{9.7}$$

where $\tau_v = 0.5\,\mathrm{s}$, $a_{max} = 1.5\,\mathrm{m/s^2}$ by default, $\tau_a = 20\,\mathrm{ms}$ and $T_{max} = 10\,\mathrm{N\,m}$, and the speed is held at or above zero. The benchmark is linearized at constant speed, and here the matrices of Equation (9.1) simply follow v as it changes, a quasi-steady simplification that the benchmark does not cover.

9.3.4 Fall, Stop and Disturbances

Three hard events act between integration steps. FALLEN latches where $|\varphi|$ crosses $\varphi_{fall} = 40^\circ$, the instant located by linear interpolation inside the step, and the tilt switch then cuts the motor while the bicycle topples about its contact line as one rigid body, $\ddot{\varphi} = m_T g h_T \sin\varphi / I_{Txx}$ with $h_T = -z_T$, until the bars touch the ground at 72°, where the run stops. The handlebar stop at $\delta_{lock} = \pm 45^\circ$ is an inelastic, one-sided constraint. On arrival the steer rate toward the stop is removed by an impulse J on the steer coordinate, which reaches the lean through the mass matrix as $\Delta\dot{\mathbf{q}} = \mathbf{M}^{-1}[0,\ J]^{\mathsf{T}}$, and while the steer is held only the lean row of Equation (9.1) is integrated, until the free steer acceleration points away from the stop. A pothole is the same kind of impulse, $J_p = 0.5\,\mathrm{N\,m\,s}$ on the steer. The lean torque is Equation (9.8).

$$T_\varphi = F h_T + m_T g h_T \arctan(s_r/100) + T_{sh} \tag{9.8}$$

where F is a side wind of up to $\pm 100\,\mathrm{N}$ plus any gust (70 N for 1.5 s) at the center of mass, s_r is a road cross slope in percent, and T_{sh} is a shove of 50 N at the saddle, 0.9 m up, for 0.4 s. Every one of these sizes is a teaching value.

9.3.5 Instruments

The controller reads the bicycle only through a measurement packet, assembled once every control period $T_s = 10\,\mathrm{ms}$ by one function, as shown in Equation (9.9).

$$\begin{aligned} \dot{\varphi}_g &= \dot{\varphi} + b_g + \sigma_g n_1, \qquad \varphi_{acc} = \varphi - \frac{v\dot{\psi} + h_s\ddot{\varphi}}{g} + \sigma_a n_2 \\ \delta_m &= q\,\mathrm{rnd}(\delta/q), \qquad q = 2\pi/2^{N_\delta} \\ y_m &= y + \sigma_y n_3, \qquad \psi_m = \psi + \sigma_\psi n_4 \end{aligned} \tag{9.9}$$

where n_1 to n_4 are unit normal deviates from the engine's generator seeded at 20261019, drawn on every call whatever the noise levels so that one changed noise does not reshuffle the others, $\ddot{\varphi}$ is the true lean rate differenced over the period, $\sigma_\psi = 0.2^\circ$, and the remaining values are in Table 9.2. An accelerometer cannot tell gravity from the bicycle's own sideways acceleration, so in a balanced turn φ_{acc} reads almost no lean. The wheel speed and the distance along the course are exact, and the tilt switch reports the fall.

9.3.6 Numerics and Initial States

Table 9.2 lists the parameters. The nine states, the lean and steer with their rates, the motor torque, the speed, the heading and the rear contact, are integrated by the engine of Section 1.6.1 in steps of 1 ms, against a 20 ms motor lag and a castering root of $-24.6\,\mathrm{s^{-1}}$ at 10 m/s. The controller runs every 10 ms and the historian logs at 50 Hz, which gives 29 samples per period of the fastest motion the trends show, the weave at 1.7 Hz at 10 m/s. A case starts upright and straight at its speed setpoint, or with the 2° lean its scenario names, the motor torque at zero. The noise is seeded, so every run replays exactly, and Reset and the reload button behave as Section 1.6.1 describes. The application passes its

49 verification checks. Among them the engine's integrated motion equals the matrix exponential of $\mathbf{A}(v)$ to 10^{-8} (relative) over 1 s at 1, 3, 5 and 8 m/s, and halving the step moves the countersteer peak of case 6 by 6×10^{-9} degree and the fall of case 1 by 0.4 ms. On the machine used here the model runs about 348 000 steps a second outside the browser, against the 5 000 a second that the top speed of 5× real time needs. At that speed the 40 s slalom takes 8 s to watch.

9.3.7 Assumptions

The Whipple bicycle. Rigid parts, ideal hinges and knife-edge wheels that roll without slip, with no tire compliance or relaxation, no frame flex, no friction in the steering head and no rider moving on the frame. Wobble, the fast steering oscillation of real machines, appears to need tire and frame compliance and is absent [1].

Linear about upright running. The lean is trusted for small angles, and FALLEN at 40° is a device for the picture. An optional nonlinear mode puts $\sin\varphi$ in the gravity term of the lean equation and nothing else, and with it on, above the capsize speed, the bicycle settles into a steady turn at 15 to 31° of lean from 6.5 to 10 m/s instead of falling, because the turn's terms stay linear. The full nonlinear benchmark has families of steady hands-free circles of its own [14], and the mode is not a model of them.

A rigid dummy. The rider's 85 kg belong to the rear frame and the dummy's hands rest on the bars without torque. Case 8's rider is a control law.

A stand-in for the road. A cross slope puts no lean torque on knife-edge wheels to first order, because the contact stays under the wheel plane, so the term in Equation (9.8) is a teaching stand-in for what real tires do. The drive has no pitch dynamics, rolling resistance or air drag.

Teaching limits. The 10 N m motor, the 20 ms lag, the 45° stop and the 40° fall angle are teaching values. A real steer motor's limits come from its data sheet and gearing, a real bicycle's steering lock from where its bars or fork crown meet the frame, and a self-balancing test bicycle cuts its motor at a tilt its builder chooses.

9.4 A Window in Speed

9.4.1 Four Modes Against Speed

At rest the bicycle is two inverted pendulums, with real roots at ± 3.13 and $\pm 5.53\,\mathrm{s}^{-1}$, as shown in Figure 9.2a. As the speed rises, the two positive roots meet at $v_d = 0.684\,\mathrm{m/s}$ and form the weave, an oscillation of lean and steer, while the negative ones become the capsize, a slow real mode dominated by lean, and the castering, a fast real mode dominated by steer [1]. The weave grows below $v_w = 4.292\,\mathrm{m/s}$ and the capsize above $v_c = 6.024\,\mathrm{m/s}$. Between the two every mode decays, the slowest at $0.32\,\mathrm{s}^{-1}$ at 5 m/s, and the bicycle steers itself into its own fall hard enough to catch it and not so hard that it holds itself in a turn. Hands off from the 2° push, it falls in 1.46 s at 1 m/s, where the weave grows at $3.53\,\mathrm{s}^{-1}$ with a period of 7.8 s, and in 2.00 s at 3 m/s, where the weave grows at $1.71\,\mathrm{s}^{-1}$ with a period of 2.7 s and the lean swings once to the other side before the fall. At 8 m/s the capsize grows at $0.143\,\mathrm{s}^{-1}$, an e-folding time of 7.0 s, and the lean reaches 40° after 20.2 s.

Pause. Try Exercise 9.1 before reading on.

> **TRY IT IN THE APP** **2 · The Window: Hands Off at 3, 5 and 8 m/s**
>
> Load the scenario and press Run with your hands off, and watch the lean. At 3 m/s it swings once to the other side and the bicycle is down at 2.00 s. Set Speed setpoint to 5 m/s, press Reset and run again: the bicycle rides upright. At 8 m/s it leans slowly and passes 40° after 20.2 s.

9.4.2 Where the Steer Gets Its Grip

The steer acts on the lean through the off-diagonal stiffness. Holding the steer at an angle δ puts a static torque $(gK_{0\varphi\delta} + v^2K_{2\varphi\delta})\delta$ on the lean, −25.5 N m per radian from gravity and a term growing

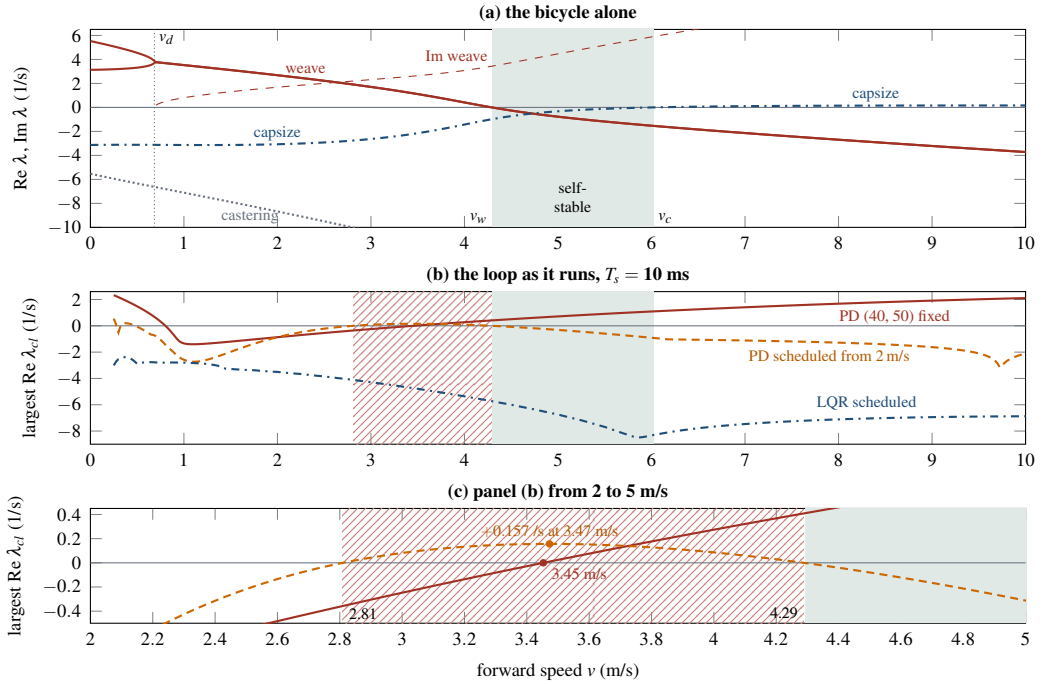


Figure 9.2 The eigenvalue map, from the simulator's own functions. (a) Real parts of the four eigenvalues of the benchmark bicycle against speed, with the weave's imaginary part dashed and the self-stable window from 4.292 to 6.024 m/s shaded. The weave's real parts are solid, the capsize dash-dotted and the castering dotted. (b) The largest real part of the closed loop as it runs, the bicycle held over each 10 ms period, for case 5's lean PD tuned at 2 m/s, fixed (solid) and scheduled (dashed), and for the scheduled LQR (dash-dotted). (c) The same from 2 to 5 m/s. The fixed PD loses stability at 3.45 m/s, and the schedule leaves a band from 2.81 to 4.29 m/s (hatched) in which the closed loop grows at up to 0.157 s^{-1}.

with the square of the speed, so that the grip is 1889 N m/rad at 5 m/s, 281 at 2 m/s, 51 at 1 m/s, zero at 0.58 m/s and reversed below. At 0.3 m/s it is −18.6 N m/rad, against −794 N m/rad of gravity on the lean, $gm_T z_T$. A lean loop's gain must therefore grow as the speed falls, and at a walking pace no gain the motor can deliver is enough. Case 9 is built on this fact. Below the zero the steer's gain on the lean has changed sign, as the gain of a braking tire does past its friction peak in Chapter 4.

🕮 **Pause.** Try Exercise 9.3 before reading on.

9.4.3 A Response That Starts the Wrong Way

Steering to the right turns the front contact to the right under a bicycle whose center of mass carries on, so the frame begins to lean to the left, and only a bicycle leaning left can turn left without falling over. From a lean or heading command to the path, the response therefore starts the wrong way, the signature of a right-half-plane zero [5, 6] (Section 1.4.7). In the linear model at 5 m/s the regulator of Section 9.5, asked for a 5° lean to the left, steers 4.0° to the right before 4.0° to the left, and the rear wheel's track moves 5.9 cm to the right before the path bends left. Any loop wrapped around the lean loop, such as the path loop of case 7, has to be slower than the lean loop for this reason. The cart loop of Chapter 7 and the position loop of Volume 3, Chapter 10, which command an angle through a wrong-way response, obey the same rule.

🕮 **Pause.** Try Exercise 9.2 before reading on.

Table 9.4 Controller Settings, Defaults and Ranges

Symbol	Setting	Default	Range
—	law	LQR	PD, rider
ε	PD sign	into the fall	wrong way
k_φ, $k_\dot{\varphi}$	PD gains (N m/rad, N m s/rad)	40, 50	0–200, 0–120
—	schedule on speed	on	off
v_d	design speed	5 m/s	0.5–10
$\bar{\varphi}$, $\bar{\delta}$, $\bar{T}$	LQR scales	2°, 5°, 5 N m	0.5–10, 1–30, 0.5–20
τ_{cf}	lean filter time constant	0.5 s	0.05–5
—	lean estimate	compensated	raw, perfect
k_r, T_d	rider reflex (N m/rad), anticipation	10, 0.4 s	0–100, 0–2
k_a, c_a	rider's arms (N m/rad, N m s/rad)	10, 0.5	0–60, 0–10
τ_r	rider's reaction	150 ms	0–400
ω_p, ζ_p	path loop	0.6 rad/s, 0.7	0.2–3, 0.3–2
t_p	curvature preview	0.4 s	0–2
v_{SP}, a_{max}	speed setpoint, drive limit	5 m/s, 1.5 m/s^2	0–10, 0.2–4

9.5 The Controller as Coded

Every 10 ms the controller reads the packet of Equation (9.9), updates its estimate, computes the lean reference, runs one law and writes a torque command clamped to $\pm T_{max}$ and held until the next period. Its own model is the declared bicycle, the parameters as set with no passenger, and no law reads the simulator's truth, which the export logs in columns prefixed `truth_`. Table 9.4 lists the settings.

9.5.1 Estimate

The laws act on $\hat{\mathbf{x}} = [\hat{\varphi}, \hat{\delta}, \hat{\dot{\varphi}}, \hat{\dot{\delta}}, \hat{T}]$. The lean is a complementary filter that integrates the gyro and pulls toward the accelerometer, corrected for the turn the encoder and the wheel speed imply and for the sensor's own swing, as shown in Equation (9.10).

$$\begin{aligned} \hat{\varphi}_k &= a\big(\hat{\varphi}_{k-1} + \dot{\varphi}_{g,k} T_s\big) + (1-a)\Big(\varphi_{acc,k} + \frac{v\dot{\psi}_{kin} + h_s\hat{\dot{\varphi}}}{g}\Big) \\ a &= \frac{\tau_{cf}}{\tau_{cf} + T_s}, \qquad \dot{\psi}_{kin} = \frac{(v\hat{\delta} + c\hat{\dot{\delta}})\cos\lambda}{w} \end{aligned} \tag{9.10}$$

where $\hat{\dot{\varphi}}$ is the gyro differenced and passed through a 20 ms first-order filter. The lean rate is the gyro as read, bias included, the steer is the encoder, the steer rate is the encoder differenced through the same 20 ms filter, and $\hat{T}$ is a replica of the motor's lag driven by the commands sent. Two options replace the filter, the raw accelerometer without the correction and the true state, which is labeled a teaching comparison.

9.5.2 The Lean PD and Its Schedule

The lean PD is Equation (9.11).

$$T_{cmd} = \varepsilon\big(k_\varphi(\hat{\varphi} - \varphi_r) + k_\dot{\varphi}\hat{\dot{\varphi}}\big) \tag{9.11}$$

with $\varepsilon = +1$ steering into the fall, a positive torque for a lean to the right, and $\varepsilon = -1$ the wrong way. On the lag-free bicycle the closed loop's characteristic polynomial is affine in the two gains, $p(s) = \det\big(\mathbf{M}s^2 + v\mathbf{C}_1 s + g\mathbf{K}_0 + v^2\mathbf{K}_2 - \mathbf{e}_2[k_\varphi + k_\dot{\varphi}s,\ 0]\big)$ with $\mathbf{e}_2 = [0,1]^\mathsf{T}$, and two gains cannot place its four roots. A schedule in the sense of Section 1.4.3, which Chapter 3 applies on angle, keeps the closed-loop polynomial of the design point, and two gains can only come close to it. The schedule therefore chooses, at the measured speed, the pair whose coefficients of s^3 to s^0 come closest in

weighted least squares to those the tuned gains give at the design speed, as shown in Equation (9.12).

$$\min_{k_\varphi, k_{\dot\varphi}} \sum_{i=0}^{3} \left(\frac{a_i(v; k_\varphi, k_{\dot\varphi}) - a_i(v_d)}{a_i(v_d)} \right)^2 \tag{9.12}$$

From case 5's design, 40 N m/rad and 50 N m s/rad at 2 m/s, it gives 7.6 and 6.2 at 5 m/s and 4.2 and −2.9 at 8 m/s. Near 0.58 m/s, where the steer's grip passes through zero, the gains run away.

9.5.3 The Full-State Regulator

The regulator works on the declared bicycle with the motor's lag as a fifth state, $\dot{\mathbf{x}} = \mathbf{A}_5(v)\mathbf{x} + \mathbf{b}_5 T_{cmd}$ with $\mathbf{x} = [\varphi, \delta, \dot\varphi, \dot\delta, T_\delta]$, discretized for a zero-order hold at T_s by the framework's Riccati module, `shared/src/framework/core/riccati.js`, with Bryson's weights (Section 1.4.5) $\mathbf{Q} = \mathrm{diag}(1/\bar\varphi^2, 1/\bar\delta^2, 0, 0, 0)$ and $R = 1/\bar T^2$ on the lean, steer and torque scales $\bar\varphi$, $\bar\delta$ and $\bar T$. The law regulates about the steady turn that a lean reference asks for, as shown in Equation (9.13).

$$\begin{aligned} T_{cmd} &= T_r - \mathbf{K}(v)(\hat{\mathbf{x}} - \mathbf{x}_r), \qquad \mathbf{x}_r = [\varphi_r,\ \delta_r,\ 0,\ 0,\ T_r] \\ \delta_r &= -K_{\varphi\varphi}\varphi_r / K_{\varphi\delta}, \qquad T_r = K_{\delta\varphi}\varphi_r + K_{\delta\delta}\delta_r \end{aligned} \tag{9.13}$$

where K_{ij} are entries of $g\mathbf{K}_0 + v^2\mathbf{K}_2$ and δ_r is clamped to $\pm 20^\circ$ near the speed where $K_{\varphi\delta}$ vanishes. Fixed, $\mathbf{K}$ is designed at v_d, and at 5 m/s it is $[-159.1,\ 69.5,\ -26.4,\ 4.6,\ 0.39]$ in N m/rad for the angles, N m s/rad for the rates and N m per N m for the torque. Scheduled, a table of designs every 0.25 m/s from 0.25 to 12 m/s is computed once per model and weighting and interpolated linearly at the measured speed. At 1 m/s the scheduled gain on the lean rate is 203 N m s/rad, so the gyro's 0.1 °/s of noise reaches the command as 0.35 N m RMS through that gain alone, and as 0.44 N m RMS in all.

9.5.4 The Rider Model

The rider's balance reflex acts on the lean error and its rate $\tau_r = 150$ ms late, while the arms' own stiffness and damping on the bars act at once, as shown in Equation (9.14).

$$T_{cmd}(t) = k_r\left[(\hat\varphi - \varphi_r) + T_d\hat{\dot\varphi}\right]_{t-\tau_r} - k_a\hat\delta - c_a\hat{\dot\delta} \tag{9.14}$$

The delay is a buffer of the controller's own samples. A delayed reflex alone cannot ride this bicycle. With the arms' stiffness k_a set to zero, none of 36 pairs of (k_r, T_d) tried stays up even at 4 m/s, while with the arms the model rides at 3 and 4 m/s and falls at 2 and 1.5 m/s. Delayed reflexes and the passive impedance of the arms are both elements of the rider models the literature reviews [9, 16].

9.5.5 Path Loop

The path loop never touches the bars. A car follows a path by steering it directly, as the kinematic bicycle of Volume 20, Chapter 2 does, while a bicycle can only follow a path by leaning. The path loop turns the lateral error $e_y = y_{ref} - y_m$ and the heading error $e_\psi = \psi_{ref} - \psi_m$ into a lean reference, with the course's curvature κ read t_p ahead, where the bicycle will be once it has leaned, as shown in Equation (9.15).

$$\varphi_r = \frac{\omega_p^2 e_y + 2\zeta_p\omega_p v e_\psi}{g} + \arctan\frac{v^2\kappa(x + vt_p)}{g} \tag{9.15}$$

limited to $|\varphi_r| \le 15^\circ$, since a balanced bicycle accelerates sideways at about $g\varphi$. The course is a cosine through eight gates 15 m apart, alternately 1 m to the right and to the left, each 0.75 m wide on either side of its center, faded in and out by a quintic envelope so that its heading and curvature are continuous.

Table 9.5 The Nine Cases and the Challenge

#	Configuration and events	Lesson
1	manual at 1, 3 and 5 m/s, released at 2°	speed decides balance
2	hands off at 3, 5 and 8 m/s from 2°	the self-stable window
3	trail 0.12, 0, −0.02 m, tilt 10 and 5°, at 5 m/s	geometry moves the window
4	PD (40, 50) at 2 m/s, wrong sign	steer into the fall
5	same PD, speed 2 to 8 m/s at 0.5 m/s^2 from 3 s	one tuning, a schedule
6	scheduled LQR at 5 m/s, lean SP −5° at 2 s	countersteering
7	scheduled LQR and path loop, eight gates at 5 m/s	cascade order
8	rider at 4 m/s, shove at 5 s	reaction time
9	scheduled LQR at 0.3 m/s from 2°	honestly infeasible
C	slalom at 7, 3, 7 m/s, gust at gate 4, pothole at gate 6	everything at once

Table 9.6 Outcomes Read From the Running Engine

Case and condition	Result
1, hands off at 1 / 3 m/s	falls 1.46 / 2.00 s
1, a tap right at 5 m/s	leans 11.1° left, turns 38° left in 5 s
2, hands off at 5 / 8 m/s	under 0.5° by 5.5 s / falls 20.2 s
3, trail 0.12 / 0 / −0.02 m	window 4.70–7.11 / 2.93–3.13 m/s / none
3, tilt 10 / 5°; 20 kg passenger	4.68–5.45 m/s / none; 4.47–5.89 m/s
4, wrong sign / hands off / right sign	falls 0.61 / 1.66 s / within 0.5° from 0.93 s
5, fixed / scheduled	saturated 13.9 of 30 s / never, ≤1.58 N m
6, first steer; settled turn	4.3° right by 2.13 s; 2.1° left, 29 m
7, ω_p 0.6 / 2.5 rad/s; raw accelerometer	8 gates within 0.09 m / 4 missed; falls 7.19 s
8, rider at 4 / 2 / 1.5 m/s	lean peak 12.0° / falls 6.00 / 1.91 s
9, any of 123 designs	all fall; shipped LQR at 1.53 s
C, shipped / suggested	2 gates, fall 17.0 s / 8 gates, 0.10 % saturated

9.5.6 Manual Mode, Transfer and the Loop as It Runs

In manual mode the station's torque, a held arrow key (2 N m, ramped over 0.05 s) and a tap on either side of the drawn bicycle (2 N m for 0.2 s) push the bars. Every law keeps running, so on the switch to automatic the law's output starts from the hands' torque and the difference decays with a 0.1 s time constant. The saturation flag compares the law's request with $\pm T_{max}$ and is cleared every period. The tilt switch cuts the motor in either mode.

The Map view, the balance card and the export report the closed loop as it runs. The bicycle with the motor's lag is held over each period, the law is applied once a period, and each discrete root z is shown as $\ln z/T_s$. For the rider, a second-order Padé approximant of the delay $\tau_r + T_s/2$ stands in for the buffer. The hold decides stability for the lean PD, whose lean-rate gain drives a closed-loop oscillation of roughly 9 to 25 rad/s between 2 and 8 m/s. Applied continuously, the fixed PD of case 5 would be stable up to 5.59 m/s, and as it runs it is unstable from 3.45 m/s, as shown in Figure 9.2b. Its schedule, which the continuous analysis rates stable on 2 to 8 m/s with the weakest root at –0.062 s^{-1} near 3.6 m/s, lets the closed loop's oscillation grow at up to 0.157 s^{-1} near 3.47 m/s as it runs. The engine is the referee. With the true state, the noise off and the speed held, the lean's envelope under the schedule grows at 0.082, 0.142 and 0.086 s^{-1} at 3, 3.5 and 4 m/s, against 0.082, 0.156 and 0.086 from the sampled analysis and −0.190, −0.066 and −0.095 from the continuous one. The scheduled LQR's largest real part stays below –2.76 s^{-1} from 0.5 to 10 m/s.

9.6 Nine Cases and a Slalom, Measured

The cases are listed in Table 9.5 and their outcomes in Table 9.6. Every number in this section was measured on the engine run headless with the page's own stepping, at the default seed, with any change a sentence names entered before Run, and every time is measured from Run. A case opened first, one chosen from the list after another case and one restarted with Reset give identical runs.

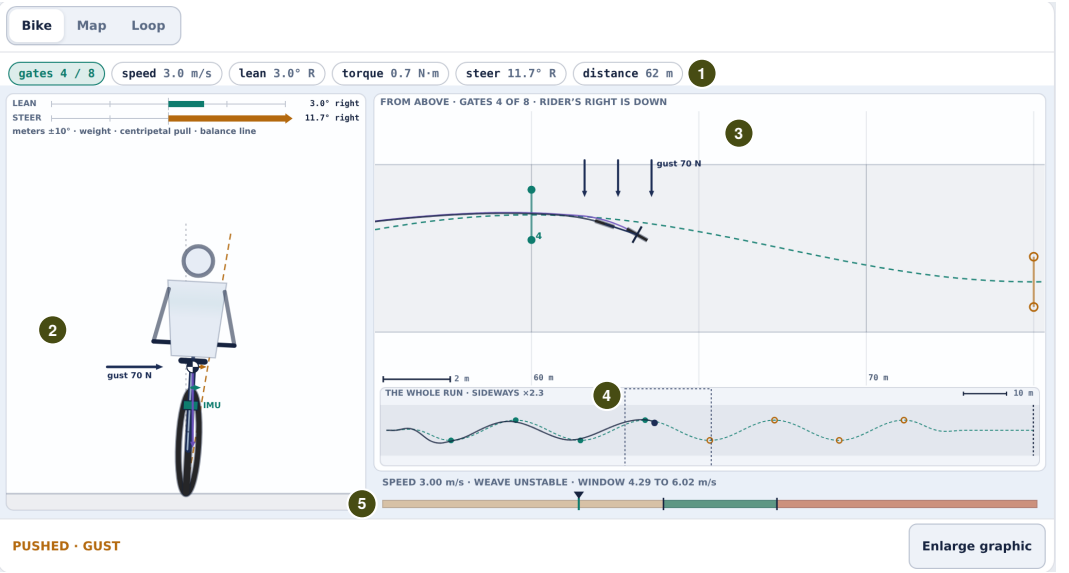


Figure 9.3 The slalom on its speed program with the suggested scheduled LQR, captured at $t = 11.4$ s on the Bike view and cropped to the plant view (the header, the control panel, the trends and the scorecard are cut). (1) The badges read 4 of 8 gates, 3.0 m/s, a lean of 3.0° to the right, a torque of 0.7 N m, a steer of 11.7° to the right and 62 m. (2) In the rear view the 70 N gust at gate 4 pushes from the left at the center of mass, and the meters show the bicycle leaning 3.0° to the right while the regulator steers 11.7° to the right into the fall. (3) From above, the three gust arrows cross the road just past gate 4. (4) The whole run, stretched sideways, with the window of the view above it. (5) The speed bar puts 3.00 m/s in the weave band below the window of 4.29 to 6.02 m/s, and its label reads that the weave is unstable.

9.6.1 By Hand, and Hands Off

Case 1 releases the bicycle at 1 m/s with the student's hands on the bars. A held key or a tap pushes with 2 N m, and pushing into the fall, to the right, brings the wheels back under the bicycle, but its fastest motion doubles every 0.20 s and hands off it is down in 1.46 s. At 3 m/s a player who steers by the lean they see, only 25 ms late, is down within 4.5 s for the best of 49 gain pairs on the keys' 2 N m. At 5 m/s the student can let go, and one tap to the right leans the bicycle 11.1° to the left and turns it 38° to the left in 5 s, a first meeting with case 6. Case 2 runs the same push with nobody steering and measures the window of Section 9.4. Case 3 moves it. The window is 4.70 to 7.11 m/s at 0.12 m of trail and 2.93 to 3.13 m/s at zero trail, so trail is one term among several. At −0.02 m there is no window, and at 5 m/s a real mode growing at 0.65 s^{-1} brings the bicycle down at 4.37 s. A steer axis tilt of 10° gives 4.68 to 5.45 m/s and 5° gives none, and a 20 kg passenger, which the controller does not know about, narrows the window to 4.47 to 5.89 m/s.

DID YOU KNOW? No Trail, Still a Window

Experiments of 1970 pointed to trail as an important factor in a bicycle's stability [1]. From the simulator: with no trail at all, the benchmark bicycle still rides itself between 2.93 and 3.13 m/s, while a steer axis tilted only 5° leaves no window at any speed (Table 9.6).

9.6.2 The Sign, and One Tuning Across Speeds

Case 4 balances at 2 m/s with the lean PD at 40 N m/rad and 50 N m s/rad, shipped with the wrong sign. It steers away from the lean and throws the bicycle down in 0.61 s, against 1.66 s with nobody steering, its largest closed-loop root at +13.7 s^{-1} against the bicycle's own +2.68 s^{-1}. With the right sign the slowest mode decays at 0.86 s^{-1} and the lean is within 0.5° of upright from 0.93 s with at most 2.78 N m.

Case 5 carries the same PD up a ramp from 2 to 8 m/s at 0.5 m/s^2 from 3 s. The torque first saturates at 11.9 s, at 6.44 m/s, and is saturated for 13.9 of the 30 s. From 20 s the bars swing ±7.7° at 2.9 Hz while the lean moves only ±1.8°, a saturated limit cycle (Section 1.4.9) of the closed loop's single oscillatory pair, which the linear analysis puts at 3.91 Hz at 8 m/s, well above the bicycle's own weave at 1.35 Hz. Scheduled, the same design rides the whole ramp with at most 1.58 N m and never saturates. The weak band of Figure 9.2c does not show, because the ramp crosses it in 3.0 s, in which a growth of at most 0.157 s^{-1} enlarges a disturbance by at most a factor of 1.6. In the band the growing pair oscillates at 1.50 Hz at 3.5 m/s with the steer swinging about five and a half times the lean, the weave as the feedback has moved it, against 0.42 Hz for the bicycle alone. Stabilizing gain pairs exist at every speed, but choosing them would be a new design rather than one design carried across speeds. The full-state LQR, which also feeds back the steer, has no such band, and the PD schedule's band is kept as a teaching result.

9.6.3 Countersteering and the Slalom

Case 6 holds the bicycle upright at 5 m/s with the scheduled LQR and asks for a 5° lean to the left at 2 s. The torque goes right first, up to 7.1 N m, the steer swings 4.3° to the right by 2.13 s and turns left only from 2.25 s, the lean reaches −4.5° at 2.42 s, and the turn settles at 2.1° of left steer on a circle of about 29 m. With the true state in the loop the radius is within 0.4 % of $v^2/(g\tan\varphi)$. The rear wheel's track moves a few centimeters to the right first, 2.6 cm in this run and 5.9 cm in the noise-free linear model.

TRY IT IN THE APP 6 · Countersteer: How a Left Turn Starts

Set the clock to 0.25×, press Run and watch the steer angle trend when the left turn is asked at 2 s. The bars swing 4.3° to the right first, by 2.13 s, and turn left only from 2.25 s. Only then does the lean follow, reaching −4.5° at 2.42 s.

Case 7 rides the eight gates at 5 m/s under the path loop and the scheduled LQR. At $\omega_p = 0.6\,\mathrm{rad/s}$ every gate is passed within 0.09 m of its center, the loop asking for at most 8.6° of lean. At 2.5 rad/s it asks for leans up to its 15° limit, the lean loop delivers each one late and after a countersteer, and the bicycle misses 4 of the 8 gates by up to 1.21 m. Linearized with the true state, that cascade has a root at +3.13 s^{-1} at 5 m/s, against −0.54 s^{-1} at the shipped 0.6 rad/s, and only the lean and torque limits hold it to a wide swing. With the raw accelerometer the slalom's turns pull the lean estimate 27° from the truth and the bicycle falls at 7.19 s.

9.6.4 A Rider, and a Walking Pace

Case 8 hands the bars to the rider model at 4 m/s. It catches a shove at the saddle with the lean peaking at 12.0° and at most 2.03 N m on the bars. At 1.5 m/s it falls at 1.91 s, and at 2 m/s at 6.00 s. At 1.5 m/s the bicycle's weave grows at 3.11 s^{-1}, doubling every 0.22 s, and 150 ms of delay is most of that time, the delay limit of Section 1.4.7. The LQR has no delay, but braking to a stop at 0.5 m/s^2

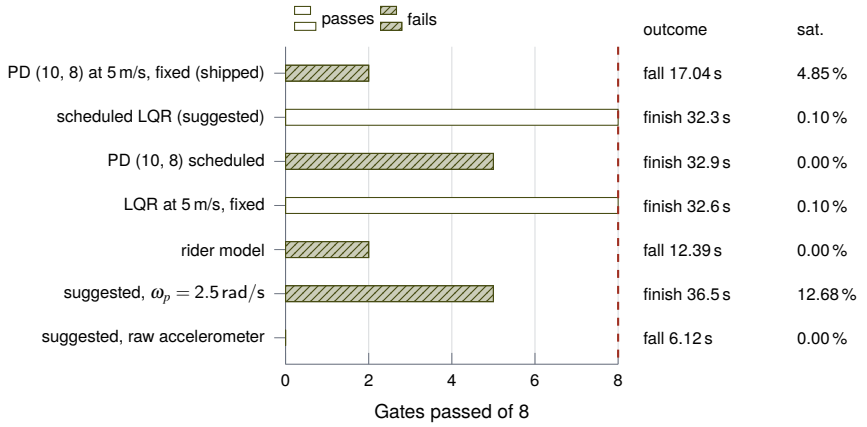


Figure 9.4 The slalom configurations of Table 9.7: gates passed of eight (the target, dashed line), with each run's outcome and the share of the 40 s with the torque command saturated. White bars pass the challenge, hatched bars fail it. Notice that both passing runs carry a regulator, while the scheduled PD finishes the course with only five gates, and the fast path loop finishes with five gates and 12.68 % of the run saturated.

it holds on only until the bicycle has all but stopped. In the last few centimeters per second it saturates the motor and puts the bars on their stop, below 0.17 m/s and at 0.05 m/s in this run, and it falls as it stops. Its gain on the lean grows from 194 N m/rad at 3 m/s to 1650 N m/rad at 0.5 m/s, as the grip of Section 9.4 predicts.

Case 9 asks the scheduled LQR to hold the bicycle at 0.3 m/s after the same 2° push. The design is stable on paper, its closed loop decaying at 2.53 s^{-1}, but from a perfect reading of the lean it asks at once for 123 N m, about twelve times what the motor has. The bars reach their stop at 0.21 s and the bicycle falls at 1.53 s. The application's own sweep from this start runs 123 designs, 80 LQR weightings, 40 PD gain pairs of either sign, the PD schedule, the rider and the LQR given the true state, and every one falls. A smaller push is another matter. From 0.25° the LQR holds the bicycle with the steer swinging to 24.8°, and the 2° push is held at 1 m/s with 20.9° of steer and at 2 m/s with 7.8°. At a walking pace riders put a foot down.

9.6.5 The Slalom on a Speed Program

The challenge rides the course of case 7 with the speed setpoint at 7 m/s, 3 m/s from 40 m and 7 m/s again from 100 m, through a 70 N gust at gate 4 and the pothole at gate 6. Its target is no fall, all eight gates, the finish line at 150 m, and the torque command saturated for at most 2 % of the 40 s, judged on the truth, with the score the count of gates. The configurations are compared in Table 9.7. As shipped, a lean PD tuned at 5 m/s, where the bicycle rides itself, passes 2 gates and falls at 17.0 s at 3.00 m/s, 11 m after the gust. The suggested scheduled LQR passes all eight within 0.40 m of their centers and finishes at 32.3 s. At seeds 1 to 5 the shipped settings fall between 14.75 and 17.04 s with 2 gates and the suggested ones pass with their worst gate 0.39 to 0.44 m off center. The state of the suggested run at 11.4 s, inside the gust and the weave band, is shown in Figure 9.3, and Figure 9.4 draws the gates and the saturation of every configuration.

9.7 Technology in Practice

What Measures It

A riderless test bicycle built to check a model of the benchmark's kind measured the roll rate, the yaw rate, the steer angle and the rotation of the rear wheel, and its eigenvalues were fitted to the measured lean rate [11]. A balance-assist bicycle reads its roll rate with a micromachined rate gyroscope and

Table 9.7 The Slalom Over 40 s, Default Seed

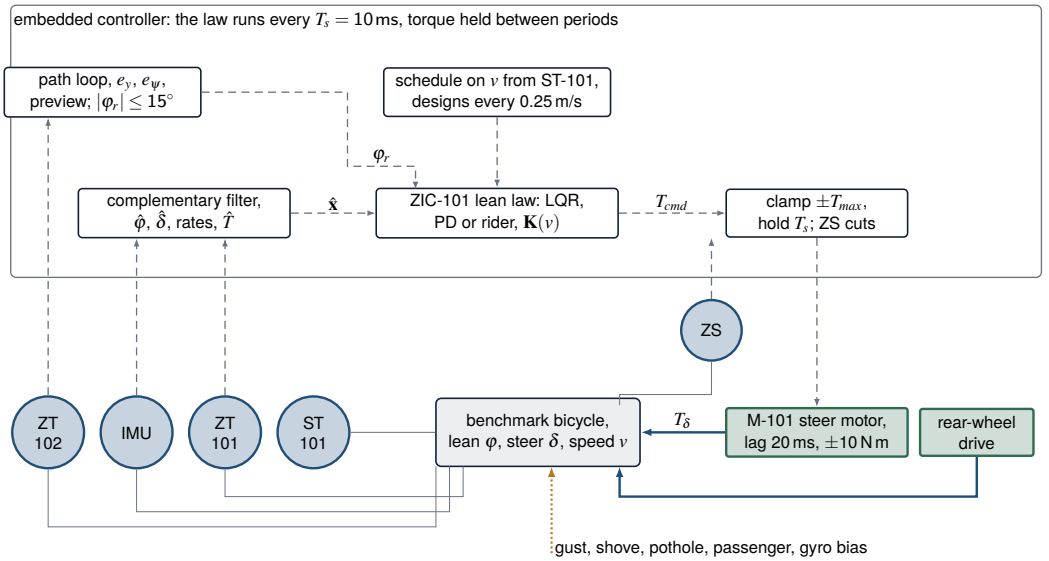
Configuration	Gates	Outcome	Sat.	Verdict
PD (10, 8) at 5 m/s, fixed (shipped)	2	fall 17.04 s	4.85 %	fails
scheduled LQR (suggested)	8	finish 32.3 s	0.10 %	passes
PD (10, 8) scheduled	5	finish 32.9 s	0.00 %	fails
LQR at 5 m/s, fixed	8	finish 32.6 s	0.10 %	passes
rider model	2	fall 12.39 s	0.00 %	fails
suggested, $\omega_p = 2.5\,\mathrm{rad/s}$	5	finish 36.5 s	12.68 %	fails
suggested, raw accelerometer	0	fall 6.12 s	0.00 %	fails

the speed of its rear wheel with a wheel speed sensor [17]. The chapter's instruments have the same shape: a gyro with noise and bias, an accelerometer that cannot tell gravity from the turn, an encoder on the steer and an exact wheel speed, merged by the complementary filter of Equation (9.10).

> **TECHNOLOGY** **Sensor: Rate Gyros on the Frame, an Angle Sensor on the Steer**
>
> One riderless test bicycle carried two angular rate gyros, rated to $\pm 100\,°/\mathrm{s}$, clamped at right angles on the seat post to read the lean rate and the yaw rate, with a spirit level to set them square. A potentiometer coupled straight to the steering axis read the steer angle, and a ten-pole magnet ring on the rear wheel gave the speed from the time between zero crossings, a staircase on the trend. Sampling was raised from 100 to 400 Hz to keep the speed signal from aliasing, but the steer signal then turned erratic [11]. The chapter's encoder has 2^{14} counts a turn, and its wheel speed is exact.

What Moves It

Real machines use two kinds of steering actuator. The balance-assist bicycle applies a torque between the head tube and the steer tube, and the motor's torque and the rider's act on the steer in sum [17]. A bicycle robot built for vehicle testing treats its steer angle as perfectly controlled, so its model runs from the steer angle to the lean and leaves the steer torque out [18]. The chapter's steer motor is of the first kind, a torque with a 20 ms lag and a limit of 10 N m, which is why case 9 can ask for more torque than the motor has.



IMU: rate gyro and accelerometer at h_s; ZT-101 steer encoder, 2^{14} counts a turn; ST-101 wheel speed; ZT-102 lane position

Figure 9.5 The balance loop on test-bicycle hardware. The inertial unit on the frame (rate gyro and accelerometer) and the steer encoder ZT-101 feed the complementary filter, the lane position ZT-102 feeds the path loop, which commands the lean, and the wheel speed ST-101 picks the gains from the schedule. The lean law ZIC-101 sends a torque command, clamped and held for one period, to the steer motor M-101, and the tilt switch ZS cuts the motor at a fall. Notice that speed enters the loop twice: it moves the bicycle's own poles and it chooses the controller's gains.

> **TECHNOLOGY** **Final Element: A Steering Motor in Parallel With the Rider**
>
> A balance-assist electric bicycle carries a motor in the steering assembly that applies up to 7 N m between the head tube and the steer tube. A motor controller turns the commanded torque into applied torque, and the rider's torque on the bars adds to it. The law is roll rate feedback with a gain that falls linearly as the speed rises toward about 4.7 m/s. With the higher gain it stabilizes the riderless bicycle's weave down to 3.4 km/h [17]. The chapter's motor is a torque source with a 20 ms lag and a limit of 10 N m, and the dummy's hands put no torque on the bars.

Where the Controller Runs

On the balance-assist bicycle the law runs on a small microprocessor carried on the bicycle, and every sensor is logged on a vehicle data bus at 100 Hz or faster [17]. An earlier test bicycle recorded through a data acquisition unit with 14-bit inputs into a laptop fixed on its rear rack, where the laptop's mass reduced the bicycle's stability and was weighed as part of the measured machine [11]. The chapter's controller runs every 10 ms on integration steps of 1 ms and holds its torque between periods, and that hold alone moves the stability edge of the lean PD tuned at 2 m/s from 5.59 to 3.45 m/s. The block that holds the chapter's structure is the schedule, a table of regulator designs interpolated at the measured speed. Figure 9.5 draws the loop on test-bicycle hardware.

What Is Modernizing It

The first development is balance assistance on consumer electric bicycles. One study added a steering motor to such a bicycle and, with 26 participants riding on a treadmill at 6 km/h and pushed at the handlebars, found the chance of recovering significantly higher with the assistance on [17]. The second is the self-balancing bicycle robot as a cyclist target for testing driver assistance systems. Unlike a stable moving platform, such a robot is unstable and so leans more naturally [18]. The third

is the schedule built from data. Gray-box identification at a few speeds replaces measuring every part, and the next steps named are adapting the model to changing loads, such as different dummies, by adaptive control and online identification [18]. The chapter's regulator trusts the declared bicycle, and its 20 kg passenger shows what such a learner would have to correct.

IN PRACTICE **Commissioning a Schedule From Closed-Loop Tests**

A bicycle robot was tuned without taking the bicycle apart. Its balance loop was closed first, and periodic multisine signals up to 25 Hz were added at 2.2, 2.4, 2.6 and 2.8 m/s. Averaging over the periods removes the bias that closing the loop puts into a frequency response. A gray-box model with the speed as a parameter was fitted to all four, and the gains were scheduled for the same bandwidth at every speed: 1.5 Hz for the inner balance loop and 0.15 Hz for the outer lean loop. The schedule then held the robot at 1.7 m/s, outside the tested range [18]. The chapter's path loop is likewise kept far slower than its lean loop.

9.8 Limitations

The bicycle is the Whipple model linearized about upright running, so it has no tires, no frame flex, no wobble and no rider moving on it, and its lean is trusted only for small angles. The matrices follow the speed quasi-steadily, which the benchmark does not cover, the nonlinear option is gravity alone and ends the capsize in a steady turn rather than a fall, and the cross slope's lean torque is a stand-in that knife-edge wheels would not feel. The motor, the stop, the fall angle, the instruments and every threshold are teaching values rather than any test bicycle's.

The Map and Loop views are linear analyses on the true state, with the motor's lag and the controller's sampling but without noise or the estimator, and the rider's delay enters them through a Padé approximant. The two-gain PD schedule leaves a band near the weave speed in which the closed loop grows slowly, which the case teaches rather than repairs. The regulator and the schedules trust the declared bicycle, and no estimator learns a passenger or a gyro bias. The rider model is a control law with passive arms, not a fitted model of a person.

OPEN CHALLENGE **How a Stabilized Bicycle Feels to Its Rider**

A steering motor that stabilizes a bicycle also changes the machine the rider must steer. On one balance-assist bicycle the controller raised the weave frequency about threefold, to almost 1 Hz, which a rider can still follow but which may feel like a normal bicycle at 25 km/h. Riders of a similar bicycle reported dissatisfaction. Fewer falls in a laboratory cannot yet be turned into fewer falls on the road, because detailed data on how people fall in natural riding are missing. A long study of normal use against a control group would give the strongest evidence [17]. The rider of this chapter is a control law, not a fitted model of a person.

9.9 Conclusion

The simulator of this chapter derives the benchmark bicycle's four matrices from its geometry and masses, to 1.4×10^{-14} of the published ones, and its eigenvalues reproduce the published table to 6.2×10^{-13}. The same derivation runs the bicycle. Hands off, it falls in 2.00 s at 3 m/s, rides itself at 5 m/s and tips over in 20.2 s at 8 m/s, and a trail of −0.02 m removes the window. A lean loop with the wrong sign falls in 0.61 s against 1.66 s with nobody steering, and gains tuned at 2 m/s are unstable above 3.45 m/s once the controller's 10 ms hold is counted. Their two-gain schedule rides a ramp to 8 m/s without saturating but leaves a band from 2.81 to 4.29 m/s where the closed loop grows at up to $0.157\ \mathrm{s}^{-1}$, and the scheduled regulator holds every speed from 0.5 to 10 m/s. A left turn starts with 4.3° of right steer, a path loop must stay slower than the lean it commands, and at 0.3 m/s, where the steer's grip on the lean has reversed, none of 123 designs holds the bicycle up. The bicycle is linear, rigid and riderless. A fuller model would add tire relaxation and frame compliance for wobble, a nonlinear model that falls as the real machine does, a rider who moves on the frame and an estimator for the gyro's bias. The simulator is a bicycle whose speed decides which controller it needs.

KEY POINTS

- A bicycle with nobody on it rides itself only between two speeds, 4.292 and 6.024 m/s.
- The steer's grip on the lean grows with the square of the speed, passes through zero at 0.58 m/s and is reversed below.
- A left turn starts with 4.3° of right steer, and a path loop must stay slower than the lean it commands.
- Gains tuned at 2 m/s are unstable above 3.45 m/s once the 10 ms hold is counted, while the scheduled regulator holds every speed from 0.5 to 10 m/s.

Where else this applies. Chapter 4 meets a gain that changes sign with the operating point, and Volume 3, Chapter 10 a wrong-way start.

References

[1] J. P. Meijaard, J. M. Papadopoulos, A. Ruina, and A. L. Schwab, "Linearized dynamics equations for the balance and steer of a bicycle: a benchmark and review," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2084, pp. 1955–1982, 2007.

[2] R. S. Sharp, "On the stability and control of the bicycle," *Applied Mechanics Reviews*, vol. 61, no. 6, p. 060803, 2008.

[3] ——, "The stability and control of motorcycles," *Journal of Mechanical Engineering Science*, vol. 13, no. 5, pp. 316–329, 1971.

[4] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[5] 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.

[6] N. Getz, "Control of balance for a nonlinear nonholonomic non-minimum phase model of a bicycle," in *Proceedings of 1994 American Control Conference - ACC '94*, vol. 1, 1994, pp. 148–151.

[7] A. L. Schwab, J. P. Meijaard, and J. D. G. Kooijman, "Lateral dynamics of a bicycle with a passive rider model: stability and controllability," *Vehicle System Dynamics*, vol. 50, no. 8, pp. 1209–1224, 2012.

[8] D. J. N. Limebeer and R. S. Sharp, "Bicycles, motorcycles, and models," *IEEE Control Systems*, vol. 26, no. 5, pp. 34–61, 2006.

[9] A. L. Schwab and J. P. Meijaard, "A review on bicycle dynamics and rider control," *Vehicle System Dynamics*, vol. 51, no. 7, pp. 1059–1090, 2013.

[10] K. J. Åström, R. E. Klein, and A. Lennartsson, "Bicycle dynamics and control: adapted bicycles for education and research," *IEEE Control Systems*, vol. 25, no. 4, pp. 26–47, 2005.

[11] J. D. G. Kooijman, A. L. Schwab, and J. P. Meijaard, "Experimental validation of a model of an uncontrolled bicycle," *Multibody System Dynamics*, vol. 19, no. 1-2, pp. 115–132, 2008.

[12] J. D. G. Kooijman, J. P. Meijaard, J. M. Papadopoulos, A. Ruina, and A. L. Schwab, "A bicycle can be self-stable without gyroscopic or caster effects," *Science*, vol. 332, no. 6027, pp. 339–342, 2011.

[13] D. E. H. Jones, "The stability of the bicycle," *Physics Today*, vol. 23, no. 4, pp. 34–40, 1970.

[14] P. Basu-Mandal, A. Chatterjee, and J. M. Papadopoulos, "Hands-free circular motions of a benchmark bicycle," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2084, pp. 1983–2003, 2007.

[15] N. H. Getz and J. E. Marsden, "Control for an autonomous bicycle," in *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, vol. 2, 1995, pp. 1397–1402.

[16] J. D. G. Kooijman and A. L. Schwab, "A review on bicycle and motorcycle rider control with a perspective on handling qualities," *Vehicle System Dynamics*, vol. 51, no. 11, pp. 1722–1764, 2013.

[17] M. T. Haitjema, L. Alizadehsaravi, and J. K. Moore, "Automatic bicycle balance assistance reduces probability of falling at low speeds when subjected to handlebar perturbations," *IATSS Research*, vol. 50, no. 1, pp. 701–711, 2026.

[18] Y. Wang, F. Bruzelius, and J. Sjöberg, "Gain-scheduled bicycle balance controller based on system identification," in *16th International Symposium on Advanced Vehicle Control (AVEC 2024)*, ser. Lecture Notes in Mechanical Engineering. Springer Nature Switzerland, 2024, pp. 679–685.

9.10 Exercises

9.1 Concept check · 2 min

In case 2 the bicycle is released hands off at 8 m/s from a 2° lean. It leans slowly into a tightening spiral and passes 40° after 20.2 s. Which mode brings it down?

(a) The weave, whose real part turns positive again above the capsize speed and grows slowly

(b) The castering mode, which at 8 m/s becomes too fast for the steer to follow without control

(c) A limit cycle of lean and steer that starts when the steer reaches its 45° stop

(d) The capsize, a slow real mode dominated by lean that grows at $0.143\,\mathrm{s}^{-1}$ at this speed

9.2 Concept check · 1 min

In case 6 the scheduled LQR holds the bicycle upright at 5 m/s, and at 2 s it is asked for a 5° lean to the left. Which way do the bars turn first, and why?

(a) Right, by about 4.3°, so that the wheels run out from under the bicycle and it starts to lean left

(b) Left, by about 4.3°, because the regulator steers toward the turn and the lean then follows the steer

(c) Right, because the spinning front wheel's gyroscopic torque reverses the steer above the weave speed

(d) Neither way: the regulator holds the steer at zero until the lean has reached the 5° requested

9.3 Concept check · 2 min

In case 9 the scheduled LQR, stable on paper at 0.3 m/s, lets the bicycle fall from a 2° push, and so do 122 other designs. What is the reason?

(a) The LQR's gain table has no design at so low a speed, so the regulator runs with the gains of a faster bicycle

(b) The steer's grip on the lean has reversed and shrunk, so catching the push needs far more torque and steer than the motor and the stop allow

(c) The gyro's noise, multiplied by the very large scheduled gains, keeps the motor saturated before the push can be caught

(d) The weave grows at $3.53\,\mathrm{s}^{-1}$ at this speed, which is faster than a controller sampling every 10 ms can follow

9.4 Theory and calculation · 6 min

Use the printed matrices of Equation (9.5) with $g = 9.81\,\mathrm{m/s^2}$, $w = 1.02\,\mathrm{m}$ and $\lambda = 18°$.

(a) Find the gravity part of the steer's grip on the lean, $gK_{0\varphi\delta}$, the grip $gK_{0\varphi\delta} + v^2K_{2\varphi\delta}$ at 5 and 1 m/s, and the speed at which the grip is zero. *Checkpoint:* −25.5 N m/rad, 1889 N m/rad, 51.1 N m/rad and 0.577 m/s.

(b) In a steady turn the rates are zero and $T_\varphi = 0$, so the lean row of Equation (9.1) fixes the steer that holds a lean. Find δ/φ at 5 m/s, and the steer for a 5° lean to the left. *Checkpoint:* 0.420, so 2.10° to the left.

(c) Find the radius of that turn from Equation (9.6), and compare it with $v^2/(g\tan\varphi)$. *Checkpoint:* 29.2 m against 29.1 m.

9.5 Theory and calculation · 6 min

Use the eigenvalues of Table 9.3.

(a) At 1 m/s, find the doubling time of the weave and its period. Explain why the fall at 1 m/s does not look like a weave. *Checkpoint:* 0.197 s and 7.8 s.

(b) At 8 m/s, find the e-folding time of the capsize, and the time a pure exponential would take to grow from 2° to 40°. Compare with the 20.2 s measured. *Checkpoint:* 6.98 s and 20.9 s.

(c) In case 5 the drive ramps the speed at 0.5 m/s^2. How long does the bicycle spend in the band from 2.81 to 4.29 m/s, and by what factor at most can a growth of 0.157 s^{-1} enlarge a disturbance there? *Checkpoint:* 2.96 s and 1.59.

9.6 Theory and calculation · 6 min

The lean estimate of Equation (9.10) runs every $T_s = 10\,\text{ms}$ with $\tau_{cf} = 0.5\,\text{s}$.

(a) Find the filter coefficient a, and the share of each new accelerometer reading that enters the estimate. *Checkpoint:* 0.980 and 2.0 %.

(b) At 1 m/s the scheduled LQR's gain on the lean rate is 203 N m s/rad. Find the RMS torque that the gyro's 0.1 °/s of noise puts on the command through that gain. *Checkpoint:* 0.35 N m.

(c) The bicycle runs a steady left turn at 5 m/s with 5° of lean and 2.10° of steer. Neglect noise and $h_s\ddot{\varphi}$, and find what the accelerometer reads, $\varphi - v\dot{\psi}/g$ with $\dot{\psi}$ from Equation (9.6). *Checkpoint:* about $-0.01°$, almost no lean.

9.7 App challenge · 8 min

Open case 4, "Steer into the fall: the lean loop's sign", at the default seed and 1× speed.

(a) Press Run as shipped and let it fall. Read the Fall tile, and say which way the bars turn relative to the lean. *Checkpoint:* 0.61 s, away from the lean.

(b) Switch to Manual and press Hands off, so the steer torque reads 0 N m (until then the hands keep the loop's last torque). Press Reset and Run, and read the Fall tile. *Checkpoint:* 1.66 s.

(c) Switch back to Auto, set Feedback sign to Into the fall, press Reset and run for 10 s. Read the Largest torque tile, and from the lean trend or the CSV column `truth_lean_deg` the time after which the lean stays within 0.5° of upright (to 0.05 s). *Checkpoint:* 2.8 N m and about 0.93 s.

9.8 App challenge · 8 min

Open case 5, "One tuning from 2 to 8 m/s", at the default seed and 2× speed. The PD tuned at 2 m/s rides a speed ramp to 8 m/s that starts at 3 s.

(a) Run the full 30 s as shipped and read the Steer saturated tile. From the CSV columns `steer_saturated` and `speed_mps`, find the time and speed of the first saturation. *Checkpoint:* 13.94 s (46.5 %); first at 11.9 s, at 6.44 m/s.

(b) From `truth_steer_deg` and `truth_lean_deg` between 20 and 30 s, find the swing of the steer and of the lean and the frequency of the steer. Is this the bicycle's own weave, at 1.35 Hz at 8 m/s? *Checkpoint:* ±7.7°, ±1.8°, about 2.9 Hz.

(c) Press Reset, set Gain schedule on speed to On, and run the 30 s again. Read the Steer saturated and Largest torque tiles. *Checkpoint:* 0.00 s and 1.6 N m.

9.9 App challenge · 8 min

Open the challenge, "The slalom on a speed program", at the default seed and 5× speed. Change nothing but the controller settings named, each before pressing Run, and let every run go the full 40 s.

(a) Run it as shipped (lean PD tuned at 5 m/s). Read the Gates, Fall and Steer saturated tiles. *Checkpoint:* 2 of 8, at 17.04 s, 4.9 %.

(b) Press Reset, set Controller to Full-state LQR and Gain schedule on speed to On, and run again. Read the same tiles and the Challenge target. *Checkpoint:* 8 of 8, none, 0.1 %, met.

(c) Press Reset and set Gain schedule on speed to Off, so the LQR is designed at 5 m/s only. Run again. Then press Reset, set Controller to Lean PD with the schedule On, and run once more. Which passes? *Checkpoint:* fixed LQR 8 of 8 and met; scheduled PD 5 of 8, not met.

9.10 Application area · 3 min

The simulated bicycle is the Whipple model with knife-edge wheels.

(a) Real bicycles and motorcycles can show wobble, a fast steering oscillation. Why does the simulator never show it, and what would a model need to show it? *Checkpoint:* no tire or frame compliance; add tire relaxation and frame compliance.

(b) The road's cross slope enters Equation (9.8) as a lean torque. Can the simulator be used to judge how a real crowned road affects a test bicycle? *Checkpoint:* no; the term is a teaching stand-in.

9.11 Application area · 4 min

A delivery bicycle with a steer assist carries a 20 kg child seat on its rear rack. The assist's model is the empty bicycle, as in the simulator.

(a) The passenger narrows the self-stable window from 4.29–6.02 m/s to 4.47–5.89 m/s. Which matrices change, and which does the controller use? *Checkpoint:* all four the physics integrates; the declared, empty ones.

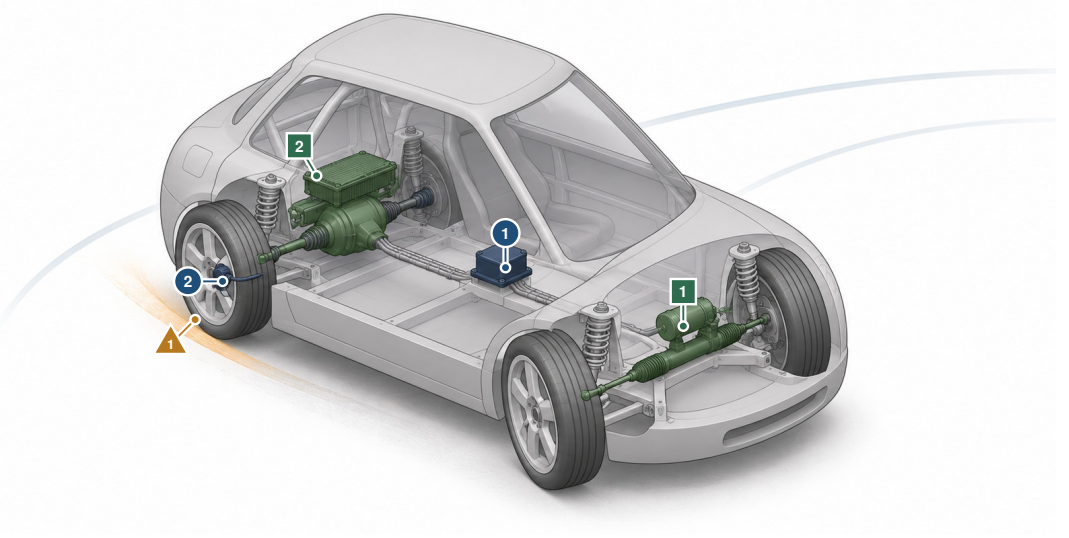
(b) How would you check, with the simulator, that the scheduled LQR still serves the loaded bicycle across its speed range, and what would you change in the product if it did not? *Checkpoint:* passenger slider, Map view and runs across speed; schedule on a measured load or design for the range.

9.12 Review the agent · 5 min

An AI assistant reviews the lean PD of case 5 for a delivery bicycle that cruises at 3.5 m/s. It writes: "Scheduled from the 2 m/s design, the closed loop is stable from 2 to 8 m/s, with its weakest root at $-0.062\ \mathrm{s}^{-1}$ near 3.6 m/s. The schedule is safe to ship." Find the flaw, and say how you would check the claim on the simulator.

CHAPTER 10

Holding a Drifting Car at an Unstable Equilibrium With Steer and Throttle



The controller moves 1 the steering servo on the front rack, which turns the front wheels
2 the electric drive on the rear axle, which sets the drive torque

The controller reads 1 the navigation unit with its gyro: sideslip, yaw rate, speed and position
2 the encoder that reads how fast the driven wheels spin

What disturbs it 1 the grip of the pad under the tires, lower on the wet patch

Notice that the rear tires slide at their grip limit, so the drive torque on the rear axle is the handle left on how hard they push sideways. Illustration, not to scale.

IN PLAIN TERMS **Holding a Car Sideways in a Slide**

The main idea. A drifting car circles a cone sideways, its rear tires sliding at the edge of their grip. A computer can hold it there by turning the front wheels and setting how hard the rear wheels are driven.

The challenge. Like a pencil balanced on its tip, the car left alone spins out within seconds. Its sliding rear tires cannot push sideways any harder, so the only handle left on them is their forward push, which takes sideways grip away as it grows.

Why it matters. Many machines must be held at a point they would leave on their own.

THREE WAYS TO READ THIS CHAPTER

Explore	11 min	the first page, the fact in the Introduction (Section 10.1), the history segment in Section 10.2, the try-its and the fact in Section 10.6, 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 10.6), Technology in Practice (Section 10.7), Limitations (Section 10.8) and the open challenge

The control question A rear-drive car can circle a cone sideways in steady state, its rear tires sliding at the edge of their friction. That drift is an unstable equilibrium whose sliding tire cannot raise its lateral force. What handle is left, and which controllers can hold the car there?

Learning objectives

- Solve the force balances of a steady drift for the rear forces and the drive force on the friction circle.
- Read the two unstable roots of the drift from its linearization, and relate the slide's root to the growth measured on a trend.
- Explain why a yaw-rate grip controller spins the car and why two single loops on throttle and steer fight when the speed target moves.
- Compare a two-input LQR with the nested loops after a kick, and schedule its equilibrium and gain on a friction estimate.
- Show that a front-drive car has no saturated-rear drift when both axles see the same friction.

Prerequisites Sections 1.4.2 (steady states), 1.4.3 (gain schedule), 1.4.4 (eigenvalues), 1.4.5 (discrete LQR, Bryson's rule), 1.4.7 (limits from an unstable pole) and 1.6.1 (engine). Chapter 9 for an LQR scheduled on speed.

The app https://apmonitor.com/apps/autonomous-drift/

See also Volume 20, Chapter 2 for a steering loop over a proportional-integral speed loop on a kinematic bicycle, whose wheels roll without slipping.

10.1 Introduction

On a dry skidpad the simulated coupe can travel at 8 m/s with its nose 35° inside its direction of travel, yawing at 0.927 rad/s around a circle of 8.63 m radius. Its rear tire slips at −42.4° and pushes with all of the 5.12 kN its load and friction allow, its front wheels are turned 25.04° toward the outside of the turn, and the motor holds 3.72 kN of drive force at the rear contact. Every acceleration of the body is zero to 10^{-14}, so this drift is an equilibrium of the vehicle model. Started on it with the steering and the throttle held, the car is 5° off at 2.98 s and spun at 4.29 s.

Operating points that are useful only because feedback holds them are common. Statically unstable aircraft, balancing robots, magnetic bearings and an exothermic reactor at its middle steady state (Chapter 2) all ask a controller to hold a state the plant leaves on its own, and an unstable pole sets limits on every such loop that no tuning removes, among them a lower bound on the loop's bandwidth and a danger from actuator saturation [1] (Section 1.4.7). A car past the limit of its rear tires belongs to the same class. A bifurcation analysis of the nonlinear single-track model shows the stable cornering equilibrium losing its stability as the tires saturate [2], and the large-sideslip "powerslide" is a steady state whose stability can be analyzed in the same terms [3]. The drift adds a feature that the textbook unstable plants lack. The actuator that would normally correct the error, the rear tire's lateral force, is the one that is saturated, so the correction has to come through the coupling between two other inputs.

Two beliefs are the simulator's targets. The first is that a sliding tire has lost its authority. In the drift the rear lateral force follows the drive force along the friction circle, so the throttle still moves it, and in case 3 the error grows at 1.98/s, close to the linearization's 1.95/s, precisely because nothing in the open loop uses that handle. The second is that a controller that steers a car well on grip will hold it at the limit. The yaw-rate and speed loop of case 2, the structure of a road car's stability control, spins the car at 3.61 s when asked for 30° of sideslip and spins it for every request of 10° or more. A transfer function about a grip operating point shows neither, because the first is a property of the saturated equilibrium and the second a consequence of designing around the wrong one.

✦ DID YOU KNOW? What Stability Control Was Expected to Save, 2007

When it made stability control mandatory in 2007, the United States highway safety agency estimated that the systems would save 5,300 to 9,600 lives and prevent 156,000 to 238,000 injuries a year once every light vehicle on the road had one. It expected single-vehicle crashes of passenger cars to fall by 34 percent [4].

The work that established drifting as a control problem was done on instrumented research cars, and a student meets its pieces apart. Equilibrium analysis and phase portraits are given for fixed speed [5], state feedback is designed at one operating point [6], and the experimental controllers follow a path with the drift inside [7, 8]. This chapter puts the pieces together on one declared car run in a browser. It asks what it takes to hold that unstable equilibrium, from bare hands through two interacting single loops to a scheduled two-input LQR, and where the equilibrium stops existing. The single-track model, the brush tire, the proportional-integral loops and the LQR are all established methods. The results are measurements on one parameter set and seed. With the speed free the drift has two unstable roots (1.95/s for the slide and 0.21/s for the speed under a held torque), where the fixed-speed analysis finds one, 1.89/s. A short proof shows that the same car with front drive has no saturated-rear equilibrium when both axles see the same friction, and each figure-eight reversal takes the rear tire off its friction circle for an instant. The simulator uses one pure rate function alike in the integrator, the equilibrium solver and the linearization, a shared tire module with the brush model and its combined-slip form, a two-input LQR computed through the series' Riccati module, and a measurement boundary that keeps every law away from the simulation's truth. Ten cases run from a car driven by hand to a scored minute of donut and figure-eight, and the app credits each case only when its own rule is met.

Section 10.2 traces the domain from the handling limit to automated drifting. Section 10.3 gives the car, its tires and its drive. Section 10.4 solves for the drift, finds its two unstable roots and shows where no drift exists. Section 10.5 describes the measurements and the laws: the grip controller, the driver's nested loops, the scheduled LQR, the friction estimate, the path loop and the reversal. Section 10.6 measures the ten cases, Section 10.7 describes the sensors, actuators and controllers of real drifting cars, Section 10.8 states the limitations and Section 10.9 concludes.

10.2 From the Handling Limit to Automated Drifting

The lateral dynamics of the car were first written as a control problem in the 1950s. Segel's linear model of the automobile's response to the steering wheel, checked against instrumented road tests, made handling a subject of dynamic analysis [9], and its reduction to sideslip and yaw on one track per axle still opens vehicle dynamics texts [10]. Its tires are linear, so such a model stops where the tires begin to slide. The tire models that carry it past that point came from the brush idea, in which the tread elements stick in the front of the contact patch and slide in the rear, so that the lateral force rises linearly with slip, bends over and settles at the friction limit [11].

What happens beyond the limit was studied first as a hazard. A bifurcation analysis of the nonlinear single-track model traced how the stable cornering equilibrium is lost as the tires saturate, and robust front-wheel steering control was designed to keep the car stable [2]. Production cars took the same view. Electronic stability control brakes individual wheels to keep the measured yaw rate and sideslip near a reference computed from the driver's steering, and a widely fitted system had five years of service experience by 2000 [12]. In that design a drift is a fault to be prevented, and the yaw-rate loop of case 2 is its simplest form.

Drifting had long been a driving technique before analysis caught up with it in the late 2000s. The handling characteristics and the stability of the steady powerslide were worked out with the tools of ordinary steady cornering [3]. Equilibrium families were computed against sideslip, phase portraits drawn at fixed speed, and a controller was built that uses the rear drive force as the handle on the saturated rear tire through the friction circle [5]. State feedback designed by LQR on the linearization at a drift equilibrium was shown to stabilize steady-state drifting of a rear-drive car [6]. A controller

Table 10.1 Signals and States of the Drifting Car at the Design Point

Variable	Role	Design	Range
β	CV, sideslip	−35°	SP 0–45°
r	CV, yaw rate	0.927 rad/s	—
ρ	CV of the path loop, distance	8.63 m	SP 7–12 m
V	speed	8 m/s	SP 4–12 m/s
δ_{cmd}	MV, steer command	−25.04°	±40°
T	MV, drive torque	1152 N m (61.9 %)	0–1860 N m
μ_F, μ_R	DV, friction	0.9	0.9 dry, 0.6 wet
kick, gust	DV, events	—	10°, 2 kN for 1 s
τ_T, θ_δ	DV, throttle lag, steer delay	0	0–300 ms, 0–200 ms
U, V_b	states, body velocities	6.55, −4.59 m/s	—
ω, κ	states, driven wheel	38.16 rad/s, 0.770	—
δ	state, road-wheel steer	−25.04°	±40°
x, y, ψ	states, pose on the pad	—	fence
θ	state, rear tire heat	0	0–1

framework for autonomous drifting, with a proof of stability, was validated experimentally on a research test car [7]. Path tracking followed, first with simultaneous stabilization and tracking of basic drifting trajectories [13] and then along a general path, which places drifting inside automated vehicle control beyond the stability limits [8]. Friction had meanwhile become a quantity a vehicle can identify in real time from GPS and motion measurements [14]. The simulator follows this sequence, from equilibrium to linearization to state feedback to path loop, as cases a student can run.

> ⧖ HISTORY **Stability Control Becomes Standard Equipment, 2007 to 2011**
>
> In April 2007 a final rule of the United States highway safety agency established Federal Motor Vehicle Safety Standard No. 126, which requires electronic stability control on light vehicles of 4536 kg gross vehicle weight rating or less. The rule describes the systems as automatic braking of individual wheels that helps the driver when the car begins to lose directional stability at the rear wheels, a spin out, or control at the front wheels, a plow out. Makers had fitted about 29 percent of light vehicles of model year 2006 with it. A phase-in from September 1, 2008 required it on every new light vehicle by September 1, 2011. The system must determine the yaw rate and estimate the sideslip or its rate, and the rule defines sideslip as the arctangent of the lateral over the longitudinal velocity of the center of gravity [4], the β of this chapter.

10.3 The Car and the Pad

The simulated vehicle is a generic rear-drive coupe of 1450 kg on a flat asphalt skidpad, with no manufacturer's data behind it, as shown in Figure 10.1. Its four tires are represented by two equivalent axle contacts, with forces, normal loads and stiffnesses summed over each axle's pair. Its center of mass lies $a = 1.0$ m behind the front axle and $b = 1.5$ m ahead of the rear, and the static loads are $F_{zf} = mgb/L = 8.53$ kN and $F_{zr} = mga/L = 5.69$ kN with $L = a + b$. The pad carries two clusters of nine cones 18 m apart, eight on a 5 m ring and one at the center of each, a painted 9 m circle around each cluster (the two circles touch, which makes the figure-eight), a fence at $-21 \le x \le 39$ m, $|y| \le 19$ m, and a quarter of circle A, from 3 to 16 m out, that can turn wet.

The variables appear in Table 10.1 by their role. The controlled variables are the sideslip and the yaw rate, with the distance from the cluster's center in an outer loop, and the two manipulated variables are the steer command and the drive torque. Ten states are integrated, the position (x, y) and heading ψ on the pad, the body velocities U and V_b, the yaw rate r, the driven wheel's speed ω and transient slip κ, the road-wheel steer δ and a rear tire heat θ. The sideslip is $\beta = \operatorname{atan2}(V_b, U)$ and the speed $V = \sqrt{U^2 + V_b^2}$. A left-hand drift has $r > 0$ and $\beta < 0$, and every target is entered as a positive sideslip.

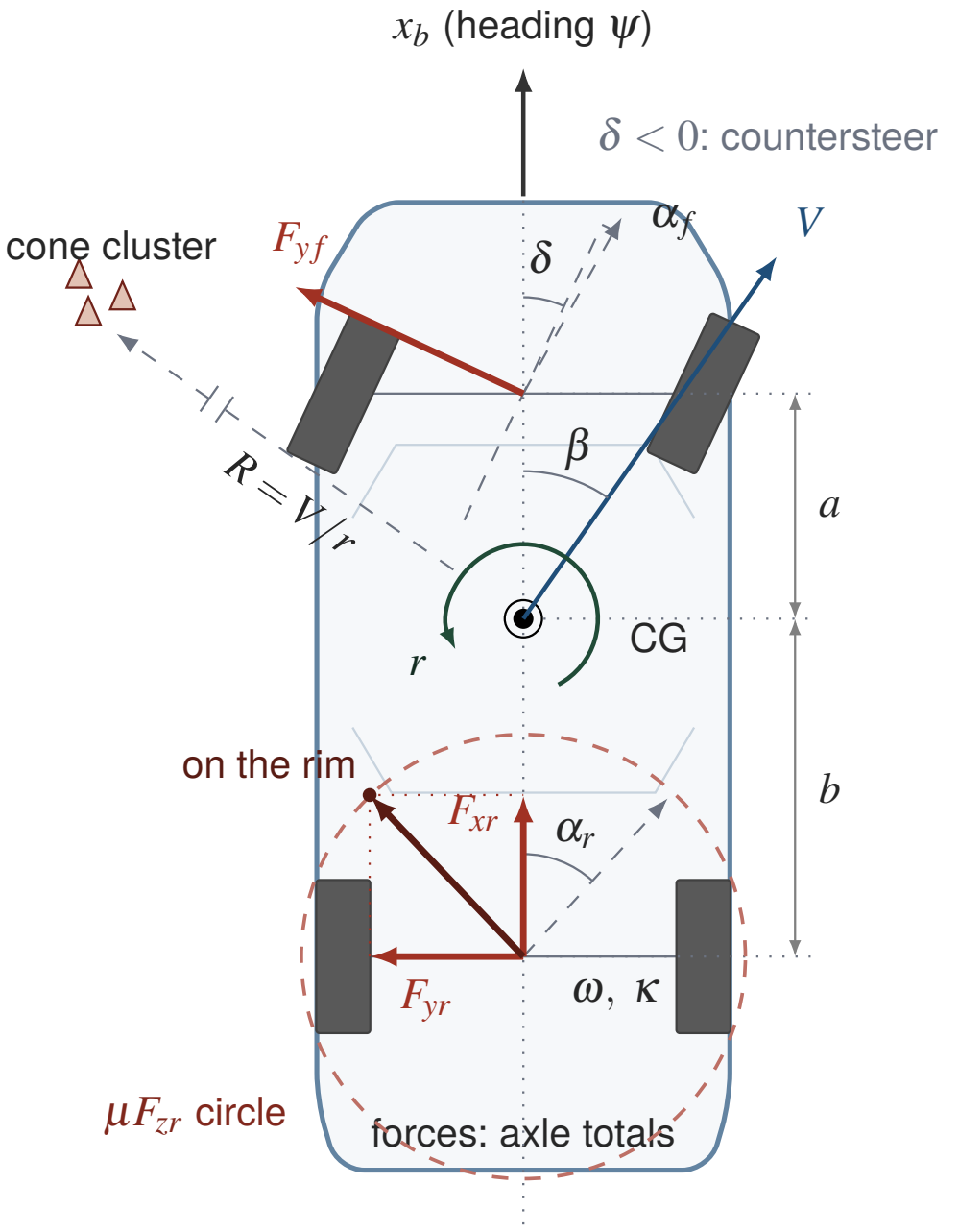


Figure 10.1 The four-tire car with the single-track model's axle-total forces, at the left-hand drift of the design point (35° at 8 m/s). The heading x_b and the velocity V differ by the sideslip β, whose components along and across x_b are U and V_b. Both front tires turn by δ against the turn. Their equivalent axle contact slips at α_f and carries the total lateral force F_{yf} within μF_{zf}. The driven rear pair shares the modeled speed ω and slip ratio κ. Its equivalent contact slips at α_r, and the axle forces F_{xr} and F_{yr} add to a resultant on the circle of radius μF_{zr}. The car yaws at r around a center $R = V/r$ away. Lengths to scale, forces at one scale, the radius broken.

10.3.1 Planar Motion

The body moves in the plane under the tire forces, aerodynamic drag and a possible lateral gust F_g, as shown in Equation (10.1).

$$\begin{aligned} m(\dot{U} - rV_b) &= F_{xr} + F_{xf}\cos\delta - F_{yf}\sin\delta \\ &\quad - \tfrac{1}{2}\rho_a C_d A\, U|U| \\ m(\dot{V}_b + rU) &= F_{yr} + F_{xf}\sin\delta + F_{yf}\cos\delta + F_g \\ I_z\dot{r} &= a(F_{yf}\cos\delta + F_{xf}\sin\delta) - bF_{yr} \\ \dot{x} &= U\cos\psi - V_b\sin\psi, \quad \dot{y} = U\sin\psi + V_b\cos\psi \\ \dot{\psi} &= r \end{aligned} \tag{10.1}$$

where F_{xf} is zero with rear drive and F_{xr} is zero with the front-drive option of case 9. The slip angles of the two contacts are the angles between each wheel's plane and its velocity, as shown in Equation (10.2).

$$\alpha_f = \operatorname{atan2}(V_b + ar,\ \hat{U}) - \delta, \qquad \alpha_r = \operatorname{atan2}(V_b - br,\ \hat{U}) \tag{10.2}$$

where $\hat{U} = \max(|U|, u_{min})$ keeps the angles defined near a standstill. At the design point $\alpha_r = -42.4^\circ$ and $\alpha_f = -4.15^\circ$, so the rear slides while the front grips.

10.3.2 Brush Tires and the Friction Circle

Both tires follow the Fiala brush model [11] from the framework's shared tire module. A tire that carries only lateral force, with cornering stiffness C and limit F, gives Equation (10.3).

$$F_y = \begin{cases} -Ct + \dfrac{C^2}{3F}|t|t - \dfrac{C^3}{27F^2}t^3, & |\alpha| < \operatorname{atan}\dfrac{3F}{C} \\ -F\operatorname{sign}\alpha, & \text{beyond} \end{cases} \tag{10.3}$$

where $t = \tan\alpha$. The undriven axle uses it with $F = \mu F_z$. The force leaves the line $-C\tan\alpha$ at once, reaches the limit with zero slope at the sliding angle, 5.48° for the rear on the dry pad, and stays there. The driven axle carries combined slip from its slip ratio and slip angle, as shown in Equation (10.4).

$$\begin{aligned} s_x &= \frac{C_x\kappa}{1+\kappa}, \quad s_y = \frac{C_\alpha\tan\alpha}{1+\kappa} \\ s &= \sqrt{s_x^2 + s_y^2}, \quad g = \frac{s}{3\mu F_z} \\ F &= \mu F_z\big(1 - (1-g)^3\big)\ \ (g<1), \qquad F = \mu F_z\ \ (g \ge 1) \\ F_x &= F\,s_x/s, \qquad F_y = -F\,s_y/s \end{aligned} \tag{10.4}$$

where C_x is the longitudinal stiffness per unit slip, C_α the cornering stiffness and the denominator $1+\kappa$ is floored at 0.1. The magnitude is the brush curve of the composite slip, so the form reduces to Equation (10.3) at $\kappa = 0$ and to the longitudinal brush curve at $\alpha = 0$, and once $g \ge 1$ the force lies on the circle of radius μF_z. Chapter 4 treats the longitudinal side alone, with a friction curve that peaks and then falls. Here the lateral force on the circle is $F_{yr} = -\sqrt{(\mu F_{zr})^2 - F_{xr}^2}\operatorname{sign}\alpha_r$, which is the friction circle as a control law: every newton of drive the rear carries is paid for in lateral grip. The force is directed along the stiffness-weighted slip (s_x, s_y), a common simplification. The full brush model points a sliding tire's force against its sliding velocity, along $(\kappa, \tan\alpha)$ [11], and at the design point the two directions are 43.48° and 49.85° from the wheel's plane, 6.37° apart. The equilibria below do not depend on the direction, because there the rear force is on its circle and the drive force is solved for directly. The slip ratio that holds it, the linearization and the closed loop do depend on it.

Table 10.2 Model Parameters. H: Hidden, A: Adjustable in the Advanced Panel

Symbol	Meaning	Value	Unit	
m, I_z	mass, yaw inertia	1450, 2300	kg, kg m^2	A
a, b	CG behind front, ahead of rear axle	1.0, 1.5	m	A
h	CG height (load transfer)	0.5	m	A
C_dA, ρ_a	drag area, air density	0.65, 1.2	m^2, kg/m^3	A, H
C_f, C_r	cornering stiffness, front, rear	120, 160	kN/rad	A
C_x	rear longitudinal stiffness	200	kN	A
μ_{dry}, μ_{wet}	friction, dry pad, wet patch	0.9, 0.6	—	A
σ	slip relaxation length	0.3	m	A
v_ε, u_{min}	slip floors	0.3, 0.5	m/s	H
$\Delta\mu_{heat}$	rear friction lost when hot	0.15	—	A
P_{heat}, τ_θ	heating power, time constant	40, 20	kW, s	A
R, J_w	wheel radius, axle inertia	0.31, 2.0	m, kg m^2	A
$F_{x,max}, P_{max}$	drive force, power limit	6, 120	kN, kW	A
$\delta_{max}, \dot{\delta}_{max}$	steer lock, servo rate	40, 400	°, °/s	A
τ_s	steer servo lag	30	ms	A

10.3.3 Driven Wheel, Steering and Drive

The driven wheel has its own speed and a transient slip that relaxes over a length σ, and the steer follows its command through a servo, as shown in Equation (10.5).

$$\begin{aligned} J_w\dot{\omega} &= T_q - RF_{x,drive} \\ T_q &= \min\left(T,\ P_{max}/\max(|\omega|,1)\right) \\ \sigma\dot{\kappa} &= R\omega - u_w - (|u_w| + v_\varepsilon)\kappa \\ \dot{\delta} &= \mathrm{clip}\left((\delta_{cmd} - \delta)/\tau_s,\ \pm\dot{\delta}_{max}\right) \end{aligned} \tag{10.5}$$

where u_w is the driven wheel's own forward speed (U at the rear), so the steady slip is the drive slip ratio $(R\omega - u_w)/(|u_w| + v_\varepsilon)$ and $v_\varepsilon = 0.3\,\mathrm{m/s}$ keeps it finite at rest. The computer clips the torque command to $\min(F_{x,max}R,\ P_{max}/\max(|\omega|,1))$, 1860 N m or the 120 kW power limit, and the plant applies the power limit again continuously, so the delivered power never passes 120.0 kW. The electric drive and its 2.0 kg m^2 axle make the throttle a fast torque source.

Two options are off in every case. Longitudinal load transfer moves hF_x/L onto the rear under drive, with the center of mass $h = 0.5\,\mathrm{m}$ high. Tire heat follows $\tau_\theta\dot{\theta} = P_{slide}/P_{heat} - \theta$, with P_{slide} the rear's force times its slip speed while it slides, and lowers the rear friction by $\Delta\mu_{heat}\min(1, \max(0, \theta))$.

10.3.4 Parameters, Numerics and Events

Table 10.2 gives every parameter. All are teaching values in the range the drifting literature reports [5–7], not any paper's vehicle. Physics runs on the framework's integrator (Section 1.6.1) with a step of 1 ms. The stiffest pair is the driven wheel against its slip, and halving the step leaves every case's fate unchanged. The growth rate of case 3 is 1.982/s at both steps, the recovery of case 5 0.914 s at both, and the challenge's scores 0 and 58 at both. Each case starts either on the drift equilibrium of its targets, placed on its circle, or on a grip turn with the kinematic steer, and a drift start is an equilibrium of the integrated model to a residual below 10^{-9}. Hard events live outside the derivatives. SPUN latches when $|\beta|$ passes 70°, with the crossing located inside the step, GRIPPED OUT when the rear has not slid for 2 s under a drift command of 10° or more outside a reversal, OFF PAD at the fence, and a cone falls when the body rectangle, grown by the cone's 0.15 m radius, touches it. These thresholds are teaching values. Real stability control works from yaw-rate and sideslip references set by the vehicle's own calibration [12], and friction values on real pavement come from tire testing [11].

10.3.5 Assumptions

The single-track model lumps each axle into one tire, so there is no lateral load transfer, no roll and no differential. A real drift car loses most of the inside rear's load and depends on a locked differential to keep both rear wheels sliding, and the model holds best where the two rear tires behave alike. The tires have constant stiffness and friction, no camber and no lateral relaxation, and the brush curve keeps its peak force while sliding, whereas a real tire's sliding friction usually falls below its peak. The combined-slip direction is the stiffness-weighted simplification described above. Axle loads are static unless load transfer is switched on. The drive is an electric torque source with a lumped axle. The pad is flat and its friction is uniform except for the wet quarter, whose edge is a step. The model is least reliable at very low speed, where the slip floors act, and at the moment a tire changes between sliding and gripping, where the brush curve's corner at the sliding angle meets a real tire's rounded peak.

10.4 The Drift as an Equilibrium

10.4.1 Solving for the Steady Drift

For a speed V and a sideslip β the car drifts in steady state when $\dot{U}$, $\dot{V}_b$ and $\dot{r}$ vanish together. With $U = V\cos\beta$ and $V_b = V\sin\beta$ fixed, Equation (10.1) gives three residuals in the three unknowns δ^*, r^* and F_{xr}^*, as shown in Equation (10.6).

$$\begin{aligned} 0 &= F_{xr} - F_{yf}\sin\delta - \tfrac{1}{2}\rho_a C_d A\, U^2 + mrV_b \\ 0 &= F_{yf}\cos\delta + F_{yr} - mrU \\ 0 &= aF_{yf}\cos\delta - bF_{yr} \end{aligned} \tag{10.6}$$

with F_{yf} from Equation (10.3) at μF_{zf} and the rear on its circle, $F_{yr} = \text{Fiala}\big(\alpha_r, C_r, \sqrt{(\mu F_{zr})^2 - F_{xr}^2}\big)$. A damped Newton iteration with a numerical Jacobian solves it from three starting guesses, with the drive force written as $\mu F_{zr}\tanh(z/\mu F_{zr})$ so that it can only be a force the tire can give. A second Newton pass on the integrated model's own rates then finds the slip ratio κ^*, wheel speed ω^* and torque T^* that hold F_{xr}^* at α_r, so the state returned is an equilibrium of Equations (10.1) to (10.5) to round-off. At the design point, 35° and 8 m/s on the dry pad, the solution is $\delta^* = -25.04°$, $r^* = 0.927\,\text{rad/s}$ (a radius of 8.63 m), $F_{xr}^* = 3716\,\text{N}$, $\kappa^* = 0.770$ and $T^* = 1152\,\text{N m}$, 44.0 kW of drive.

The family of these equilibria at 8 m/s is shown in Figure 10.2a. Between 5° and 45° the drive force rises from 2.10 to 4.17 kN and the radius from 7.92 to 8.84 m, while the steer passes from 8.5° into the turn to 36.2° against it, crossing zero between 12° and 13°. Countersteer is therefore a consequence of large sideslip rather than its definition. A deeper drift needs more drive, because the rear's lateral share must fall as the front's steer angle rises. On the wet pad the design target moves to $\delta^* = -27.88°$, $r^* = 0.647\,\text{rad/s}$, a radius of 12.37 m and 2369 N, which is why a design fixed on the dry pad's numbers fails on the wet patch of case 6.

10.4.2 Two Unstable Roots

The model is linearized about the equilibrium by central differences of its rate function, as in Section 1.4.4, in the six states $\tilde{x} = [U, V_b, r, \omega, \kappa, \delta] - x^*$ and the two inputs $\tilde{u} = [\delta_{cmd}, T] - u^*$, giving $\dot{\tilde{x}} = A\tilde{x} + B\tilde{u}$. At the design point A has the eigenvalues $+1.955$, $+0.206$, -8.55, $-11.60 \pm 15.85i$ and -33.33 1/s, the last being the steer servo's $1/\tau_s$. Two roots are unstable, at every sideslip from 5° to 45°, as shown in Figure 10.2b. The faster, the slide, doubles an error every 0.355 s and peaks at 1.956/s near 37°. The slower is the speed drifting under a held torque. A three-state car written from the same tire functions, with (U, V_b, r) as states and the steer and drive force held, has both, 1.966 and 0.215/s at the design point, while its lateral pair (V_b, r) at a fixed speed has only one, 1.889/s. That single root is the one an analysis at fixed speed reports [5], and the second appears only when

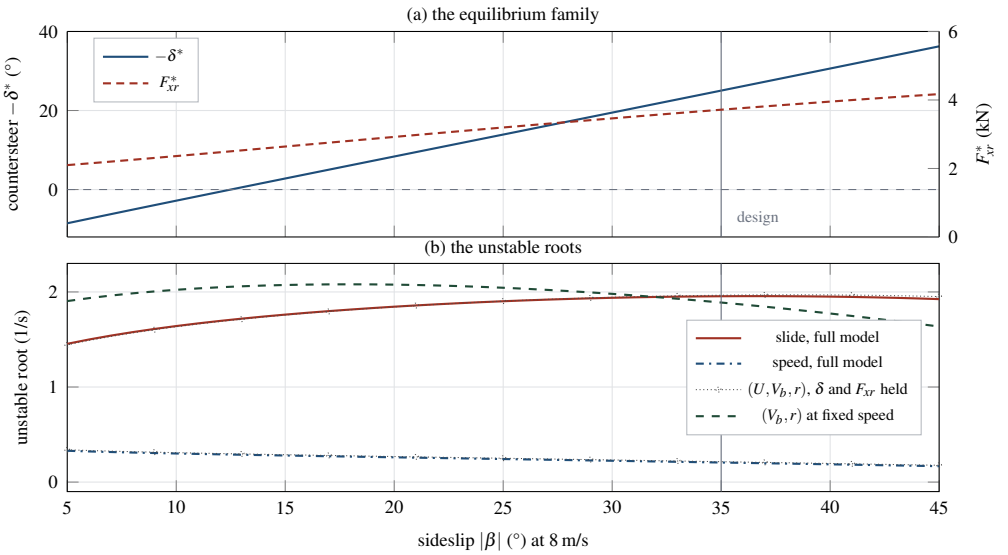


Figure 10.2 The drift equilibrium family at 8 m/s on the dry pad, from the simulator's own solver and linearization. (a) Countersteer (solid) and rear drive force (densely dashed) against sideslip. The steer points into the turn below 12.5°. (b) The unstable roots. The full model has two at every sideslip, the slide (solid) and a slow speed mode (dash-dotted), and the three-state car with steer and drive force held reproduces both (dotted with circles). Holding the speed fixed leaves one (dashed). The vertical line marks the design point.

the speed is free. At 10 m/s the pair moves to 2.284 and 0.100/s, and on the wet pad at 8 m/s to 1.991 and 0.090/s.

Because the rear is saturated, its lateral force cannot respond to its slip angle, and the yaw moment that would restore a gripping car is missing. The linearization says how the two inputs act instead. Steering left raises the yaw acceleration through the gripping front, and more torque spins the wheel up, which moves the rear force along its circle toward drive and away from lateral grip, so the rear swings out further. Both signs are checked by the app's own tests.

Pause. Try Exercise 10.1 before reading on.

10.4.3 Where No Drift Exists

Case 9 drives the front axle and lets the rear roll. With the rear saturated and undriven, $|F_{yr}| = \mu_R F_{zr}$, and the yaw balance of Equation (10.6) asks the front for a body-lateral force $(b/a)\mu_R F_{zr}$. With static loads this is $\mu_R mgab/(aL) = \mu_R F_{zf}$. When both axles see the same friction, that is the front's whole friction circle, so the front can carry no body-longitudinal force. The longitudinal balance then needs $mr|V_b| + \frac{1}{2}\rho_a C_d A\, U^2 = 0$, which no drift can satisfy, because both terms are positive. The result holds for any weight distribution, and load transfer under forward drive only lowers F_{zf} against the same demand. When the front has more grip than the rear, the lateral balance gives $rU = \mu_R g$ (drag neglected), the front needs a longitudinal force $m\mu_R g \tan|\beta|$ beside its lateral $\mu_R F_{zf}$, and its circle allows a saturated-rear drift only up to

$$\tan|\beta| \le \frac{b}{L}\,\frac{\sqrt{\mu_F^2 - \mu_R^2}}{\mu_R}. \tag{10.7}$$

For $\mu_F = 0.9$ on $\mu_R = 0.6$ the bound is 33.9°, and the app's solver finds such drifts from 2° to 33° at 8 m/s. With equal friction the bound is zero. Over 1230 targets, every 0.5 m/s from 5 to 12 m/s and

every degree from 5° to 45° on both surfaces, the front-drive solver never converges, and its smallest residual is 0.0343 m/s^2. The rear-drive solver converges at 1228 of the same targets with residuals below 2×10^{-11}. The two exceptions, cold starts at low speed and small sideslip, are discussed in Section 10.8.

🕮 **Pause.** Try Exercise 10.3 before reading on.

10.5 Implemented Control System

Every law runs every 10 ms and reads only a measurement packet. A dual-antenna navigation unit gives the speed (± 0.03 m/s), the sideslip ($\pm 0.15°$), the position (± 2 cm) and the heading ($\pm 0.1°$), a gyro the yaw rate ($\pm 0.2°/\text{s}$), an accelerometer the lateral acceleration (± 0.05 m/s^2), and encoders the wheel speed (± 0.05 rad/s) and the steer ($\pm 0.05°$). Each noise is colored with a 20 ms correlation time and seeded, so a run replays bit for bit (Section 1.6.1). The body velocities follow as $U = V\cos\beta$ and $V_b = V\sin\beta$, and the transient slip, which no instrument measures, is replaced by its quasi-steady value $\hat{\kappa} = (R\omega - U)/(|U| + v_\varepsilon)$. The surface, the tire forces, the true slip, the heat and any gust never reach a law. Each law is written for a left-hand drift in direction-normalized signals (the sideslip $b = -s\beta$, the yaw rate sr, the steer $s\delta$, with $s = \pm 1$), so one law serves both directions.

10.5.1 The Grip Controller

The grip controller is the stability-control pattern of the lane-keeping simulator of Volume 20, Chapter 2. A proportional-integral (PI) loop steers toward a yaw-rate target, and a second PI loop sets the torque from the speed error, as shown in Equation (10.8).

$$r_{ref} = \frac{V_{SP}}{R_{path}} + K_{\beta r}\max\big(0,\ \beta_{SP} - \max(b,0)\big) \tag{10.8}$$

where the second term asks for extra yaw rate while sideslip is missing. The request is one-sided because a two-sided one added yaw rate on grip, where b is negative, and drove the car into the cones. The defaults are $K_g = 0.5$ and $\tau_{I,g} = 0.5$ s on the steer, $K_v = 600$ N m per m/s and $\tau_{I,v} = 2$ s on the torque, and $K_{\beta r} = 1.0/\text{s}$.

10.5.2 The Driver's Nested Loops

Two PI loops reproduce the way a driver holds a slide, each on top of the dry pad's equilibrium as a feedforward, as shown in Equation (10.9).

$$T = T^* + \mathrm{PI}_\beta(\beta_{SP} - b), \qquad \delta_{cmd} = \delta^* + \mathrm{PI}_r(r^* - sr) \tag{10.9}$$

The throttle holds the sideslip because more drive takes lateral force from the rear, which slides out further, and the steer holds the yaw rate. Each loop moves the other's variable. The reference tuning is $K_\beta = 104.7$ N m/° (6000 N m/rad), $\tau_{I,\beta} = 0.5$ s, $K_r = 1.0$ and $\tau_{I,r} = 0.5$ s. Case 4 and the challenge ship a loose tuning with half the gains and twice the integral times.

Every loop is the series' velocity-form PID of Section 1.6.2, with back-calculation on the integral alone and a tracking time equal to the integral time τ_I. Its output is clamped to 0 to the torque limit and to ±40° of steer. Every loop not in charge tracks the applied commands, so Manual to Auto and any change of law are bumpless.

10.5.3 LQR About the Equilibrium

The LQR acts on the whole measured state, as shown in Equation (10.10).

$$\begin{aligned} u &= u^* - K(\hat{x} - x^*), \qquad J = \sum_k \tilde{x}_k^{\mathsf{T}} Q \tilde{x}_k + \tilde{u}_k^{\mathsf{T}} R_u \tilde{u}_k \\ K &= (R_u + \Gamma^{\mathsf{T}} P \Gamma)^{-1} \Gamma^{\mathsf{T}} P \Phi \end{aligned} \tag{10.10}$$

where Φ and Γ are the zero-order-hold discretization of (A,B) at 10 ms and P solves the discrete algebraic Riccati equation (Section 1.4.5). The series' Riccati module solves the single-input regulator, so the two-input equation is solved by its Kalman routine on the dual system $(\Phi^{\mathsf{T}},\Gamma^{\mathsf{T}})$, whose a-priori covariance satisfies the same equation [15]. The weights are diagonal by Bryson's rule, $Q=\mathrm{diag}(1/s_i^2)$ with scales of 1 m/s on U, $V\beta_s$ on V_b ($\beta_s=2^\circ$), 0.1 rad/s on r, 20 rad/s on ω, 1 on κ and 0.5 rad on δ, and $R_u=\mathrm{diag}(1/\delta_s^2,1/T_s^2)$ with $\delta_s=11.5^\circ$ and $T_s=800\,\mathrm{N\,m}$. At the design point the gain's steer row is $(-0.187,\ 0.008,\ 1.589,\ 0.002,\ 0.054,\ 0.724)$ and its torque row $(29.6,\ -2381,\ 4647,\ 45.0,\ 153.6,\ 349.9)$ in SI units, and the closed loop's roots are -0.81, -7.54, $-19.84\pm18.28i$ and $-33.06\pm20.64i$ 1/s. An independent Riccati recursion reproduces the gain to 2.6×10^{-8}.

Rather than solve the Riccati equation at every control period, the gain is scheduled, here on speed and friction where Chapter 9 schedules its LQR on speed alone. Nodes sit every 1 m/s from 4 to 14 m/s and every 0.1 of friction from 0.3 to 1.3, each designed at the exact equilibrium of its node, and the gain is interpolated bilinearly between the four nodes around the target. The equilibrium itself is solved exactly at the target speed (on a 0.05 m/s grid) and at the friction the schedule names, either the dry pad's 0.9 or the estimate $\hat{\mu}$ rounded to 0.02. When the LQR takes over from another law it adds the last command's difference as a bias that decays with a 0.5 s time constant. Where the solver finds no equilibrium, as in case 9, the LQR hands the car to the nested loops without feedforward.

10.5.4 Friction Estimate, Path Loop and Reversal

While the rear slides its force is on its circle, so its magnitude over its load is the friction. The lateral and yaw balances give the rear lateral force from the inertial measurements, and the drive force follows from the torque command less the wheel's spin-up, as shown in Equation (10.11).

$$\begin{aligned}\hat{F}_{yr}&=\frac{a\,m\,\bar{a}_y-I_z\bar{\dot{r}}}{L},\qquad \hat{F}_{xr}=\frac{T-J_w\bar{\dot{\omega}}}{R}\\ \hat{\mu}_{inst}&=\sqrt{\hat{F}_{xr}^2+\hat{F}_{yr}^2}\,\big/\,F_{zr}\end{aligned}\tag{10.11}$$

where the bars are 50 ms filters (the acceleration's started from its first reading). The estimate $\hat{\mu}$ follows $\hat{\mu}_{inst}$, clipped to 0.2–1.3, through a 150 ms filter only while the rear's slip speed exceeds 1 m/s, and holds otherwise, starting from 0.9. With front drive $\hat{F}_{xr}=0$. The idea follows the friction identification of motion-based estimators [14], simplified to a saturated tire.

The path loop trims the target speed on the distance ρ from the cluster's center and its rate, $V_T=V_{SP}-K_{p,\rho}(\rho-R_{path})-K_{d,\rho}\dot{\rho}$, clipped to $V_{SP}\pm2.5\,\mathrm{m/s}$, with $K_{p,\rho}=0.5\,(\mathrm{m/s})/\mathrm{m}$, $K_{d,\rho}=1.0\,\mathrm{s}$ and $R_{path}=9\,\mathrm{m}$. Because the equilibrium's radius grows with its speed (8.63 m at 8 m/s and 13.12 m at 10 m/s), choosing the speed chooses the circle, which is the idea of tracking a path by choosing among drift equilibria [13] in its simplest form. The figure-eight reversal is open loop. 10° of arc before the circles' common point the throttle is lifted and the steer goes to full opposite lock until the yaw rate has reversed past 1.5 rad/s, then 1200 N m is applied with the lock held until the new side's sideslip passes 5°, and the new side's law takes over. A transition counts once the sideslip has stayed 0.5 s in the new band within 5 s of the flick.

The application keeps three models apart. The plant is everything in Section 10.3. The identified model is the growth rate a learner reads from the trend of case 3. The controller's model is the solver's and the linearization's copy of the declared parameters, which knows neither the wet patch nor a gust, and which meets the true friction only through $\hat{\mu}$.

10.6 Ten Cases, Measured

The cases appear in Table 10.3. All share the reference settings (the LQR scheduled on $\hat{\mu}$, the path loop on, 35° at 8 m/s) and depart from them as their rows say. Each number below is taken from the simulator's engine run without a display at the seed it ships with, advanced in the same 1 ms steps the

Table 10.3 The Ten Cases

#	Case and settings	Lesson
1	Manual on the drift, stations at its inputs	the drift is unstable
2	Grip controller asked for 30°, started on grip	grip control at the limit
3	No controller, started on the equilibrium	the unstable root
4	Loose nested loops, speed target 8 to 10 m/s at 10 s, 30 s	two loops that fight
5	LQR on the dry schedule, 10° kick at 8 s	both actuators together
6	LQR on the dry schedule, path loop, wet quarter from 0, 30 s	the equilibrium moves
7	Figure-eight from 5 s	the reversal
8	LQR, path loop off, circle 2.5 m off center	the path loop
9	Front drive	no equilibrium exists
10	Challenge: loose nested loops, figure-eight from 30 s, wet from 25 s, 60 s	the whole problem

browser uses, with each named setting applied before the first step and a run length of 20 s unless a row names another. "In band" counts the seconds with the true sideslip within 5° of its target outside a reversal. Every case loads identically when chosen first, from the list after another case, or after Reset (Section 1.6.4). Figure 10.3 draws the time in band of every run that reports one.

10.6.1 By Hand and With No Controller

Case 1 starts in the drift with the steer and throttle stations at the equilibrium's inputs rounded to their resolution, −25.0° and 61.9 %, which leaves them 0.041° and 0.029 % away. Untouched, the car is 5° off at 2.98 s and spins at 4.29 s. Flooring the throttle spins it at 0.97 s, and closing it lets the rear grip again, GRIPPED OUT at 2.86 s, after which the car turns the other way with 13° to 14.5° of sideslip. The nested loops in Auto hold the band for 19.99 of the 20 seconds. Case 3 is the same start with no law. The sideslip error passes 0.5° at 1.82 s and 5° at 2.98 s, a factor of ten in 1.16 s and a growth of 1.98/s against the root's 1.95/s. With the inputs set exactly before Run the car stays on the equilibrium for the whole run, because the simulated state is then a fixed point to the last bit, and the Nudge event (1° of velocity) at 2 s takes it 5° off at 3.22 s. A real car always carries that nudge. The other targets of the family leave as fast, at 3.16 s from 25°, 3.13 s from 45° and 3.54 s from 35° at 10 m/s.

> **TRY IT IN THE APP** **3 · The Equilibrium, With No Controller**
>
> Load the scenario, press Run without touching anything and watch the sideslip badge. The car starts on its drift equilibrium, yet the sideslip error passes 0.5° at 1.82 s and 5° at 2.98 s, a factor of ten in 1.16 s. That growth is the slide's unstable root, read off a trend.

10.6.2 Grip Control and the Two Loops

The grip controller of case 2 steers into the turn for more yaw rate, the rear breaks away and the car spins at 3.61 s. Every request of 10° or more spins it, at 9.91 s for 10°, 5.62 s for 15° and 3.53 s for 35°, and a larger request gain spins it sooner (3.41 s at $K_{\beta r} = 2$). With the request at zero it holds the 9 m circle on grip with 7.23° to 7.30° of sideslip, the nose outside the velocity and no cone touched. Its loops assume a tire whose force rises with slip, which the drift does not have.

Pause. Try Exercise 10.2 before reading on.

The loose nested loops of case 4 hold 34.7° to 35.4° until the speed target steps to 10 m/s at 10 s. After it they pass the disturbance back and forth in a swing with a 3.26 s period that grows, leaving the band at 17.23 s and spanning 25.4° to 44.6° by 30 s, as listed in Table 10.4. Doubling either gain alone stops the swing. With the reference tuning the sideslip stays within 32.2° to 36.9° and is back within 2° for good at 10.76 s, and the LQR on the same step keeps 34.9° to 38.9°.

Table 10.4 Measured Results at the Default Seed

Case	Law	In band (s)	Cones	Outcome
1	by hand, untouched	—	1	SPUN 4.29 s
	nested, Auto	19.99	0	held
2	grip, 30°	—	2	SPUN 3.61 s
4, 30 s	nested, loose	24.0	0	swing 25.4–44.6°
	nested, reference	—	0	32.2–36.9°
5	LQR, dry	19.7	0	±2° in 0.92 s
	nested, reference	18.5	4	peak 42.1°, 3.74 s
	nested, loose	11.0	4	peak 47.5°
6, 30 s	LQR, dry schedule	3.9	3	SPUN 10.06 s
	LQR, $\hat{\mu}$ schedule	30.0	0	32.8–39.9°
	nested, reference	12.0	—	on the pad
7	LQR, $\hat{\mu}$, path	16.6	0	2 transitions
	nested, reference	10.1	5	1 transition
	LQR, path off	17.1	3	2 transitions
8	LQR, path off	—	4	ρ 6.08–11.19 m
	LQR, path on	—	0	ρ 8.81–8.82 m
9	LQR (front drive)	—	—	OFF PAD 3.18 s
	grip (front drive)	—	—	GRIPPED OUT 2.30 s
	LQR, rear drive	20.0	0	held

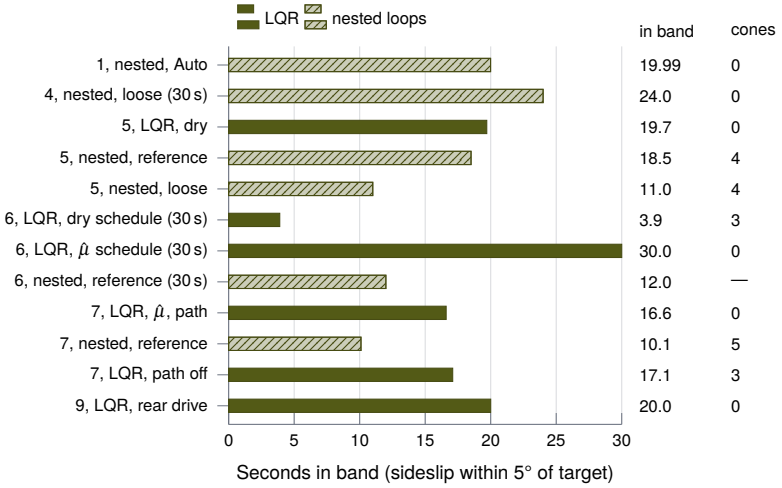


Figure 10.3 The rows of Table 10.4 that report time in band, with their cones. Solid bars are LQR runs, hatched bars nested-loop runs. Runs last 20 s except those marked 30 s. Notice that on the wet patch of case 6 the same LQR holds the band for all 30 s when its schedule follows $\hat{\mu}$ and for only 3.9 s on the dry schedule.

10.6.3 LQR, a Kick and the Wet Patch

The LQR of case 5 brings the sideslip back within 2° of 35° 0.92 s after the 10° kick and never overshoots, reaching 35.04° at most. The reference nested loops overshoot to 42.1°, take 3.74 s and knock four cones. The LQR moves steer and throttle together the moment the kick arrives, because its gain couples every state to both inputs, while each nested loop waits for its own variable to move. In case 6 the wet quarter is on from the start. The LQR on the dry pad's schedule spins the car at 10.06 s. Scheduled on $\hat{\mu}$, the equilibrium and the gain follow the friction, and the car holds 32.8° to 39.9° and 8.62 to 10.10 m from the cluster for all 30 s. The estimate falls to 0.59 on the patch, stays within 0.016 of the true friction on average while the rear slides, and is up to 0.30 off for a moment at the patch's edges, where the friction steps. Without the path loop the fixed design survives but

Table 10.5 The Challenge Over 60 s, Default Seed

Configuration	End	Score
nested, loose, no path loop (shipped)	SPUN 32.54 s	0
LQR, $\hat{\mu}$ schedule, path loop (suggested)	60 s, 2 transitions	58
LQR, dry schedule, path loop	SPUN 37.41 s	0
LQR, $\hat{\mu}$ schedule, no path loop	OFF PAD 46.98 s	0
nested, reference, path loop	60 s, 11 cones	0
nested, reference, no path loop	60 s, 7 cones	0

wanders and knocks eight cones.

> **TRY IT IN THE APP** **6 · The Wet Patch**
>
> Press Run and watch the car each time it crosses the wet quarter. On the dry pad's schedule it spins at 10.06 s. Press Reset, set Friction in the schedule to Estimated $\hat{\mu}$ and run again. The sideslip now stays between 32.8 and 39.9° for all 30 s, while $\hat{\mu}$ falls to 0.59 on the patch.

10.6.4 The Figure-Eight and the Path Loop

Case 7 flicks at 8.37 s and 15.81 s, and each reversal takes 1.13 s and 1.14 s from the flick to the handover. The run completes both transitions, keeps the band for 16.6 of the 20 seconds and touches no cone. The rear tire leaves its friction circle in each reversal, its force falling to 78 % of the limit in the first and 84 % in the second, as shown in Figure 10.4. While the throttle is lifted and the lock is opposite, the rear slip angle swings through zero, from −23.3° at 8.80 s to 2.0° at 9.00 s, and for an instant the rear grips. That instant is what lets the nose cross to the other side. The reference nested loops complete one of the two reversals and knock five cones. In case 8 the car starts on a circle 2.5 m off center. Without the path loop it passes between 6.08 and 11.19 m from the cluster and knocks four cones. With it the distance is within 0.5 m of 9 m from 2.08 s and between 8.81 and 8.82 m from 15 s.

10.6.5 Front Drive and the Challenge

With front drive (case 9) the solver finds no equilibrium, and the LQR falls back to the nested loops. They floor the throttle, the front plows, the rear stops sliding (GRIPPED OUT at 2.29 s) and the car slides wide off the pad at 3.18 s. The grip controller grips out at 2.30 s. No held pair of steer and throttle keeps the rear sliding near 35° either. Of 187 pairs, 147 leave the pad, 39 grip out and one runs on with no time in the band. The same car with rear drive, from the same Reset, holds the band for all 20 s.

> **DID YOU KNOW?** **Front Drive Cannot Hold This Drift**
>
> From the simulator: with front drive (case 9) the solver finds no drift, and the LQR's fallback leaves the pad at 3.18 s. The same car with rear drive holds the band for all 20.0 s (Table 10.4). Which axle is driven decides whether the drift exists at all.

The challenge asks for 60 s of donut and figure-eight with the wet quarter from 25 s, scoring the seconds in band plus the transitions, and zero on a spin or a cone. As shipped, the loose nested loops hold the band until the wet quarter reaches the car. They leave it at 25.28 s, the sideslip peaks at 50.9°, the first cone falls at 27.15 s, and the car spins at 32.54 s, before the figure-eight begins. The suggested settings, the LQR scheduled on $\hat{\mu}$ with the path loop, hold the band for 56.2 s, complete both transitions and touch no cone, a score of 58. Table 10.5 shows that each part is needed. Over seeds 1 to 20 the shipped settings spin between 32.37 s and 32.58 s every time, and the suggested ones pass every time with 56.09 to 56.42 s in the band. At five times real speed the suggested 60 s run plays in 12.1 s of wall time in a desktop browser.

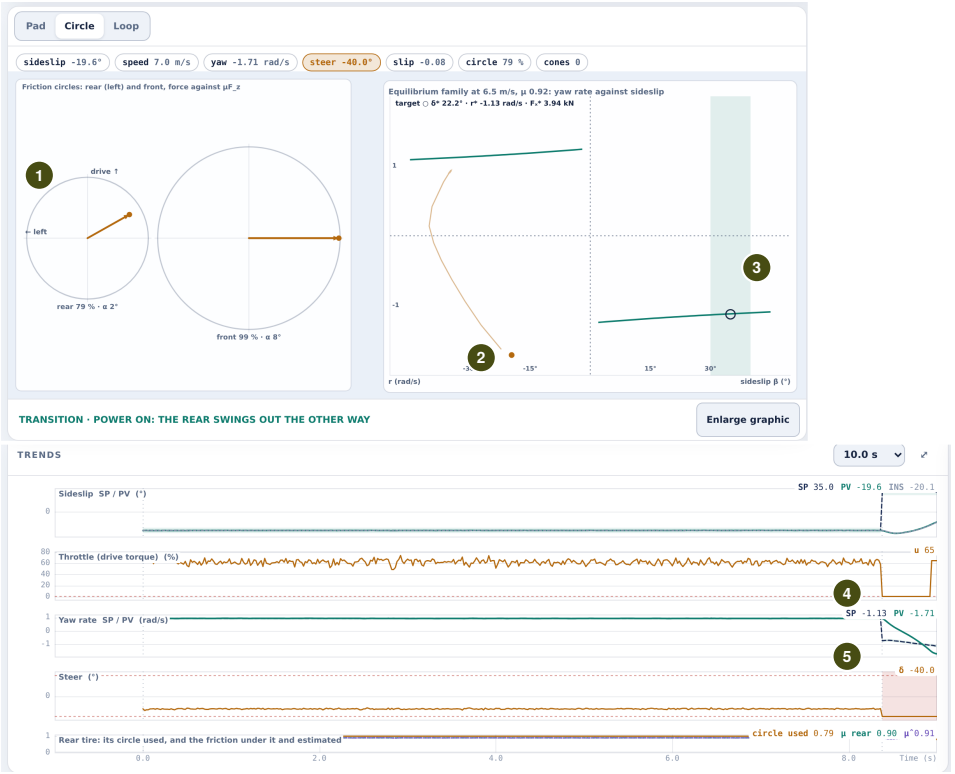


Figure 10.4 The first reversal of the figure-eight (case 7) at $t = 9.00\,\mathrm{s}$ on the Circle view, under the LQR scheduled on $\hat{\mu}$ with the path loop. The picture joins two parts of the app's screen, the plant view above the trends. (1) The rear tire's force has fallen inside its friction circle, to 79 % at a slip angle of 2°, while the front works at 99 %. (2) The state (dot, with its trail) crosses from the left-hand branch of the equilibrium family toward (3) the right-hand target, computed at the path loop's speed target of 6.46 m/s (the plot's title rounds it) and the schedule's friction 0.92 ($\delta^* = 22.2°$, $r^* = -1.13\,\mathrm{rad/s}$, $F_{xr}^* = 3.94\,\mathrm{kN}$). The badges read −19.6° of sideslip, 7.0 m/s, −1.71 rad/s and full opposite lock, −40°, with (4) the throttle back on at 65 % in the power phase. The 10 s trend window shows 35° held from the start, the lift at the flick (8.37 s) and (5) the yaw rate reversing.

10.7 Technology in Practice

What Measures It

A drift controller needs the sideslip and the yaw rate. The yaw rate comes from a gyro, and in common cars the gyro and the lateral accelerometer are already on board for the stability system [16]. The sideslip is harder. Production systems estimate it with extended or unscented Kalman filters, the industrial state of the art [17], while drifting test cars measure it with a satellite-aided inertial unit, as the chapter's packet does with its navigation unit, gyro, accelerometer and encoders.

> **TECHNOLOGY Sensor: Satellite-Aided Inertial Navigation for Sideslip**
>
> Sideslip has no inexpensive direct measurement. An inertial navigation system aided by satellite positioning gives position to the centimeter and measures the slip angle directly [16], and one drifting test car read its state from such a unit at 100 Hz [18]. Trees, tall buildings, mountains and tunnels interrupt the satellite signal, and the sideslip then turns unstable and unreliable [16]. An optical speed sensor gave one test car its reference sideslip, to within $\pm 0.2^\circ$ [17]. The chapter's navigation unit reads it to $\pm 0.15^\circ$.

What Moves It

The two handles are the steer and the drive torque. Experimental controllers that drift along a path ran on a research car whose independent electric motors acted like a fully locked differential, a layout uncommon in standard cars [19]. A production car with a combustion engine couples its rear wheels through a differential and delivers engine torque with a delay, so it cannot set each wheel's speed [18]. The chapter's drive is the research car's kind: an electric torque source on one lumped axle, limited to 1860 N m and 120 kW. Its steer servo follows its command with a 30 ms lag, limited by its rate and its lock.

> **TECHNOLOGY Final Element: A Steering Actuator at Its Limit**
>
> On one production car the electric power steering motor, built for low-torque assistance such as automatic lane changes, gives at most about 30 N m. In an automated drift the steer stayed on its limit and the car spun, most likely because the motor could not overcome the front tires' self-aligning moment, so a driver steered while the controller set the drive torque [19]. On another production car the steering lagged its command by a mean phase delay of 140 ms, and sideslip targets above 0.65 rad were unreliable because the steer command regularly passed the 34° lock, leading to spin-outs [18]. The chapter's servo has a 30 ms lag, a 400 °/s rate and a 40° lock.

Where the Controller Runs

On one test car the controller's generated code ran on a rapid-prototyping computer that also bridged the vehicle's data buses, where measured and estimated states arrived every 10 ms. A nonlinear predictive controller solved its problem at 50 Hz over 25 steps and sent its inputs through the car's driving-assistance interface, which safety features limited [19]. On another, an onboard computer ran the drift controller at a fixed 100 Hz and commanded the production actuators over the vehicle bus [18]. The chapter's laws run every 10 ms, the same 100 Hz, on integration steps of 1 ms. The block that holds its structure is the schedule: a table of LQR gains, each designed at the exact equilibrium of its node and interpolated at the target. Figure 10.5 draws the loop on test-car hardware.

What Is Modernizing It

Three developments carry the chapter's ideas toward production cars. The first is predictive control on a standard passenger car. A nonlinear predictive controller tracking equilibria computed offline brought such a car to 30° of sideslip and held it for several rounds on a dry surface and on a watered skidpad, with a driver steering and the controller setting the drive torque [19]. The second is latency compensation: a state predictor that phase-aligns steer and torque let a combustion-engined sports car hold circular and figure-eight drifts [18]. The third is learning. Agents trained in simulation discover drifting strategies without detailed tire models, but most results remain in simulation because transfer to the real car, robustness to model mismatch and training data are still hard, and a learned diffusion model that needs little data runs on a high-performance processor [18]. The chapter's friction schedule is the simplest form of the adaptation all three point to.

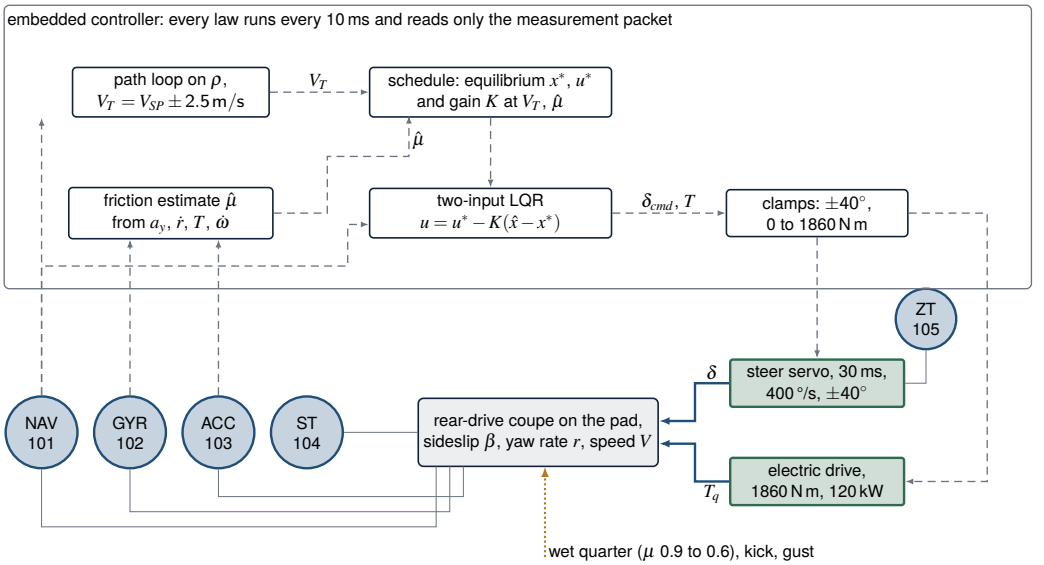


Figure 10.5 The drift loop on test-car hardware. The navigation unit NAV-101 gives the speed, sideslip and pose, the gyro GYR-102 and accelerometer ACC-103 the yaw rate and lateral acceleration, and the encoders ST-104 and ZT-105 the wheel speed and steer. The friction estimate and the path loop's speed target pick the equilibrium and the gain from the schedule, and the LQR sends both inputs, clamped, to the steer servo and the rear electric drive. Notice that one law moves both actuators at once, and that the friction, which no instrument reads, reaches it only through $\hat{\mu}$.

> **IN PRACTICE** **How a Stability System Proves It Stops a Spin**
>
> The 2007 rule adds a performance test. In the sine with dwell maneuver the car coasts at an initial 50 miles per hour while a steering machine turns the wheel through a fixed pattern, repeated at rising steering amplitudes, a maneuver severe enough to spin most cars without the system. The car passes if its yaw rate falls to no more than 35 percent of its peak one second after the wheel returns to straight ahead, and to 20 percent after 1.75 seconds. A responsiveness criterion keeps a car that barely turns from passing [4]. The test checks that a car stops rotating, the opposite of the slide this chapter holds on purpose.

10.8 Limitations

The car is a single track with brush tires at constant friction, so it cannot represent the inside rear wheel unloading, a differential, a tire's fall from peak to sliding friction, or camber. The combined-slip force points along the stiffness-weighted slip rather than the sliding velocity, 6.37° apart at the design point. At the same slip the drive force would be 3302 N instead of 3716 N, so the slip ratio, the linearization and the closed loop would move while the equilibria would not. The wet patch is a step in friction, which the estimate follows with errors up to 0.30 at its edges. The drive is an ideal electric torque source. The reversal is a fixed open-loop sequence tuned for the declared car at 35°, not a planned trajectory, and the challenge's thresholds and the loss rules are teaching values.

The controller's equilibrium is started cold at each target, and at low speed and small sideslip the equations also admit a mirror-image root that turns the other way. The solver accepts only a root whose yaw rate turns the commanded way, and the controller treats a solution past the 40° lock as no equilibrium. Over learner-settable targets (4 m/s to 12 m/s, 1° to 45°) it finds the commanded drift at all 1485 on the dry pad, and on the wet pad at all but five at 45° and 11 m/s or above, whose steer is past the lock. Of the 9090 targets the shipped cases can reach (25° to 45°, the path loop's 5.5 m/s to

10.5 m/s and the schedule's friction from 0.58 to 0.92), only three corner targets at 45° change, their steer a hair past the lock. No shipped run visits them, so every run reported here is unchanged. The estimator meets only the friction, not a gust or a load transfer, and the gain schedule assumes the declared mass, inertia and stiffnesses.

OPEN CHALLENGE Holding the Slide Through Slow Actuators

Path-following drift control was first validated on a purpose-built electric research car with near-instant torque and independently driven rear wheels. In simulation, added actuator latency set the sideslip oscillating, and above 95 ms the oscillations became catastrophic. A predictor that runs the model ahead over the mean latency let a production sports car hold circles and figure-eights despite actuator delays over 250 ms. It still lacks a formal proof of closed-loop stability, it was tested only on circles and figure-eights, and tire model fidelity and friction estimation still limit it [18]. Progress would be a delay-tolerant drift law with a proof, shown on arbitrary paths. The chapter's throttle lag and steer delay let a reader meet the same problem.

10.9 Conclusion

A drifting car is held at an equilibrium it leaves by itself, through a tire that has no lateral force left to give, and this simulator lets a learner see both facts. Started on the 35° drift at 8 m/s with its inputs held, the car is 5° off in 2.98 s, at the rate of the slide's root of 1.95/s, and the speed adds a second root of 0.21/s that an analysis at fixed speed does not see. A grip controller spins it at every request of 10° or more. Two driver-like loops hold it until a speed change sets them fighting, while an LQR on the two-input linearization recovers a 10° kick in 0.92 s without overshoot and, scheduled on a friction estimate, holds the drift across a wet patch that spins the fixed design at 10.06 s. A path loop that trims the speed holds the circle within 0.5 m, each figure-eight reversal takes the rear off its friction circle for an instant, and the same car with front drive has no saturated-rear equilibrium when both axles see the same friction. The tire, the drive and the reversal are idealized. A fuller simulator would add the two rear wheels with a differential, a tire whose friction falls as it slides, a reversal planned as a trajectory between equilibria, and an estimator for the vehicle's mass and stiffness. The drift shows that a saturated actuator still leaves a handle when another input shares its limit.

KEY POINTS

- A drifting car is held at a point it leaves by itself, through rear tires that have no sideways force left to give.
- Started on the 35° drift at 8 m/s with its inputs held, the car is 5° off in 2.98 s, at the rate of the slide's root of 1.95/s.
- A grip controller spins the car at every request of 10° or more, and two driver-like loops hold it until a speed change sets them fighting.
- An LQR on the two-input linearization recovers a 10° kick in 0.92 s and, scheduled on a friction estimate, holds the drift across a wet patch.

Where else this applies. Chapter 4 also works a tire at the peak of its force curve, and Chapter 2 holds a reactor at a steady state it leaves by itself.

References

[1] G. Stein, "Respect the unstable," *IEEE Control Systems*, vol. 23, no. 4, pp. 12–25, 2003.

[2] E. Ono, S. Hosoe, H. D. Tuan, and S. Doi, "Bifurcation in vehicle dynamics and robust front wheel steering control," *IEEE Transactions on Control Systems Technology*, vol. 6, no. 3, pp. 412–420, 1998.

[3] J. Edelmann and M. Plöchl, "Handling characteristics and stability of the steady-state powerslide motion of an automobile," *Regular and Chaotic Dynamics*, vol. 14, no. 6, pp. 682–692, 2009.

[4] National Highway Traffic Safety Administration, "Federal motor vehicle safety standards; electronic stability control systems; controls and displays, final rule, docket no. NHTSA-2007-27662," Federal Register, vol. 72, no. 66, pp. 17236–17322, 6 April 2007, 2007. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2007-04-06/pdf/07-1649.pdf

[5] C. Voser, R. Y. Hindiyeh, and J. C. Gerdes, "Analysis and control of high sideslip manoeuvres," *Vehicle System Dynamics*, vol. 48, no. sup1, pp. 317–336, 2010.

[6] E. Velenis, D. Katzourakis, E. Frazzoli, P. Tsiotras, and R. Happee, "Steady-state drifting stabilization of RWD vehicles," *Control Engineering Practice*, vol. 19, no. 11, pp. 1363–1376, 2011.

[7] R. Y. Hindiyeh and J. C. Gerdes, "A controller framework for autonomous drifting: Design, stability, and experimental validation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 136, no. 5, p. 051015, 2014.

[8] J. Y. Goh, T. Goel, and J. C. Gerdes, "Toward automated vehicle control beyond the stability limits: Drifting along a general path," *Journal of Dynamic Systems, Measurement, and Control*, vol. 142, no. 2, p. 021004, 2020.

[9] L. Segel, "Theoretical prediction and experimental substantiation of the response of the automobile to steering control," *Proceedings of the Institution of Mechanical Engineers: Automobile Division*, vol. 10, no. 1, pp. 310–330, 1956.

[10] R. Rajamani, *Vehicle Dynamics and Control*, ser. Mechanical Engineering Series. Springer US, 2012.

[11] H. B. Pacejka, "Theory of steady-state slip force and moment generation," in *Tire and Vehicle Dynamics*. Elsevier, 2012, pp. 87–147.

[12] A. T. van Zanten, "Bosch ESP systems: 5 years of experience," SAE International, SAE Technical Paper 2000-01-1633, 2000.

[13] J. Y. Goh and J. C. Gerdes, "Simultaneous stabilization and tracking of basic automobile drifting trajectories," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 597–602.

[14] J.-O. Hahn, R. Rajamani, and L. Alexander, "GPS-based real-time identification of tire-road friction coefficient," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 3, pp. 331–343, 2002.

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

[16] A. Leanza, G. Reina, and J.-L. Blanco-Claraco, "A factor-graph-based approach to vehicle sideslip angle estimation," *Sensors*, vol. 21, no. 16, p. 5409, 2021.

[17] A. Bertipaglia, D. de Mol, M. Alirezaei, R. Happee, and B. Shyrokau, "Model-based vs data-driven estimation of vehicle sideslip angle and benefits of tyre force measurements," 15th International Symposium on Advanced Vehicle Control (AVEC), Kanagawa, 2022; arXiv preprint 2206.15119, 2022. [Online]. Available: https://arxiv.org/abs/2206.15119

[18] F. Werner, T. Heintzenberg, M. Lienkamp, and J. Betz, "Drifting in the future: Stabilizing path following drifting on high-latency vehicle systems," in *2026 IEEE International Conference on Robotics and Automation (ICRA)*, 2026, pp. 12 763–12 770.

[19] S. Meijer, A. Bertipaglia, and B. Shyrokau, "A nonlinear model predictive control for automated drifting with a standard passenger vehicle," in *2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, 2024, pp. 284–289.

10.10 Exercises

10.1 ○ Concept check · 2 min

In the drift of case 3 the rear tire slips at −42.4°, far past its sliding angle of 5.48°, so its force sits on its friction circle. How does the drive torque still move the rear's lateral force?

(a) More torque moves load onto the rear axle, so its friction circle grows and its lateral force grows with it

(b) More torque raises the rear slip angle, and the brush curve's lateral force keeps rising past the sliding angle

(c) The resultant stays on the circle, so more drive force leaves less lateral force and the rear swings further out

(d) It does not: torque changes only the speed, and the lateral force answers only to the steer through the front

10.2 ○ Concept check · 2 min

In case 2 the grip controller, asked for 30° of sideslip, spins the car at 3.61 s, and it spins at every request of 10° or more. What is the reason?

(a) Its integral terms wind up against the steer lock before the sideslip builds, and the stored error throws the car past 70°

(b) Its 10 ms control period is too slow for the slide's root of 1.95/s, so any law at that period loses the car

(c) Its speed loop lifts the throttle as the car slows in the turn, so the rear grips again and snaps the nose around

(d) It steers into the turn for more yaw rate, which assumes a rear force that rises with slip, and a sliding rear cannot check the yaw

10.3 Concept check · 1 min

In case 9 the same car drives its front axle, and on equal friction the solver finds no drift at any target from 5 to 12 m/s and 5° to 45°. Why?

(a) With the rear saturated and undriven, the yaw balance takes the front's whole friction circle for lateral force and leaves none to drive

(b) The front's cornering stiffness of 120 kN/rad is too low for any countersteer to balance the yaw moment of the sliding rear

(c) The LQR's gain schedule has no nodes for front drive, so the equilibrium solver is never called with the front axle driven

(d) The front's drive force pulls the nose toward the velocity, which limits any front-drive drift to the rear's sliding angle

10.4 Theory and calculation · 6 min

At the design point the car drifts on the dry pad ($\mu = 0.9$) at $V = 8\,\mathrm{m/s}$ with $|\beta| = 35°$ and $r^* = 0.927\,\mathrm{rad/s}$. Use $m = 1450\,\mathrm{kg}$, $a = 1.0\,\mathrm{m}$, $b = 1.5\,\mathrm{m}$ and $g = 9.81\,\mathrm{m/s^2}$.

(a) Find the static axle loads and the radius μF_{zr} of the rear friction circle. *Checkpoint:* 8.53 kN, 5.69 kN and 5.12 kN.

(b) In magnitude, the lateral and yaw balances of Equation (10.6) read $F_{yf}\cos\delta + F_{yr} = mrU$ and $aF_{yf}\cos\delta = bF_{yr}$. Find U, the rear lateral force and $F_{yf}\cos\delta$. *Checkpoint:* 6.55 m/s, 3.52 kN and 5.29 kN.

(c) Put the rear force on its circle and find the drive force F_{xr}^*. Then find the drive power from $T^* = 1152\,\mathrm{N\,m}$ and $\omega^* = 38.16\,\mathrm{rad/s}$. *Checkpoint:* 3.72 kN and 44.0 kW.

10.5 Theory and calculation · 6 min

The linearization at the design point has the unstable roots $p_1 = 1.955/\mathrm{s}$ and $p_2 = 0.206/\mathrm{s}$.

(a) Find the doubling time of each mode, and the time the slide's mode takes to grow an error from 0.5° to 5°. *Checkpoint:* 0.355 s and 3.36 s, then 1.18 s.

(b) In case 3 the sideslip error passes 0.5° at 1.82 s and 5° at 2.98 s. Find the growth rate these give, and say which root it matches. *Checkpoint:* 1.98/s, the slide's.

(c) Section 1.4.7 states that feedback on an unstable mode through a delay τ_d is stable for some gain only if $p\tau_d < 1$, and that a loop around an unstable pole needs a bandwidth above about $2p$. Find the longest delay the slide allows, $p_1\tau_d$ at the app's largest steer delay of 200 ms, and the bandwidth floor. *Checkpoint:* 0.51 s, 0.39 and 3.9 rad/s.

10.6 Theory and calculation · 6 min

Equation (10.7) bounds the sideslip of a saturated-rear drift on front drive.

(a) Show that with static loads the front's lateral demand $(b/a)\mu_R F_{zr}$ equals $\mu_R F_{zf}$, and evaluate it on the dry pad. *Checkpoint:* both 7.68 kN.

(b) Evaluate the bound for $\mu_F = 0.9$ on $\mu_R = 0.6$, and for equal friction. *Checkpoint:* 33.9° and 0°.

(c) A designer moves the center of mass to mid-wheelbase, $a = b = 1.25\,\text{m}$, keeping $\mu_F = 0.9$ and $\mu_R = 0.6$. Find the largest saturated-rear drift, and say whether a front-drive drift at 35° is possible on either car. *Checkpoint:* 29.2°; no.

10.7 App challenge · 8 min

Open case 3, "The equilibrium, with no controller", at the default seed and 1× speed. It starts in Manual with Steer and Throttle at the drift's values to the stations' resolution.

(a) Before pressing Run, read δ^*, r^* and F_x^* on the Circle view, and the unstable roots on the Loop view. *Checkpoint:* −25.0°, 0.93 rad/s, 3.72 kN; +1.95 and +0.21/s.

(b) Press Run and let the app pause itself. Read the Run tile. From the exported CSV column `beta_deg`, find the first rows at which $|\beta|$ differs from 35° by more than 0.5° and by more than 5°, and the growth rate the two times give. *Checkpoint:* SPUN at 4.29 s; 1.84 s and 3.00 s (within 0.02 s); about 1.98/s.

(c) Set Target sideslip to 25°, press Put the car on its drift equilibrium, press Reset and Run. When is the sideslip 5° from 25°? *Checkpoint:* about 3.16 s.

10.8 App challenge · 7 min

Open case 5, "LQR about the equilibrium, and a kick", at the default seed and 1× speed. A 10° kick arrives at 8 s.

(a) Run the 20 s as shipped. Read the In the band and Cones hit tiles. From the CSV column `beta_deg` after 8 s, find the largest $|\beta|$ and the time after the kick from which $|\beta|$ stays within 2° of 35°. *Checkpoint:* 19.7 s, 0; 35.0°; 0.92 s (within 0.05 s).

(b) Press Reset, set Controller to Nested, and run the 20 s again. Read the same quantities. *Checkpoint:* 18.5 s, 4; 42.1°; 3.76 s (within 0.05 s).

(c) Explain why the LQR recovers sooner and without overshoot. *Checkpoint:* its gain moves both actuators at once.

10.9 App challenge · 8 min

Open the challenge, "Donut and figure-eight", at the default seed and 5× speed. Before each run after the first, press Reset and change only the setting named, keeping the earlier changes. Run until the app pauses itself or 60 s pass.

(a) Run it as shipped. Read the Run, Cones hit and Score tiles. *Checkpoint:* SPUN at 32.54 s, 6 cones, score 0.

(b) Set Controller to LQR. The schedule already reads Estimated $\hat{\mu}$. Read the clock when the app pauses, and the Run, Cones hit and Transitions tiles. *Checkpoint:* about 46.98 s, OFF PAD, 7 cones, 2 of 2.

(c) Set Path loop to On, and read the Score tile and the Challenge target. *Checkpoint:* 58 (56.2 s and 2 transitions), 0 cones, met.

(d) Set Friction in the schedule to Dry pad, and read the Run tile. Which weakness of the shipped settings did each of the three changes address? *Checkpoint:* SPUN at 37.41 s.

10.10 Application area · 3 min

A team wants to start the tuning of a real rear-drive drift car with an open differential from the simulator's scheduled LQR.

(a) Name two assumptions of the model that bear most on whether the equilibrium carries over to that car, and say which way each would move it. *Checkpoint:* single track with no lateral load transfer or differential; sliding friction that does not fall below the peak.

(b) Which quantity in the simulator would you compare against a first test on the real car before trusting the gain? *Checkpoint:* the equilibrium: steer, yaw rate and drive at a steady drift.

10.11 Application area · 3 min

The friction estimate of Equation (10.11) updates only while the rear's slip speed exceeds 1 m/s.

(a) Why does it hold its value while the rear grips, and what does that mean for a road car whose stability control keeps every tire gripping? *Checkpoint:* on grip the force is inside the circle, so the estimate cannot learn μ.

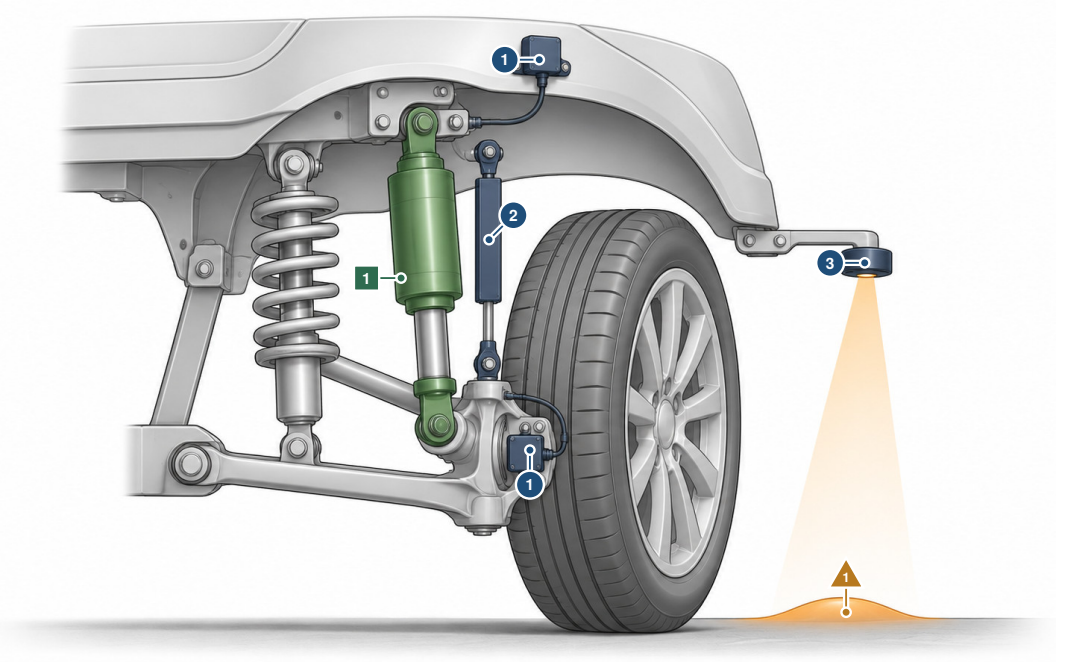
(b) At the wet patch's edges the estimate is up to 0.30 off for a moment. What does a schedule on $\hat{\mu}$ do during that moment? *Checkpoint:* it uses the wrong equilibrium and gain until the estimate catches up.

10.12 Review the agent · 5 min

An AI assistant reviews the LQR of case 6 for a test track whose wet quarter has $\mu = 0.6$. It writes: "The LQR keeps at least 60° of phase margin and tolerates a gain reduction to one half. A friction change from 0.9 to 0.6 is well inside those margins, so the dry pad's design will hold the drift on the wet quarter, and scheduling on $\hat{\mu}$ adds complexity for no benefit." Find the flaw, and say how you would check the claim on the simulator.

CHAPTER 11

Active Suspension Simulator for Comfort and Road Holding Through One Strut Force



The controller moves 1 the force actuator in the strut, which pushes the body up and the wheel down

The controller reads 1 the accelerometers on the body and on the wheel hub

2 the ride-height sensor across the strut's travel 3 the preview sensor that reads the road ahead of the tire

What disturbs it 1 the bump in the road ahead

Notice that the actuator joins the body to the wheel, so every push it gives one it also gives the other, while the sensor ahead sees the bump before the tire reaches it. Illustration, not to scale.

IN PLAIN TERMS **One Push Between the Body and the Wheel**

The main idea. A car body should stay still for the people inside, while the wheel under it must follow the road to keep its grip. A controlled strut between them can push the body up, but the same push drives the wheel down.

The challenge. At the wheel's own bouncing rhythm, a strut tuned for comfort calms the body mainly by letting the tire leave the road. A tire can only push on the road, and a plain damper can only soak up motion.

Why it matters. A tire in the air cannot brake or steer, so a smooth ride can hide a car that is losing its grip.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact in the Introduction (Section 11.1), the two history segments in Section 11.2, the fact and the try-its in Section 11.6, 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 11.6), Technology in Practice (Section 11.7), Limitations (Section 11.8) and the open challenges

The control question One strut force pushes the body up and the wheel down, and near wheel hop the tire nearly fixes the body acceleration. How much comfort can a passive damper, a semi-active damper and a force actuator each buy before the wheel leaves the road or the strut reaches its stop?

Learning objectives

- Derive the constraint no strut force changes, and compute the two modes and two invariant points of the quarter car.
- Convert a road pattern in distance into a frequency and place it against the body mode and wheel hop.
- Explain why a semi-active damper can only dissipate, and how its valve follows a skyhook and groundhook request.
- Compare comfort, tire load, travel and actuator work across the struts, with and without road preview, against a passive twin.
- Explain why load leveling must stay slow and why actuator slew decides road holding at wheel hop.

Prerequisites Sections 1.4.4 (modes and eigenvalues), 1.4.8 (resonance), 1.4.7 (lag and delay), 1.6.2 (velocity-form PID) and 1.6.1 (engine), with 1.6.3 for the predictive force law.

The app https://apmonitor.com/apps/active-suspension/

See also Volume 20, Chapter 12 for preview read from a buffer keyed by travel distance. Volume 20, Chapter 5 for a ride height held through waves with a look-ahead, and Volume 20, Chapter 2 for a vehicle loop whose behavior changes with speed.

11.1 Introduction

A suspension stands between something that must be kept still and something that must be followed. A car body carries passengers who feel its acceleration, while the wheel beneath it must keep its tire pressed on the road, because the grip available for braking and steering rests on the tire load. The quarter car, one body and one wheel joined by a spring and a strut and standing on a tire, is the standard model for weighing passenger discomfort, suspension working space and tire load variation against one another [1, 2]. A controlled strut adds a force between the two masses, and the whole difficulty is that the same force pushes the body up and the wheel down.

The equations of the two masses, added together, lose the strut force entirely. What remains is a constraint that no strut can change. It fixes the body acceleration at one frequency and the suspension deflection at another, so that only one of the three transfer functions of the quarter car can be chosen freely [3]. The first of these frequencies lies close to wheel hop, the resonance of the wheel on its tire, which is 11.9 Hz in the simulator and which a washboard of 1.25 m ripples reaches at 15 m/s. There a strut tuned for comfort leaves the wheel free to bounce, and it lowers the body acceleration mainly by letting the tire leave the road.

The simulator addresses the belief that a better controller improves every measure of a suspension at once, and the habit of judging a suspension by one comfort number. The class of problem is general. A truck cab on its chassis and a railcar body on its bogie each place a suspension force between a payload and an intermediate mass that follows a moving base, and each inherits the same constraint, because the constraint equations of interconnected subsystems do not depend on the nature of the interconnection [3]. The semi-active force generator itself was first proposed for vibration control in general rather than for cars alone [4].

DID YOU KNOW? Semi-Active Damping Well Beyond Cars

A survey of semi-active suspension control lists applications well beyond the car: the cabins of trucks and tractors, seats, trains, washing machines, buildings and bridges, and artificial legs. In each the semi-active device obeys the passivity constraint of the damper in this chapter: it can only dissipate energy [5].

However, the quarter car usually reaches a control course as a linear transfer function, which shows the two modes and hides three facts that decide the outcome on a real road. The tire can push on the road but never pull, a damper can take energy out but never put it in [1], and feedback cannot answer a bump before the wheel reaches it, which is the gap preview control fills [6]. A comfort number read on its own can hide a wheel that spent 1.62 s of 5 s in the air, as case 3 shows.

This chapter establishes how much comfort a passive damper, a semi-active damper and a force actuator can each buy before the wheel leaves the road or the strut reaches its stop. It works with a quarter-car simulator in which all three, driven by one control law, are compared at every instant with a passive twin on the same road. The quarter car, skyhook and groundhook damping, preview feedforward and the invariant points are all established material. The simulator is an eleven-state model that treats contact loss and bump-stop contact as counted events, integrates the passive twin beside the controlled car, keeps the semi-active damper dissipative by construction and reads the road ahead from a noisy buffer indexed by distance. Its passive response agrees with the linear transfer function to within 0.1 % at sixteen speeds. Ten scenarios each change one road, one piece of hardware or one setting.

Section 11.2 reviews suspension control from fixed dampers to road preview. Section 11.3 gives the quarter car, its tire, bump stops and roads, and Section 11.4 derives the constraint that no strut force can change, its two invariant points and the limits a pothole edge sets. Section 11.5 describes the force request, the three struts, the load-leveling loop and, in Section 11.5.7, a predictive force law added beside the original one. Section 11.6 reports the measured scenarios, Section 11.7 the sensors, dampers, actuators and control units of real suspensions, Section 11.8 the limitations and Section 11.9 the conclusions. The exercises of Section 11.10 work through the scenarios.

11.2 Suspension Control From Fixed Dampers to Road Preview

A passive spring and damper fix the compromise between ride and road holding when the car is built, and reviews of road vehicle suspension design start from that compromise [2]. Judging a design needs a model of the road as well as of the car. A road profile is described statistically by a displacement spectral density that falls with wavenumber [7], the form behind the ISO 8608 roughness classes, so that a suspension can be assessed on a random road rather than on a single bump.

The first history box below tells how road preview was proposed for suspensions and then solved as an optimal control problem.

HISTORY Reading the Road Ahead, 1968 and 1976

Preview control, in which the input is sensed before it reaches the plant, was proposed for vehicle suspensions in 1968. That study optimized the trade-off between vibration and the clearance needed for suspension travel on statistically described surfaces. Preview could lower the rms vibration by as much as a factor of sixteen at constant clearance, or the clearance by a factor of 223 at constant vibration [6]. The problem was solved again in 1976 as a discrete optimal control problem, with a controller structure suited to practical use [8]. Optimal linear state feedback for a fully active suspension appeared in the same year, and it needed only the body's absolute velocity and its displacements relative to the axle and the road [9].

A cheaper route than a powered actuator followed in 1974, and its ideal law, a damper from the body to a fixed reference in the sky, is the skyhook term of this simulator, as the second history box on the semi-active damper tells.

Table 11.1 Signals, States and Their Roles

Symbol	Role	Nominal	Range
$\ddot{z}_s$	CV, body acceleration	SP 0	—
x	CV, ride height	SP 0 mm	SP ± 40
F_t	output, tire load	3924 N	≥ 0
F_{req}	MV, strut force request	0 N	$\pm F_{av}$
z_s, $\dot{z}_s$, z_u, $\dot{z}_u$	states, body and wheel	0	—
F_a, c_d	states, actuator, damper	0, 500 N s/m	± 2500, 300–4500
$\tilde{z}_s$, $\dot{\tilde{z}}_s$, $\tilde{z}_u$, $\dot{\tilde{z}}_u$	states, passive twin	0	—
s	state, distance	0 m	≥ 0
v	DV, speed	12 m/s	1–35
m_{load}	DV, load on the corner	60 kg	0–250
$z_r(s)$	DV, road (six choices)	humps	—
σ_p	DV, preview noise (1σ)	0.5 mm	0–30
ε_w	DV, damper wear	0 %	0–90
d	DV, actuator derating	0 %	0–90
F_p	DV, bounce-test push	−800 N, 0.4 s	event
—	DV, bump 4 m ahead	60 mm, 3 m	event

HISTORY A Damper That Only Dissipates, 1974 to 1986

The cheaper route of 1974 was a force generator that responds to general feedback signals from a vibrating system yet needs no servomechanism power supply. In computer simulation it performed comparably to fully active vibration control [4]. A 1986 comparison of passive, active and semi-active cars on the quarter-car model derived the semi-active laws from the active ones. Able only to dissipate energy, the semi-active damper switched off wherever the active system would have supplied energy [1], and the theoretical limitations of active vehicle suspensions were examined in the same year [10]. A later survey calls this saturation of the requested force to what the damper can deliver the clipped approach, and finds it very widespread in semi-active control [5].

Constraint equations derived in 1990 showed that the body acceleration and the suspension deflection of the quarter car each have an invariant point within the frequency range of interest, that all three measures can be improved together only near the unsprung natural frequency, and that severe trade-offs exist at every other frequency [3]. Road holding gained a semi-active law of its own when the ground-hook concept, a damper from the wheel to the road, was extended for truck suspensions in 1997 [11], and magnetorheological fluid devices offered another way to vary damping electrically [12]. Linear-quadratic studies clarified the performance potential, the bandwidth requirements and the optimal structure of advanced suspensions, and practical developments followed [13].

Semi-active control laws were later compared on a common single-corner model [5], preview controllers for an active suspension were designed and implemented on a vehicle [14], and a 2015 survey reviewed the state of the art in active and semi-active suspension control [15]. The comfort side of the problem rests on how people perceive vibration in a vehicle [16], which is why a comfort number in this simulator is always read beside a tire load and a travel.

11.3 Quarter Car on a Road Defined in Distance

One corner of a car is simulated in vertical motion only: the quarter body and its load, the wheel with its hub and brake, a spring and a strut unit between them, a tire below and bump stops at the ends of the travel. The car drives at speed v along a road whose height $z_r(s)$ depends on the distance s, so a feature of wavelength λ arrives at the frequency $f = v/\lambda$. Each symbol of the equations is placed on the car in Figure 11.1.

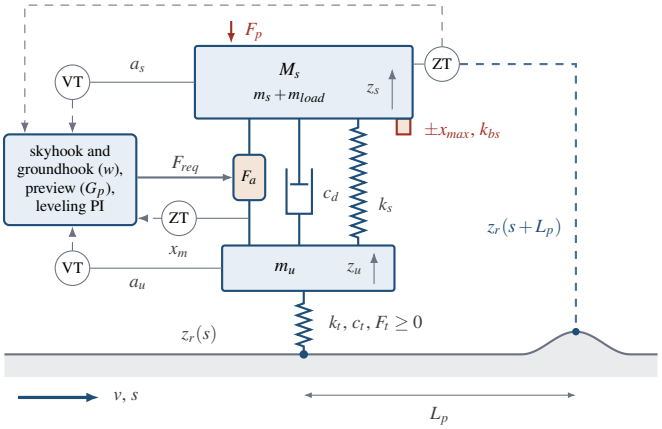


Figure 11.1 One corner of a car as simulated. The body M_s, the quarter body m_s plus the load m_{load}, and the wheel m_u are measured upward from rest at the design load as z_s and z_u, and are joined by the spring k_s and the strut unit, a damper c_d beside an actuator F_a, with bump stops k_{bs} beyond $\pm x_{max}$ of the ride height $x = z_s - z_u$. The tire k_t, c_t pushes on the road $z_r(s)$ and cannot pull, and F_p is the push of the bounce test. Two accelerometers VT and a ride-height sensor ZT reach the controller through noise and a common delay, and the preview sensor ZT reads the road L_p ahead of the wheel. The request F_{req} becomes an actuator force command or a damper valve command. The passive twin, not drawn, is the same car with the fixed damper c_p on the same road.

The signals are classified in Table 11.1. Eleven states are integrated: the body and wheel heights and velocities, the actuator force F_a, the damper coefficient c_d, the four motion states of a passive twin and the distance s. The tire load, the bump-stop force and every event are computed from them after each step.

11.3.1 Body, Wheel and Strut Unit

Heights are measured upward from where the body and the wheel rest on a level road with the design load m_{ref} aboard, and the spring and tire preloads $F_{s0} = (m_s + m_{ref})g$ and $F_{t0} = (m_s + m_{ref} + m_u)g$ hold the car there. The body mass is $M_s = m_s + m_{load}$, so a heavier load settles the body lower until the spring takes it. The two masses obey Equation (11.1).

$$\begin{aligned} M_s\ddot{z}_s &= F_{s0} - k_s x + F_u + F_{bs} + F_p - M_s g \\ m_u\ddot{z}_u &= F_t - F_{s0} + k_s x - F_u - F_{bs} - m_u g \end{aligned} \tag{11.1}$$

where $x = z_s - z_u$ is the ride height measured from the design height, F_u the force of the strut unit, F_{bs} the bump-stop force and F_p the bounce-test push, −800 N for 0.4 s of simulated time when the student presses the body down. A positive F_u pushes the body up and the wheel down by the same amount.

The strut unit is a damper beside an actuator, and both have dynamics of their own, Equation (11.2).

$$\begin{aligned} F_u &= F_a - \eta\, c_d\,(\dot{z}_s - \dot{z}_u) \\ \tau_d\dot{c}_d = c_{cmd} - c_d, &\qquad \tau_a\dot{F}_a = F_{cmd} - F_a \end{aligned} \tag{11.2}$$

where $\eta = 1 - \varepsilon_w/100$ is the fraction of damping left after the wear ε_w in percent, $\tau_d = 15$ ms is the valve lag and $\tau_a = 10$ ms the actuator lag. The passive strut commands $c_{cmd} = c_p = 1800$ N s/m and $F_{cmd} = 0$. The semi-active strut commands c_{cmd} within $[c_{min}, c_{max}] = [300, 4500]$ N s/m and $F_{cmd} = 0$. The active strut commands a force within $\pm F_{av}$ and keeps a residual damper $c_{cmd} = c_{res} = 500$ N s/m that does not wear ($\eta = 1$). A lag of a command that stays inside $[c_{min}, c_{max}]$ stays inside it, so the

damper term is $-\eta c_d(\dot{z}_s - \dot{z}_u)$ with $c_d > 0$ at every instant, including the intermediate stages of the integrator, and the damper can never feed energy in.

11.3.2 Tire, Bump Stops and Events

The tire is a spring and a light damper that can push on the road but not pull, Equation (11.3).

$$F_t = \begin{cases} \max\left(0,\ k_t\delta + c_t\dot{\delta}\right), & \delta > 0 \\ 0, & \delta \le 0 \end{cases} \qquad \delta = F_{t0}/k_t + z_r(s) - z_u \tag{11.3}$$

where δ is the true compression, 19.6 mm at rest, and $\dot{\delta} = v\,z_r'(s) - \dot{z}_u$. Beyond the free travel the bump stops take over, Equation (11.4).

$$F_{bs} = \begin{cases} k_{bs}(-x_{max} - x), & x < -x_{max} \\ -k_{bs}(x - x_{max}), & x > x_{max} \\ 0, & \text{otherwise} \end{cases} \tag{11.4}$$

with $x_{max} = 80\,\text{mm}$ and $k_{bs} = 300\,\text{kN/m}$, twelve times the spring rate. Both nonlinearities enter the right-hand side as force laws, never as clipped coordinates, and both are counted as events. After every step the model checks whether F_t has reached zero or F_{bs} has become nonzero, counts each new liftoff and each new contact, accumulates the time the wheel spends in the air, and marks the place on the road.

11.3.3 Roads

Every road is a function of distance. Humps, potholes and the washboard start 20 m in and the rough road ramps in from 15 m, so each run begins on level road, for 1.67 s over the humps at 12 m/s. A hump of height H and length L, a pothole of depth D and length L_h with rounded edges of length e, and a washboard of amplitude A and wavelength λ follow Equation (11.5).

$$\begin{aligned} z_{hump} &= \tfrac{H}{2}\left[1 - \cos(2\pi u/L)\right], \quad 0 \le u \le L \\ z_{hole} &= -\tfrac{D}{2}\left[1 - \cos(\pi u/e)\right], \quad 0 \le u < e \\ z_{wash} &= A\,r(s)\sin[2\pi(s-20)/\lambda] \end{aligned} \tag{11.5}$$

where $u = (s - 20) \bmod 40$ repeats the hump and the pothole every 40 m, the pothole bottom is flat at $-D$ and its far edge mirrors the near one, $e = \min(0.15\,\text{m}, L_h/2)$ is the width a tire envelops, and $r(s)$ is a raised-cosine ramp over the first 2λ. The defaults are $H = 60\,\text{mm}$, $L = 3\,\text{m}$, $D = 50\,\text{mm}$, $L_h = 0.9\,\text{m}$, $A = 7\,\text{mm}$ and $\lambda = 1.25\,\text{m}$. The rough road is a first-order random process in distance, Equation (11.6).

$$\begin{aligned} z_k &= a z_{k-1} + \sigma\sqrt{1 - a^2}\,\xi_k \\ a &= \mathrm{e}^{-2\pi n_c \Delta s}, \qquad \sigma^2 = \frac{\pi G_{ref}\, n_0^2}{2 n_c} \end{aligned} \tag{11.6}$$

with ξ_k standard normal, $\Delta s = 0.05\,\text{m}$, $n_0 = 0.1$ and $n_c = 0.05$ cycles/m and $G_{ref} = 64 \times 10^{-6}\,\text{m}^3$. Its one-sided spectral density falls as n^{-2} above n_c, and the table is scaled at lookup by $\sqrt{G_d(n_0)/G_{ref}}$, so the ISO 8608 roughness $G_d(n_0) = 256 \times 10^{-6}\,\text{m}^3$, the class C geometric mean, gives a process standard deviation of 8.97 mm. The table holds 1000 m, is blended over its last 20 m into its first point, has its mean removed, is interpolated linearly and is drawn afresh from the seeded generator at every reset. The challenge route joins 100 m of rough road from 15 to 115 m, washboard from 120 to 132 m, each ramped in and out over 5 m and 2λ, a smooth dip 40 mm deep and 2.5 m long at 150 m and a hump at 165 m, and it ends at 185 m. A “Bump ahead” event places a hump of the configured size 4 m in front of the wheel on any road.

11.3.4 Passive Twin, Metrics and Numerics

The passive twin obeys Equations (11.1), (11.3) and (11.4) with states $\tilde{z}_s$, $\tilde{z}_u$ and the unworn damper $F_u = -c_p(\dot{\tilde{z}}_s - \dot{\tilde{z}}_u)$, and it shares the road, the speed, the load and the push. Both cars keep separate metrics, accumulated at every physics step from the reset and frozen when the challenge route ends, Equation (11.7).

$$a_{rms} = \sqrt{\frac{1}{T}\int_0^T \ddot{z}_s^2\, dt}$$
$$\Delta F_{t,rms} = \frac{100}{F_{st}}\sqrt{\frac{1}{T}\int_0^T (F_t - F_{st})^2 dt} \tag{11.7}$$

where $F_{st} = (M_s + m_u)g$ is the static tire load, 3924 N at the design load. Comfort against the twin is $100(a_{rms}/\tilde{a}_{rms} - 1)$ in percent. The scorecard also reports the peak travel $\max|x|$, the liftoffs and the time in the air, the bump-stop hits, the positive mechanical work of the actuator $\int \max[0, F_a(\dot{z}_s - \dot{z}_u)]\, dt$ or the heat taken out by a damper, and the ride time. Comfort is unweighted, and the passenger drawn in the scene takes words adapted from the ISO 2631-1 comfort bands at fixed cut points of a running 1 s rms, as an indication only.

The car starts at rest in its static position on level road, which is the origin at the design load, with c_d at the resting value of the chosen strut and, for an active strut in automatic, F_a holding any load difference from m_{ref} at the ride height setpoint. The engine of Section 1.6.1 advances the states by the fourth-order Runge-Kutta method on a fixed 0.5 ms step, the controller runs every 2 ms, and the historian logs 100 samples per second. The equilibrium residual at the design load is 2.1×10^{-14}, and halving the step changes the rms, the peak and the tire-load rms of case 1 over 5 s by a relative 1.6×10^{-5}, both from the application's own tests. The noise generator is seeded with 20261013 (20261014 for case 8 and 20261015 for the challenge), so every run replays exactly.

The parameters and the default settings are listed in Table 11.2. The simulation runs at 0.1, 0.25, 0.5, 1 and 2 times real time, 0.5 by default, and the trends offer windows of 2, 6, 15 and 30 s.

11.3.5 Assumptions

The car is one corner in vertical motion. Pitch, roll and the load transfer from the other wheels are absent and the passenger moves rigidly with the body without a seat, which breaks down in braking and cornering and for any control of pitch and roll.

The spring and the tire are linear, the bump stops are stiff linear one-sided springs and the strut has no friction. The tire touches the road at one point and follows the profile exactly except at the pothole edges, which are rounded to the width a tire envelops. The idealization breaks down for features shorter than the contact patch, which a real tire bridges, and for edges it flattens.

The road is a single rigid track and the speed is constant between changes. Longitudinal and lateral tire forces are not modeled, so grip is represented by the vertical tire load alone, the quantity the braking and steering forces scale with, and the model does not hold during hard braking.

The actuator is a first-order lag with a slew limit and a force limit, which leaves out valve and pressure dynamics. Its energy counts positive mechanical work only, with no electrical or hydraulic losses, no regeneration and no power for holding a static force, which a real hydraulic unit does spend.

The sensors carry white noise and one common delay, and the preview reading is the true profile plus independent noise, which a camera's biased errors would violate. The speed and the distance are known exactly. Comfort is the unweighted rms body acceleration, a comparison between struts on the same road rather than a rating in the sense of ISO 2631-1.

Table 11.2 Parameters, Adjustable in the Model Panel Unless Marked Hidden, and Default Settings of the Control Cards Below the Rule

Symbol	Meaning	Value	Range
m_s	quarter body, empty	300 kg	200–500
m_{ref}	design load	60 kg	0–200
m_u	wheel, hub and brake	40 kg	25–80
k_s	suspension spring	25 kN/m	10–60
k_t, c_t	tire stiffness, damping	200 kN/m, 100 N s/m	100–350, 0–1000
x_{max}, k_{bs}	free travel, bump stop	80 mm, 300 kN/m	40–150, 50–1000
g	gravity (hidden)	9.81 m/s^2	9.7–9.9
c_p	passive damper	1800 N s/m	300–6000
c_{min}, c_{max}	semi-active range	300, 4500 N s/m	100–1500, 1500–9000
τ_d	damper valve lag	15 ms	2–80
c_{res}	residual damper, active	500 N s/m	0–3000
F_{max}	actuator force limit	2500 N	300–6000
$\dot{F}_{max}$	actuator slew limit	400 kN/s	20–1000
τ_a	actuator lag	10 ms	2–100
t_d	sensor delay	2 ms	0–50
σ_{as}, σ_{au}	accelerometer noise	0.02, 0.2 m/s^2	0–0.5, 0–3
σ_x	ride-height noise	0.5 mm	0–5
τ_v, τ_h	velocity leak, height filter	1.0, 1.0 s	0.2–5, 0.1–5
L_p, t_{lead}	preview distance, lead	5 m, 20 ms	1–15, 0–100
H, L	hump height, length	60 mm, 3 m	10–120, 0.5–6
D, L_h	pothole depth, length	50 mm, 0.9 m	10–120, 0.4–3
A, λ	washboard	7 mm, 1.25 m	2–30, 0.3–6
$G_d(n_0)$	roughness	256×10^{-6} m^3	16–1024 (10^{-6})
—	feature spacing (hidden)	40 m	20–100
w	comfort preference	70 %	0–100
C_{sky}, C_{gnd}	skyhook, groundhook gains	4000, 2500 N s/m	0–10000
K_c, τ_I	leveling gain, integral time	5 N/mm, 0.5 s	0–50, 0.1–20
h_{SP}	ride height setpoint	0 mm	−40–40
G_p, W	preview gain, window	1, 0.3 m	0–1.5, 0.05–2

11.4 One Force on Two Masses

11.4.1 Modes and the Constraint

At the design load the body bounces on the spring and tire in series at $\sqrt{k_s k_t/[(k_s+k_t)M_s]}/2\pi = 1.25$ Hz, and the wheel hops between the spring and the tire at $\sqrt{(k_s+k_t)/m_u}/2\pi = 11.94$ Hz. The coupled undamped frequencies are 1.2496 and 11.945 Hz, so the two approximations the application displays are within 0.1 %. With the damper at c_p the eigenvalues of the linear car (Section 1.4.4) give damping ratios of 0.25 for the body mode and 0.33 for wheel hop.

Adding the two lines of Equation (11.1) removes the spring, the strut and the bump stops. With the tire on the road and the load at its design value, what is left is Equation (11.8).

$$M_s\ddot{z}_s + m_u\ddot{z}_u = k_t(z_r - z_u) + c_t(\dot{z}_r - \dot{z}_u) + F_p \tag{11.8}$$

Apart from the push of the bounce test, only the tire can change the total momentum of the corner. Whatever the strut does for one mass it undoes for the other, and the controller chooses only how the tire force is shared between them.

11.4.2 Two Invariant Points

For a harmonic road $z_r = Z_r e^{j\omega t}$, and with c_t neglected, Equation (11.8) gives the two relations of Equation (11.9).

$$\begin{aligned} -\omega^2 M_s Z_s &= k_t Z_r - (k_t - \omega^2 m_u) Z_u \\ k_t(Z_r - Z_u) &= -\omega^2 \left[(M_s + m_u)Z_u + M_s X\right] \end{aligned} \tag{11.9}$$

where $X = Z_s - Z_u$ is the amplitude of the ride height. At $\omega_1 = \sqrt{k_t/m_u}$ the wheel drops out of the first line, and the body acceleration is $k_t Z_r/M_s$ whatever the strut does. At $\omega_2 = \sqrt{k_t/(M_s+m_u)}$ the

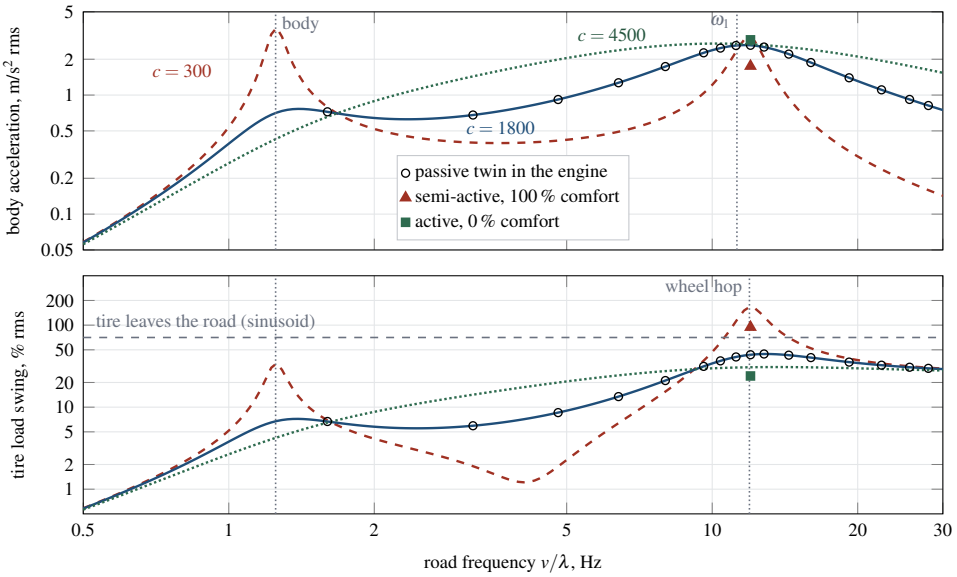


Figure 11.2 Steady response of the quarter car to the 7 mm washboard against road frequency. The lines are the linear model with a damper of 300 (dashed), 1800 (solid) or 4500 N s/m (densely dotted), the circles the passive twin in the running engine at 2 to 35 m/s, and the filled triangle and square the semi-active strut at 100 % comfort (case 3) and the active strut at 0 % with its full force (case 7 before the derating), both at 15 m/s. The dotted lines mark the body mode, the invariant point ω_1 and wheel hop, and above the horizontal dashed line a sinusoidal tire load would reach zero.

wheel drops out of the second, and the ride height is $X = -(M_s + m_u)Z_r/M_s$. These are the invariant points of the quarter car [3]. For the default car $\omega_1/2\pi = 11.25$ Hz, just below wheel hop, where the body acceleration is 0.556 m/s^2 per millimeter of road, and $\omega_2/2\pi = 3.56$ Hz, where the travel is 1.11 times the road amplitude. The tire damping of 100 N s/m blurs the first point slightly, so passive dampers from 300 to 4500 N s/m give 0.483 to 0.544 m/s^2 per millimeter there.

Both relations hold only while $F_t > 0$, because Equation (11.3) is linear only in contact. A strut can therefore lower the body acceleration at wheel hop by more than the blur only by unloading the tire until it leaves the road. In the linear model, a pure skyhook law of 4000 N s/m with the residual damper brings the washboard response at 11.25 Hz from 2.61 to 2.39 m/s^2 rms but needs a tire-load swing of 94.2 % rms, an amplitude of 133 % of the static load, which a tire that cannot pull does not have. A lag-free version of the default blend gains 2 %, 2.56 against 2.61 m/s^2, and raises the tire-load swing from 41.1 to 52.8 %.

🕮 **Pause.** Try Exercise 11.1 before reading on.

11.4.3 Frequency Is Speed Over Wavelength

Because the road is a pattern in distance, the speed decides which mode a road excites. The steady response of the linear passive car to the 7 mm washboard is drawn against $f = v/\lambda$ in Figure 11.2 for the softest, the nominal and the firmest damper. The circles are the passive twin in the running engine at sixteen speeds from 2 to 35 m/s, measured over 5 s once the washboard is fully under the wheel. They agree with the linear curve to within 0.08 % in body acceleration and in tire load, because the passive car is linear while its tire stays down and its travel stays inside the stops, which it does at every one of these speeds. The largest difference is at 2 m/s, where the start-up transient of the body mode has not quite died out.

Three features of Figure 11.2 organize the scenarios. The body acceleration peaks near wheel hop, not near the body mode, because the road input is a displacement and its acceleration grows with the square of the frequency, so the passive twin sees 0.92 m/s^2 rms at 6 m/s, 2.62 m/s^2 at 15 m/s and 1.11 m/s^2 at 28 m/s. The three dampers nearly meet at ω_1 and separate on either side, soft damping calming the body away from the two resonances and firm damping calming it at them. The soft damper's tire load crosses the liftoff line around wheel hop, which is where the comfort-biased struts of case 3 lose the road.

11.4.4 What No Strut Can Follow

A steep pothole edge asks more of the wheel than any strut can supply. Following the entry edge of Equation (11.5) needs a downward acceleration of $(D/2)(\pi v/e)^2$, which is 1579 m/s^2 at 12 m/s. Once the tire unloads, the spring preload and the wheel's own weight accelerate it downward at most at $(F_{s0}+m_u g)/m_u = 98.1\,\mathrm{m/s^2}$, less whatever the damper resists, and a fully pushing actuator adds $F_{max}/m_u = 62.5\,\mathrm{m/s^2}$. The wheel can follow the edge only below 2.99 m/s, or 3.83 m/s with the whole actuator pushing it down, and at 12 m/s every strut leaves the road. The edge speed, not the controller, decides what happens next.

The semi-active strut has the opposite limit. Its damper force times the relative velocity is $-\eta c_d(\dot z_s-\dot z_u)^2 \le 0$ at every instant, so it can oppose motion with any force its valve range allows and can never assist it. A shipped test finds the largest value of that product over 8 s on the washboard to be -6.6×10^{-36} W, so it never turns positive.

11.5 The Force Request and Its Three Struts

One control law serves every strut, and the hardware decides how much of it is delivered. The controller executes every $\Delta t_c = 2\,\mathrm{ms}$, a 500 Hz loop, before each fourth physics step, and holds its commands over the next 2 ms. It never sees the true state.

11.5.1 Instruments and Estimates

Each scan reads a body accelerometer, a wheel accelerometer and a ride-height sensor, adds white noise of $\sigma_{as} = 0.02\,\mathrm{m/s^2}$, $\sigma_{au} = 0.2\,\mathrm{m/s^2}$ and $\sigma_x = 0.5\,\mathrm{mm}$, and passes all three through one delay line of $t_d = 2\,\mathrm{ms}$ with linear interpolation. Velocities come from leaky integrators and the ride height from a first-order filter, Equation (11.10).

$$\begin{aligned} \hat{\dot z}_{s,k} &= \hat{\dot z}_{s,k-1} + \left(a_{s,m} - \hat{\dot z}_{s,k-1}/\tau_v\right)\Delta t_c \\ \hat x_k &= \hat x_{k-1}\,\mathrm{e}^{-\Delta t_c/\tau_h} + x_m\left(1-\mathrm{e}^{-\Delta t_c/\tau_h}\right) \end{aligned} \tag{11.10}$$

with the same integrator for $\hat{\dot z}_u$ from $a_{u,m}$, $\tau_v = 1\,\mathrm{s}$ and $\tau_h = 1\,\mathrm{s}$. The leak keeps the integrated noise from drifting, attenuates motion below about 0.16 Hz and adds 7° of phase lead at the body mode. The preview sensor reads the road $L_p = 5\,\mathrm{m}$ ahead of the wheel on every scan, adds noise of standard deviation σ_p, and stores the reading against the distance at which it was taken. When the wheel approaches that distance, the controller fits a least-squares line through the stored readings within $\pm W/2$ of the point $q = s + v t_{lead}$, which is where the wheel will be one lead time $t_{lead} = 20\,\mathrm{ms}$ later, and takes its height $\hat z_r$ and slope $\hat z_r'$. With fewer than three readings in the window it interpolates between the two that bracket q. Volume 20, Chapter 12 reads a seam from a buffer indexed by travel distance in the same way, because there too the sensor and the actuator are separated by distance rather than time.

11.5.2 Skyhook, Groundhook and Preview

The request is the sum of three terms, Equation (11.11).

$$\begin{aligned} F_{sky} &= -w C_{sky} \hat{\dot{z}}_s, \qquad F_{gnd} = (1-w) C_{gnd} \hat{\dot{z}}_u \\ F_{prev} &= -G_p \left(k_s \hat{z}_r + c_\pi v \hat{z}_r' \right) \\ F_{dyn} &= F_{sky} + F_{gnd} + F_{prev} \end{aligned} \tag{11.11}$$

where w is the comfort preference, from 0 for pure groundhook to 1 for pure skyhook, and c_π is the passive damping the strut keeps, c_{res} for the active strut and c_{min} for the semi-active one. The skyhook term is a damper from the body to the sky and pulls a rising body down, the groundhook term is a damper from the wheel to the road and pushes a rising wheel down, and a shipped test checks both signs. The preview term cancels what the road under the wheel is about to push through the spring and the damper, so the strut shortens over a hump and lengthens into a dip while the body stays level. It is off by default.

11.5.3 Passive, Semi-Active and Active Hardware

The passive strut ignores the request. The semi-active strut clips it to the rating $\pm F_{max}$, because a damper has no actuator to derate, and sets its valve to produce it where a damper can, Equation (11.12).

$$c_{cmd} = \begin{cases} \text{clip}_{[c_{min},\, c_{max}]} \left(-F_{req} / \hat{v}_{rel} \right), & F_{req} \hat{v}_{rel} < 0 \\ c_{min}, & \text{else} \end{cases} \tag{11.12}$$

with $\hat{v}_{rel} = \hat{\dot{z}}_s - \hat{\dot{z}}_u$ and the first case taken only when $|\hat{v}_{rel}| > 10^{-4}\,\text{m/s}$. When the request has the sign that would feed energy in, the valve goes soft and the delivered force parts from the requested one on the force trend. The strut counts as saturated whenever $|F_{req}| > 25\,\text{N}$ and the damper cannot deliver it.

Pause. Try Exercise 11.2 before reading on.

The active strut delivers the request through its lag and limits, and it alone carries a load-leveling loop, because a damper at rest carries no load. The available force is $F_{av} = F_{max}(1 - d/100)$, and when a derating lowers it the held output drops to the new limit at once rather than slewing down.

11.5.4 Load-Leveling PI in Velocity Form

The leveling loop is the framework's shared PID in its velocity, or incremental, form, with the dynamic request entering as the output bias, Equation (11.13).

$$\begin{aligned} e_k &= h_{SP} - \hat{x}_k, \qquad P_k = K_c e_k \\ S_k &= S_{k-1} + (P_k - P_{k-1}) \\ u_k &= F_{dyn,k} + S_k + I_{k-1} \\ \bar{u}_k &= \text{clamp}(u_k, -F_{av}, F_{av}) \\ F_{cmd,k} &= \text{slew}\left(\bar{u}_k,\ F_{cmd,k-1},\ \dot{F}_{max} \Delta t_c \right) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{F_{cmd,k} - u_k}{T_t} \right] \Delta t_c \end{aligned} \tag{11.13}$$

The terms are those of Equations (1.12) to (1.15) in Section 1.6.2, with the derivative time zero and the dynamic request as the bias, so the running sum S_k telescopes to $K_c e_k$ and is never clamped. Back-calculation with the tracking time $T_t = \tau_I = 0.5\,\text{s}$ acts on the integral only, and the clamp and the slew limit of 400 kN/s act on the emitted command. The integral line runs only while $K_c/\tau_I \neq 0$, so a leveling gain of zero freezes the integral, and case 5 switches the leveling off by setting $K_c = 0$ before the run, which leaves the integral at its seed of 0 N. Because the fast terms enter as the bias, clipping of the whole request also reaches the leveling integral through the back-calculation. The

Table 11.3 Scenario Sequence

#	Scenario	Setting	Teaching point
1	Isolated bump	12 m/s, active	feedback and preview
2	Pothole	12 m/s, active	contact loss
3	Washboard road	15 m/s, semi, 100 %	wheel hop by tuning
4	Higher speed	passive, 6 to 15 m/s	wheel hop by speed
5	Heavier passenger load	+140 kg at 0.2 s	leveling and travel
6	Reduced damping	passive, 60 % worn	a quiet fault
7	Actuator saturation	0 %, 500 N at 4 s	lost authority
8	Noisy preview	rough, 25 mm at 8 s	trusting a sensor
9	A force plan inside the camera horizon	active, MPC, preview	force and contact
10	Challenge, three surfaces	route, 100 %, no preview	graded run

loop is reverse acting, because a positive force raises the ride height, so the command falls as the measured height rises, and a shipped test checks that a sagging body receives more force. The reset seeds the integral with the resting actuator force, and on a passive or semi-active strut the loop tracks with zero contribution, so switching to the active strut starts it without a jump.

11.5.5 Why the Leveling Is Slow and the Actuator Fast

The defaults are $K_c = 5\,\mathrm{N/mm}$ and $\tau_I = 0.5\,\mathrm{s}$, an integral-dominant loop behind a 1 s height filter. At the body frequency ω_b the filter lags the height by nearly 90°, which turns a proportional term into a force in phase with the body velocity, a negative damping of about $K_c/(\omega_b^2 \tau_h)$. That is 81 N s/m at the default gain and 811 N s/m at the slider's maximum of 50 N/mm. At 70 % comfort the skyhook term dominates and the body stays stable at every gain the slider allows. At 0 % comfort the groundhook's reaction on the body cancels most of the residual damper. A push test on level road with the sensor noise switched off gives the body mode a damping ratio of 0.070 with the residual damper alone, 0.026 once the groundhook term is added, and 0.015 with the leveling at its default gain. Over 40 s the body still decays at 11 N/mm, grows slowly at 12 N/mm, and from 13 N/mm rings into its bump stops.

🕮 **Pause.** Try Exercise 11.3 before reading on.

The actuator is sized for wheel hop. A 400 kN/s slew limit lets it follow a full 2500 N sinusoid up to $\dot{F}_{max}/(2\pi F_{max}) = 25.5\,\mathrm{Hz}$, and its 10 ms lag costs 37° of phase at 11.9 Hz. On the washboard at 15 m/s and 0 % comfort the default actuator holds the tire-load swing to 24.0 % rms with no liftoff. With a 20 ms lag alone it holds 26.7 %, while a slew limit of 50 kN/s alone lets the groundhook force arrive late and pump the wheel, to 97.7 % rms and 60 liftoffs in 5 s. The slew rate, not the lag, decides whether the road-holding law works at wheel hop.

11.5.6 Manual Mode and Model Boundaries

In manual mode the student moves the actuator force of the active strut within $\pm F_{av}$, through the same slew limit, or the valve setting of the semi-active strut within $[c_{min}, c_{max}]$. The leveling integral tracks the manual force, so the return to automatic is bumpless, as a framework test confirms (Section 1.6.2). The simulation model is Equations (11.1) to (11.6). The original controller's internal model is limited to k_s and c_π in the preview term, since the skyhook and groundhook laws need no model, and no model is identified. The linear transfer functions of Section 11.4 are an analysis model of the passive car in contact and are not evaluated by the original control law.

11.5.7 A Nominal Preview Force Optimization

The app also offers a predictive force law, selected as MPC under Active law in place of skyhook and preview. It follows the predictive control option of Section 1.6.3, and this subsection gives

Table 11.4 Measured Results. Values in Parentheses Are the Passive Twin on the Same Road at the Same Instant

#	Configuration and window	Body acceleration	Tire load swing	Travel, events and other
1	70 %, first hump	peak 4.72 (5.90) m/s^2	—	rise 36.7 (47.7) mm, undershoot 5.3 (20.8) mm
1	100 %, first hump	peak 3.09 m/s^2	—	rise 28.4 mm
1	70 % with preview, first hump	peak 1.92 m/s^2	—	rise 5.9 mm, strut compressed 61.1 mm
1	0–10 s: 70 %, 100 %, preview	0.885, 0.596, 0.345 (1.197) rms	8.1, 5.7, 4.3 (11.0) %	actuator work 1, 30, 138 J
2	12 m/s, first pothole	peak 11.45 (13.65) m/s^2	—	2 liftoffs, 42.0 ms (1, 29.0 ms)
2	12 m/s with preview	peak 9.73 m/s^2	—	2 liftoffs, 32.0 ms
2	3 m/s, without and with preview	—	—	12.0 ms (19.5 ms), then none
3	semi-active, 100 %, 3–8 s	1.75 (2.62) rms	94.4 (43.6) %	60 liftoffs, 1.62 s airborne, valve $\leq$ 840 N s/m
3	semi-active, 50 % and 0 %	2.60, 2.64 rms	54.0, 37.5 %	no liftoff, lowest tire load 898 N at 50 %
3	active, 70 %	2.91 rms	52.6 %	no liftoff
4	passive, 6, 15, 28 m/s	0.92, 2.62, 1.11 rms	8.6, 43.6, 32.5 %	road at 4.8, 12.0, 22.4 Hz
5	+140 kg, 0–14 s	—	—	0 (7) stop hits, lowest x −76.4 mm, twin settles −54.9 mm
5	leveling off, or $h_{SP} = +20$ mm	—	—	3 stop hits, or peak compression 56.4 mm
6	60 % worn, first hump, 0–10 s	peak 3.83 (5.90), 1.234 (1.197) rms	—	±13.9 (±2.0) mm 1.6–2.4 s after the hump
6	bounce test, level road	—	—	62.2 to 32.9 mm (51.8 to 9.9 mm) in one cycle
7	0 %, 2.5–4 s	—	24.0 (43.6) %	request peak 1438 N
7	derated to 500 N, 4.5–9 s	2.93 (2.62) rms	75.2 (43.6) %	clipped 92.1 % of the time, 54 liftoffs
7	derating removed at 9 s, 10–15 s	—	24.1 %	no liftoff
8	clean preview, 2–8 s	0.76, off 0.90 (1.09) rms	—	—
8	25 mm noise, 9–20 s	1.22, off 1.01 (1.23) rms	—	window 1 m 0.87, gain 0.5 1.00
10	100 %, no preview	−35.9 % against twin	—	6 liftoffs, 0 stop hits, 72 J
10	80 %, no preview	−17.4 %	—	no liftoff
10	80 % with preview	−28.5 %	—	no liftoff, 0 stop hits, 191 J

its plant-specific model, observer and weights. It applies only to the active strut. Its model uses travel $x = z_s - z_u$, velocities v_s, v_u, wheel height relative to road $\eta = z_u - z_r$, actuator force f and a body-force disturbance d. Force variables f, d, u are in kN, and the nominal contact-mode balances are

$$\begin{aligned} \dot{x} &= v_s - v_u, & \dot{\eta} &= v_u - v_r, \\ M_s \dot{v}_s &= -k_s x - c(v_s - v_u) + 10^3 (f + d), & & \\ m_u \dot{v}_u &= k_s x + c(v_s - v_u) - k_t \eta & & \\ &\quad - c_t (v_u - v_r) - 10^3 f, & & \\ \tau_a \dot{f} &= u - f, & \dot{d} &= 0. \end{aligned} \tag{11.14}$$

The declared values are $M_s = 360\,\text{kg}$, $m_u = 40\,\text{kg}$, $k_s = 25\,\text{kN/m}$, $k_t = 200\,\text{kN/m}$, $c = 500\,\text{N s/m}$, $c_t = 100\,\text{N s/m}$ and $\tau_a = 10\,\text{ms}$. These remain fixed when load, stiffness or actuator lag changes. Flight and stops are absent from prediction. Against the original nominal plant derivatives at two signed contact states, the matrix calculation agrees within 2.2×10^{-14}.

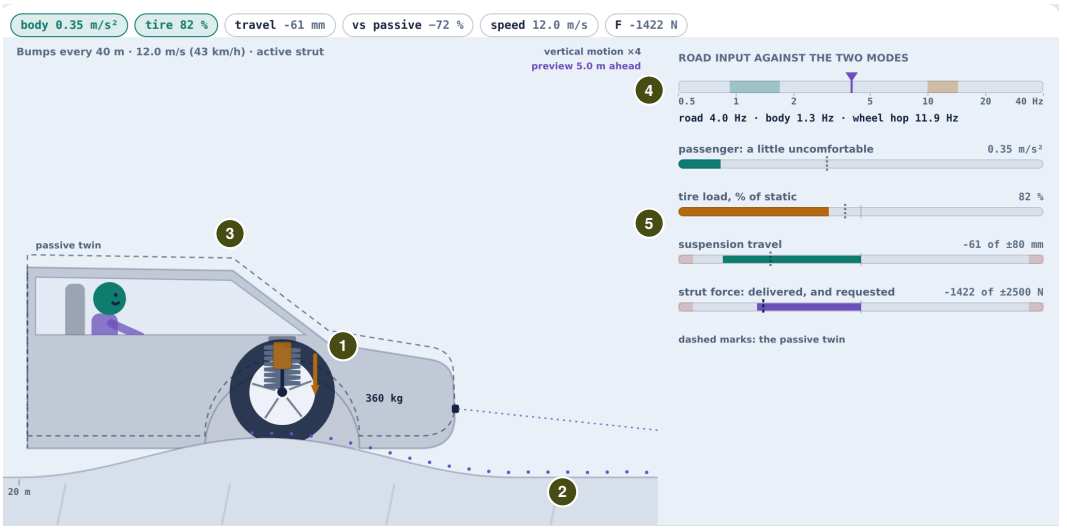


Figure 11.3 Case 1, the isolated hump, with road preview on, paused in automatic mode at $t = 1.80$ s. The picture is the plant view of the screen. The wheel is 1.6 m into the first 60 mm hump. In the cutaway the actuator arrow (1) points down as the strut retracts, a row of dots (2) marks the preview readings of the road ahead, and the dashed outline of the passive twin (3) rides above the active body. Beside it the road input at 4.0 Hz lies between the body band at 1.3 Hz and the wheel-hop band at 11.9 Hz (4), and the bars (5) read a running comfort of 0.35 m/s^2, a tire load of 82 %, 61 mm of the 80 mm compression travel and a delivered strut force of −1422 N, with the request marked beside it.

A six-state steady-state Kalman gain is applied to the existing delayed body/wheel accelerometers and raw height sensor. The observer corrects the preceding nominal 2 ms state before propagating, accounting for one report-delay step. Its discrete process covariance diagonal is $[10^{-9}, 10^{-5}, 10^{-9}, 10^{-4}, 10^{-6}, 10^{-7}]$ in the stated state units, and measurement covariance is $[0.02^2, 0.2^2, 0.0005^2]$. The covariance iteration converges in 528 steps to relative change 9.98×10^{-12}. The disturbance state is bounded at ± 5 kN and provides constant-force offset correction [17], without identifying passenger mass. On the new case the travel and tire-deflection estimation RMS are 0.108 and 0.0587 mm, and actuator-force estimation RMS is 0.183 N. Those values do not characterize flight or noisy-preview operation.

Only slopes interpolated from camera samples already stored at the current decision enter v_r. The controller integrates its own distance from the applied speed. Outside camera support, road height is held. At 12 m/s, the five-meter preview covers about 21 of 25 future intervals, not the entire horizon. Preview Off supplies no road input. Neither true position, physical state, route function, future disturbance nor current physical coefficient enters the new model. Volume 20, Chapter 5 plans a foil's flap in the same way, on water the bow has already measured.

Every 20 ms, a shared quadratic optimizer predicts 25 steps, 0.5 s, in ten move blocks [18]. Outputs are body acceleration, travel in cm and elastic tire-load variation in kN. Tracking weights are $[0.1 + 4w, 0.02, 0.1 + 4(1 - w)]$, with independent MPC comfort $w = 0.7$, terminal travel weight 0.2, force weight 0.002 and move weight 0.02. The ride-height target is retained. Force availability and moves up to the smaller of 400 N and the physical two-millisecond slew allowance are hard. Nominal travel within ±80 mm and elastic load $(M_s + m_u)g - k_t\eta \geq 0$ are soft, each with quadratic price 100 and linear price 1. The elastic surrogate omits tire damping, so it is not an actual contact guarantee. The leveling PI tracks each applied command once. A rejected plan holds force on handover, marks

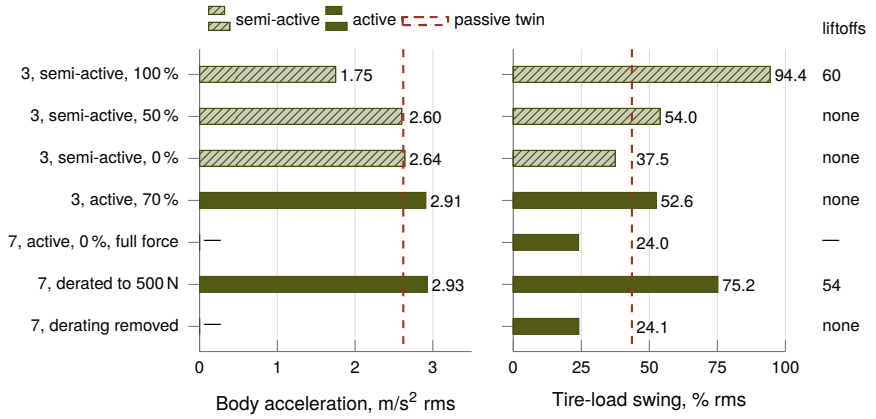


Figure 11.4 The washboard rows of Table 11.4, cases 3 and 7 at 15 m/s, with the passive twin's 2.62 m/s^2 and 43.6 % as dashed lines. Hatched bars are the semi-active strut at the comfort preference named, solid bars the active strut, and a dash marks a value the table does not report. Notice that the one setting that beats the twin's comfort by much is also the one that more than doubles its tire-load swing and leaves the road 60 times.

the historian and returns to the original law until recovery. The first solve primes the run.

11.6 Scenarios and Measured Results

The ten scenarios are listed in Table 11.3 and the measured outcomes in Table 11.4. Scenarios, completion rules and the data export work as Section 1.6.4 describes. All values come from runs of the engine at the scenario's seed with the settings named, changed at $t = 0$ unless a time is given, and windows are simulated time from the start of the run. A setting chosen before Run gives the same run as the same setting carried through Reset. Values over a window, and peaks, are taken at every 0.5 ms physics step. Every scenario starts from the same base settings, the active strut at 70 % comfort with $C_{sky} = 4000\,\mathrm{N\,s/m}$, $C_{gnd} = 2500\,\mathrm{N\,s/m}$, leveling on and preview off, and changes only what Table 11.3 lists. Figure 11.4 draws the washboard rows of Table 11.4, cases 3 and 7, against the passive twin.

> ✦ **DID YOU KNOW? A Worn Damper Feels Softer**
>
> From the simulator: on the first hump a damper that has lost 60 % of its damping peaks at 3.83 m/s^2 against 5.90 m/s^2 for the twin, and over 10 s it reads 1.234 m/s^2 rms against 1.197 m/s^2 (Table 11.4). The comfort number barely moves while the body still swings ±13.9 mm against ±2.0 mm after the hump.

11.6.1 Comfort, Grip and Work With a Finite Preview

Case 9 compares the two active laws with preview On over 8 s, seed 20261013 (Figure 11.5). MPC body acceleration RMS is 0.1550 versus 0.3148 m/s^2, and peak travel 55.4 versus 61.1 mm. Tire-load RMS is 3.78 versus 3.89% of static, with no liftoff or stop under either. Positive mechanical work increases to 148.2 from 92.4 J. A six-point hindsight grid includes a classical 100% comfort, $C_{sky} = 5000\,\mathrm{N\,s/m}$ setting with still lower RMS 0.1352 m/s^2 and 139.2 J, at more travel 63.6 mm. Thus the nominal MPC is not uniformly superior. Replacing only the original velocity estimates with Kalman estimates gives RMS 0.3281 m/s^2, so the main improvement is not explained by that estimator substitution alone. Both laws use the same camera buffer, and no additional route preview is supplied.

Table 11.5 Matched Active-Law Runs, Original / MPC

Case	Body RMS	Travel mm	Lifts	Work J
New, 8 s	.315 / .155	61.1 / 55.4	0 / 0	92 / 148
No preview, 8 s	.808 / .957	46.8 / 44.3	0 / 0	1 / 15
Pothole, 8 s	1.182 / 1.001	57.9 / 65.0	4 / 6	12 / 158
Load, 8 s	.667 / .511	76.4 / 78.5	0 / 0	89 / 162
Noisy, 20 s	1.042 / 2.184	38.4 / 85.9	0 / 4	481 / 1888
Suggested route	.885 / .653	66.7 / 65.4	0 / 5	191 / 369

RMS in m/s^2. MPC plans/mean Node microseconds: new 400/130, pothole 400/66, load 400/50, noisy 1000/80, route 617/68. No rejected plans in these runs.

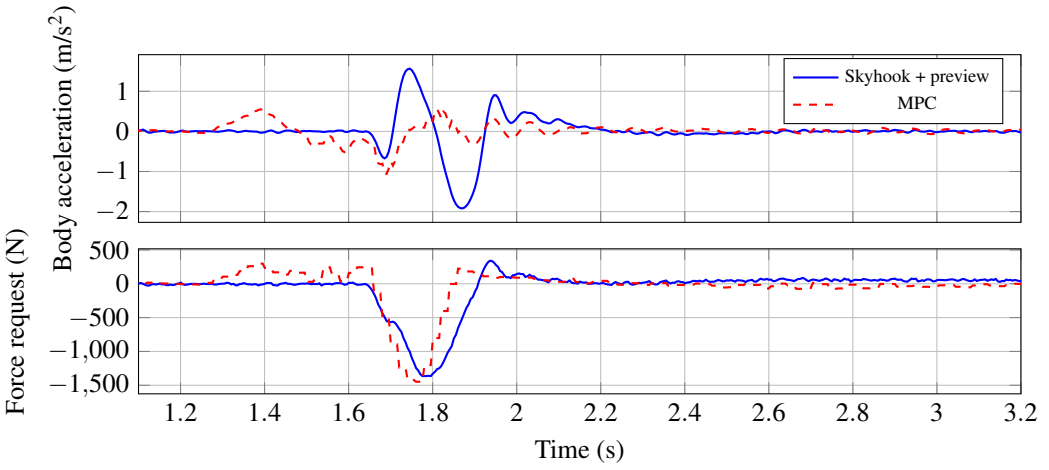


Figure 11.5 The first hump of case 9 under the same road camera, skyhook with preview as solid lines and the MPC as dashed lines. The force plan reduces body acceleration, while its greater positive mechanical work and its losses on contact-limited cases remain separate outcomes.

Without preview, MPC loses on the bump, RMS 0.9566 versus 0.8077 m/s^2. On a pothole it lowers RMS but raises tire-load variation and liftoffs from four to six. On the suggested route it lowers RMS to 0.6534 from 0.8854 m/s^2, but five liftoffs make it fail while the original suggested law passes. With a 0.3 m smoothing window in the noisy-preview case, MPC RMS is 2.184 versus 1.042 m/s^2, with four liftoffs and one stop versus none. A 1 m window improves MPC RMS to 0.6131 versus 0.8027 m/s^2, but tire-load variation is 20.99 versus 15.71% and work 682 versus 244.5 J. Modeled contact and soft bands do not preserve grip when the forecast is wrong.

The measurements behind this comparison fill 36 rows. They include passive and semi-active alternatives, active hardware on the washboard, six hindsight settings and half-step runs. The semi-active washboard results are identical under both laws because the MPC is inactive on that hardware. Halving the physics step retains the new-case ordering. Table 11.5 reports the separate objectives and planner workload. An unconstrained one-step instance of this actual nominal model agrees exactly with its hand normal equation.

Timing 125 plans in the browser per viewport gives mean solve times of 0.397 to 0.704 ms across six viewport and theme runs. At real-time playback the planner occupies 1.98 to 3.52 % of a CPU core, and twice that at the top two-times speed. Observer and rendering costs are excluded.

11.6.2 A Hump, With and Without the Road Ahead

Case 1 crosses a hump 60 mm high and 3 m long every 40 m at 12 m/s, the road input arriving at 4.0 Hz. Over the first hump the passive twin's body rises 47.7 mm, swings 20.8 mm below level on

the rebound and peaks at 5.90 m/s^2. The active strut at 70 % comfort rises 36.7 mm, undershoots 5.3 mm and peaks at 4.72 m/s^2, and 1.5 s after the wheel meets the hump the active body is 0.2 mm from level while the twin is still 3.7 mm below. At 100 % comfort the rise is 28.4 mm and the peak 3.09 m/s^2. The skyhook term pushes against the body's own velocity rather than against the relative velocity a damper sees, so it damps the rebound without passing the road's velocity to the body, and the body is still carried over the hump by the spring.

Preview changes the order of events. The strut starts retracting before the wheel reaches the hump, absorbs 61.1 mm of travel itself, and the body rises only 5.9 mm with a peak of 1.92 m/s^2 (Table 11.4), and Figure 11.3 shows the strut retracting over the hump. Over the first 10 s, which include three humps, the body acceleration is 0.885 m/s^2 rms against 1.197 m/s^2 for the twin, and 0.345 m/s^2 with preview. The price is actuator work, 138 J with preview against 1 J without, because the actuator now does the work of compressing the spring over each hump, where without preview it mostly absorbs power.

11.6.3 A Pothole No Strut Can Follow

Case 2 meets a pothole 50 mm deep and 0.9 m long with 0.15 m edges at 12 m/s, where the entry edge asks for 1579 m/s^2 of downward wheel acceleration and the wheel has 98.1 m/s^2. Dropping in is gentle for the body because the tire simply leaves the road, once for 29.0 ms on the twin and twice for 42.0 ms in all on the active strut. The jolt comes on the far side, where the falling wheel meets the rising road 0.15 m past the edge for the twin and 0.22 m past it for the active strut, at 13.65 m/s^2 and 11.45 m/s^2. Moving the preference to road holding barely helps, with one flight of 27.5 ms at 0 % comfort, because no damping law pulls the wheel down faster than its spring does. Preview presses the tire down just before the edge and cuts the flight to 32.0 ms and the peak to 9.73 m/s^2. Slowing to 3 m/s, where the edge asks for 98.7 m/s^2, still leaves the active strut airborne for 12.0 ms and the twin for 19.5 ms, and only with preview does the tire stay down, consistent with the 2.99 m/s and 3.83 m/s limits of Section 11.4.

11.6.4 Wheel Hop Found by Tuning and by Speed

Case 3 drives the washboard at 15 m/s, 12.0 Hz against the 11.9 Hz wheel hop, on the semi-active strut at 100 % comfort. From 3 to 8 s the body acceleration is 1.75 m/s^2 rms against 2.62 m/s^2 for the twin, a third lower, while the tire-load swing is 94.4 % against 43.6 % and the wheel leaves the road 60 times, once per ripple, for 1.62 s of the 5 s. The skyhook request is small at 12 Hz, so the valve never rises above 840 N s/m and the wheel is damped by little more than c_{min}, which is the soft curve of Figure 11.2 crossing the liftoff line. At 50 % the swing falls to 54.0 % with the tire load never below 898 N, and at 0 % to 37.5 %, better than the twin, with the body at 2.64 m/s^2, level with the twin as the nearby invariant point of Equation (11.9) leads one to expect. Only the setting that let the wheel fly moved the body acceleration by more than 1 %. The default active strut does worse than the passive twin on both measures here, 2.91 m/s^2 and 52.6 %. This is likely because its instantaneous damping is only the 500 N s/m residual damper, while the controller's force arrives through a lag and a delay that alone cost 46° at wheel hop.

> **TRY IT IN THE APP** **3 · Washboard Road**
>
> Load the scenario, press Run and watch the wheel and the red paint it leaves on the road. Set for comfort, the semi-active strut holds the body acceleration to 1.75 m/s^2 rms against 2.62 m/s^2 for the passive twin, a third lower. From 3 to 8 s the wheel pays for it: it leaves the road 60 times, airborne for 1.62 s of the 5 s.

Case 4 reaches the same band with speed alone. The passive car on the washboard sees 0.92 m/s^2 rms and a 8.6 % swing at 6 m/s once the ripples are fully under the wheel, and 2.62 m/s^2 and 43.6 % after the step to 15 m/s. A shipped test counts the tire load crossing its static value at 12.00 Hz, which is v/λ. At 28 m/s the ripples arrive at 22.4 Hz, above the band, and the ride improves to 1.11 m/s^2

and 32.5 %, because the wheel's own mass now filters the ripples. Worse is not a function of speed but of where the speed puts the road frequency, which Figure 11.2 draws.

11.6.5 Load and the Travel It Takes

Case 5 adds 140 kg to the corner at 0.2 s on the hump road at 10 m/s. The passive twin settles with its ride height 54.9 mm below the design value, which is $\Delta Mg/k_s$, and hits its bump stop 7 times in 14 s, once as the load lands and twice at each of the three humps. The active body plunges to −70.5 mm, is still 32.6 mm low when the first hump arrives at 2.0 s, and is back within 5 mm of its setpoint 5.1 s after the step. It never touches the stop, coming within 3.6 mm of it at −76.4 mm on the first hump, because the leveling force already carries most of the new weight and the skyhook term resists the rest of the plunge. At the end of the run the leveling loop holds 1373 N, which is 140 kg times g and 55 % of the actuator rating, authority the comfort terms no longer have. With the leveling off the active strut hits the stop 3 times. Raising the setpoint to +20 mm before the run starts the car 20 mm higher and keeps the peak compression to 56.4 mm, 20 mm less than at the design setpoint.

11.6.6 A Worn Damper and a Derated Actuator

Case 6 runs the hump road on a passive damper that has lost 60 % of its damping. The first hump feels softer, 3.83 m/s^2 against 5.90 m/s^2 for the healthy twin, because a weak damper passes less of the wheel's sudden motion to the body. The cost comes afterward. From 1.6 to 2.4 s after the wheel meets the hump the worn body still swings ±13.9 mm against ±2.0 mm, yet the body acceleration over 10 s, 1.234 m/s^2 against 1.197 m/s^2, barely records the fault. In the bounce test on level road the healthy damper cuts the bounce from 51.8 to 9.9 mm in one cycle and the worn one only from 62.2 to 32.9 mm, and on the washboard at 15 m/s the worn damper lets the wheel leave the road 60 times in 5 s. The active strut is unaffected by wear, and in the same push test its body stops at −43.0 mm and rebounds 10.8 mm.

Case 7 runs the active strut at 0 % comfort on the washboard. From 2.5 to 4 s the groundhook requests peak at 1438 N and hold the tire-load swing to 24.0 % against 43.6 % for the twin. At 4 s a thermal derating cuts the available force to 500 N. From 4.5 to 9 s the request is clipped 92.1 % of the time, the swing rises to 75.2 %, the body to 2.93 m/s^2 against 2.62 m/s^2, and the wheel leaves the road 54 times. A clipped force damps the wheel less the larger its motion becomes, so the hop grows until the tire leaves the road. With the derating removed at 9 s the swing returns to 24.1 % with no liftoff.

> **TRY IT IN THE APP** **7 · Actuator Saturation**
>
> Press Run and watch the force and Tire load strips. At 4 s the actuator is derated to 500 N. From 4.5 to 9 s the request is clipped 92.1 % of the time, the tire-load swing rises from 24.0 to 75.2 % rms and the wheel leaves the road 54 times. Set Actuator derating back to 0 % near 9 s, and the swing returns to 24.1 %.

11.6.7 Preview Through Spray

Case 8 drives rough asphalt at 15 m/s with preview on. From 2 to 8 s, with a clean reading, the body acceleration is 0.76 m/s^2 rms against 0.90 m/s^2 with preview off and 1.09 m/s^2 for the twin. At 8 s the reading turns noisy, 25 mm on a road that varies by a few millimeters, and from 9 to 20 s the noisy preview gives 1.22 m/s^2, as bad as the twin's 1.23 m/s^2, against 1.01 m/s^2 with preview switched off. The slope fitted to a noisy reading is mostly noise, and the preview force grows from 242 N rms on the clean reading to 740 N. Widening the smoothing window from 0.3 to 1 m averages the noise and keeps the longer road features, cuts the preview force to 290 N and brings the body to 0.87 m/s^2, while halving the preview gain reaches only 1.00 m/s^2. The 1 m window is also slightly better on the clean reading, 0.74 m/s^2 against 0.76 m/s^2.

11.6.8 The Graded Route

The challenge drives the 185 m route at 15 m/s and asks for a body acceleration at least 25 % below the twin's with no liftoff and no bump-stop contact. The verdict is taken when the route ends, and it counts only if the speed stays at 15 m/s and the load at 60 kg throughout and no model parameter or disturbance is changed. At 10 m/s the shipped tune would meet the comfort, liftoff and stop criteria at −47.1 %, which is why the speed is part of the case. Reset keeps the learner's setting and redraws the same seeded road, so each retry drives the same route. The twin finishes in 12.33 s at 1.238 m/s^2 rms, so the target is 0.928 m/s^2. As shipped, at 100 % comfort without preview, the body is 35.9 % better than the twin and the wheel leaves the road 6 times on the washboard, which fails the target. Moving to 80 % stops the liftoffs and leaves only 17.4 %. With preview at 80 % the body is 28.5 % better, a margin of 3.5 percentage points over the target, with no liftoff and no stop contact, for 191 J of actuator work against 72 J as shipped. A sweep in steps of 5 % passes at 75, 80 and 85 % with preview and at no setting without it. Without preview the only setting that reaches −25 %, 90 % at −25.1 %, leaves the road 5 times. At 0 % the body acceleration more than doubles, 104.0 % above the twin's, with one stop contact on the final hump, while on the rough section the body swings at about 1.16 Hz, the lightly damped body mode of Section 11.5.

11.7 Technology in Practice

What Measures It

Like the controller of this chapter, a production semi-active corner reads two accelerometers or an accelerometer and a stroke sensor [5]. The chapter's leaky integrators rebuild both velocities from the accelerations, and their leak hides motion slower than about 0.16 Hz. The road ahead can be scanned directly by a lidar, which costs much, or reconstructed from camera images as a height map [19]. The chapter's preview sensor stands for either and reads the road 5 m ahead with independent noise.

> **TECHNOLOGY** **Sensor: Accelerometers on the Body and the Wheel**
>
> A semi-active corner is usually read by two sensors per axle, either two accelerometers or an accelerometer and a stroke sensor that measures the suspension travel. Sensors weigh heavily on the cost, so the trend is toward one sensor per corner, with the second signal reconstructed by filtering, and toward sharing the sensors with stability and rollover control [5]. On one test car with semi-active dampers a prototype control unit read three vertical accelerometers on the body and two on the wheels [20]. The chapter's corner carries both accelerometers and the ride-height sensor.

What Moves It

The semi-active strut moves nothing itself: its valve only sets how hard the damper resists. The active strut must supply force. Hydraulic actuators are durable, deliver high forces and withstand high impulses, but they are complex to model, and in a heavy machine's hydraulic driveline the delay between a command and the actuator's response is large and depends on what the other actuators are doing [21]. The chapter's actuator is a first-order lag of 10 ms with a force limit of 2500 N and a slew limit of 400 kN/s.

> **TECHNOLOGY** **Final Element: A Damper Whose Valve Sets Its Force**
>
> Most semi-active dampers use solenoid valves in an electro-hydraulic circuit, and magnetorheological and electrorheological ones were expected to grow [5]. In a magnetorheological damper, micron-scale ferrous particles in a carrier fluid form chains under a magnetic field and raise its yield stress, and the damper can switch from near-minimal to high damping within milliseconds [22]. A real damper's force is not one coefficient times the stroke velocity: it lies in a region between a minimum and a maximum force curve, each a nonlinear function of that velocity, and the command interpolates between them [20]. The chapter's damper is linear, 300 to 4500 N s/m, behind a 15 ms valve lag.

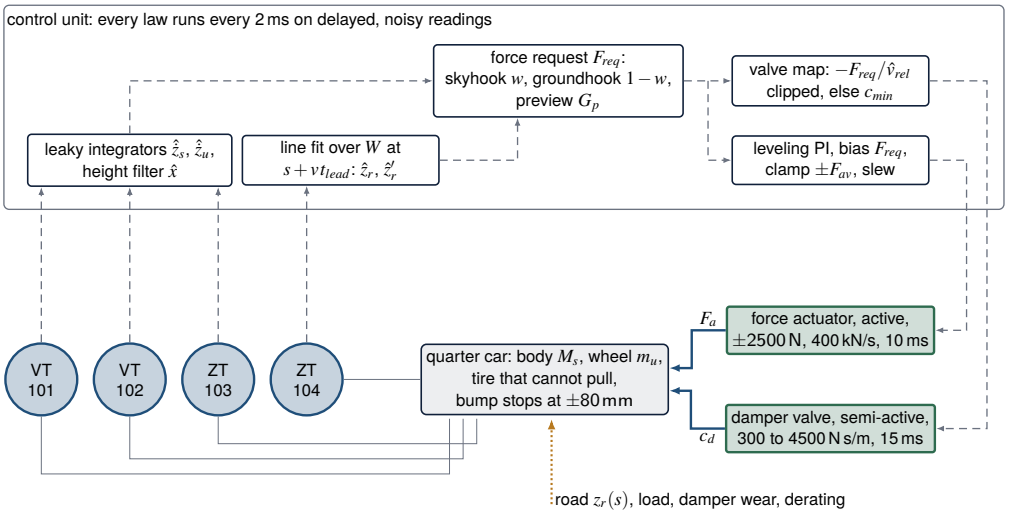


Figure 11.6 The suspension loop on vehicle hardware. The accelerometers VT-101 and VT-102 and the ride-height sensor ZT-103 feed the estimates, and the preview sensor ZT-104 feeds a line fit of the road the wheel is about to meet. One force request serves both struts: on the semi-active strut the valve map turns it into a damping coefficient, and on the active strut it enters the leveling PI as its bias before the clamp and the slew limit. Notice that the same request reaches two very different final elements, and that only the actuator can deliver a force with the sign that feeds energy in.

Where the Controller Runs

On one test car with semi-active dampers the control software ran on a prototype electronic control unit that sent a target current to each damper. Its parameters could be changed while driving from a laptop on the vehicle's network, and a commercial controller of this kind may hold hundreds of tunable parameters [20]. The chapter's controller runs every 2 ms, a 500 Hz loop. Its structure is one force request, the sum of the skyhook, groundhook and preview terms, which the semi-active valve map turns into a damping coefficient and which the active strut's leveling PI takes as its bias before the clamp and the slew limit. Figure 11.6 draws the loop on that hardware.

> **IN PRACTICE** **Tuning a Semi-Active Suspension on the Road**
>
> Fine-tuning a semi-active suspension is costly. It needs a calibration engineer who understands the control software and a professional driver who judges each proposed calibration, many hours of driving, and on a production vehicle it may take months. Most makers rely on the driver's subjective evaluation, and the comfort index of ISO 2631-1 is seldom used for suspensions in industry. Indoor alternatives place the driver in a hardware-in-the-loop simulator whose road is emulated by a robotic arm, or the car on four moving platforms. Each test on the road follows the same segment at the same speed profile, so that the road excites the same frequencies [20].

What Is Modernizing It

Three developments are changing the hardware around this loop. Road reconstruction from a top-down view of camera images reached elevation errors of 1.83 cm with one camera and 0.50 cm with a stereo pair on a real-world data set, which its authors judge promising for practical road preview [19]. Data-driven calibration with Bayesian optimization tuned a production car's semi-active controller in about an hour of experiments, where tuning by hand may take months [20]. Learned controllers for

active suspensions have moved from simulation to a real hydraulic machine, and they transferred well only when trained with actuator delays and a penalty on erratic actions [21]. The predictive force law of Section 11.5.7 is a fourth path, and its failures on the route and under a noisy camera show what a wrong forecast costs.

11.8 Limitations

The model is one corner in vertical motion. Without pitch, roll, load transfer or a seat, it has nothing to say about the heave, pitch and roll control that production active systems perform.

The tire is a point contact with a linear spring and a light damper. It does not envelop short features or bridge ripples, and grip is represented by the vertical load alone rather than by longitudinal and lateral forces. Chapter 4 models the longitudinal force against wheel slip that this chapter leaves out.

> **OPEN CHALLENGE** **Semi-Active Control Beyond Steady Driving**
>
> The bulk of semi-active control laws was designed for quasi-steady driving, with no hard acceleration, braking or steering. A semi-active suspension is fully exploited only when every working condition is considered, and much current research seeks the best handling of transients, where the damper can assist stability control, traction control and anti-lock braking [5]. Braking and cornering bring in pitch, roll and load transfer, which is also where this chapter's single corner breaks down. Progress would be a damper law that keeps the tire load up through braking and cornering as well as on a straight road.

The actuator energy counts positive mechanical work only. A hydraulic unit also spends power holding a static leveling force, so the 1373 N of case 5 costs nothing here.

Comfort is unweighted rms acceleration. It weights every frequency equally while discomfort depends on frequency, and the ISO 2631-1 words on the passenger are indicative only [16].

The preview reading is the true profile plus independent noise. A camera also misreads shadows, water and painted lines, which bias the reading rather than scatter it, and speed and distance are known exactly. The added MPC fixes nominal masses, stiffnesses, damping and report delay, omits flight and stops, and extrapolates no unseen road feature. Its disturbance estimate is not physical identification, and soft nominal contact and travel limits cannot ensure grip.

> **OPEN CHALLENGE** **Switching Without Shaking**
>
> A blended skyhook and groundhook law works well in a narrow frequency band but can pass higher vibration at other frequencies, and switching between its skyhook and groundhook phases can change the force abruptly and excite high-frequency vibration [22]. Learned controllers show the same trait. A policy rewarded for performance alone switched near its limits like a bang-bang law, which wears equipment, wastes energy and excites higher-order dynamics, and it transferred poorly from simulation to a real hydraulic machine [21]. Progress would be switching laws whose smoothness is designed in and shown on hardware. In the chapter only the valve lag and the actuator's slew limit smooth a switch.

11.9 Conclusion

This chapter studied one force serving two masses on a quarter car with a tire that cannot pull, bump stops, three kinds of strut and a passive twin on the same road. The constraint left when the strut force cancels fixes the body acceleration at 11.25 Hz, just below the 11.9 Hz wheel hop, and the running engine agrees with the linear transfer function to within 0.1 %. Over a 60 mm hump the active strut holds the body rise to 36.7 mm against 47.7 mm, and preview to 5.9 mm. No strut follows a pothole edge at 12 m/s, which asks for 1579 m/s^2 against 98.1 m/s^2 available. On the washboard at wheel hop a semi-active strut set for comfort lowers the body acceleration by a third only by leaving the road 60 times in 5 s, and under the original law on the challenge route only preview with 75 to 85 % comfort meets the target, reaching −28.5 % at 80 %. The leveling loop must stay integral-dominant, since a proportional gain above about 11 N/mm destabilizes the body at 0 % comfort. The model

has one corner, a point tire and unweighted comfort. A fuller model would add pitch and roll, a tire that envelops the road, frequency-weighted comfort and a preview sensor with biased errors. The simulator gives practice in trading competing objectives through a single actuator. The preview MPC adds a measured comfort and travel improvement at greater work, while its route and noisy-camera failures expose the limits of a contact-mode forecast.

KEY POINTS

- The strut pushes the body up and the wheel down with the same force, so it chooses only how the tire's force is shared between them.
- Just below wheel hop the body acceleration is fixed whatever the strut does, and a semi-active strut set for comfort lowers it by a third only by leaving the road 60 times in 5 s.
- A semi-active damper can only dissipate: when the request would feed energy in, its valve goes soft and the delivered force parts from the request.
- Clipped to 500 N, the road-holding force damps the wheel less the larger its motion grows, until the tire leaves the road.

Where else this applies. Volume 17, Chapter 5 sets a plain damper against reactive control, whose power flows both ways, on a heaving buoy, the same line that separates the semi-active strut here from the active one, and its predictive controller acts on a forecast of the waves as preview acts on the road.

References

[1] R. S. Sharp and S. A. Hassan, "The Relative Performance Capabilities of Passive, Active and Semi-Active Car Suspension Systems," *Proceedings of the Institution of Mechanical Engineers, Part D: Transport Engineering*, vol. 200, no. 3, pp. 219–228, 1986.

[2] R. S. Sharp and D. A. Crolla, "Road Vehicle Suspension System Design - a review," *Vehicle System Dynamics*, vol. 16, no. 3, pp. 167–192, 1987.

[3] J. K. Hedrick and T. Butsuen, "Invariant Properties of Automotive Suspensions," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 204, no. 1, pp. 21–27, 1990.

[4] D. Karnopp, M. J. Crosby, and R. A. Harwood, "Vibration Control Using Semi-Active Force Generators," *Journal of Engineering for Industry*, vol. 96, no. 2, pp. 619–626, 1974.

[5] C. Poussot-Vassal, C. Spelta, O. Sename, S. M. Savaresi, and L. Dugard, "Survey and performance evaluation on some automotive semi-active suspension control methods: A comparative study on a single-corner model," *Annual Reviews in Control*, vol. 36, no. 1, pp. 148–160, 2012.

[6] E. K. Bender, "Optimum Linear Preview Control With Application to Vehicle Suspension," *Journal of Basic Engineering*, vol. 90, no. 2, pp. 213–221, 1968.

[7] C. J. Dodds and J. D. Robson, "The description of road surface roughness," *Journal of Sound and Vibration*, vol. 31, no. 2, pp. 175–183, 1973.

[8] M. Tomizuka, "'Optimum Linear Preview Control With Application to Vehicle Suspension'—Revisited," *Journal of Dynamic Systems, Measurement, and Control*, vol. 98, no. 3, pp. 309–315, 1976.

[9] A. G. Thompson, "An Active Suspension with Optimal Linear State Feedback," *Vehicle System Dynamics*, vol. 5, no. 4, pp. 187–203, 1976.

[10] D. Karnopp, "Theoretical Limitations in Active Vehicle Suspensions," *Vehicle System Dynamics*, vol. 15, no. 1, pp. 41–54, 1986.

[11] M. Valášek, M. Novák, Z. Šika, and O. Vaculín, "Extended Ground-Hook - New Concept of Semi-Active Control of Truck's Suspension," *Vehicle System Dynamics*, vol. 27, no. 5–6, pp. 289–303, 1997.

[12] J. D. Carlson and M. R. Jolly, "MR fluid, foam and elastomer devices," *Mechatronics*, vol. 10, no. 4–5, pp. 555–569, 2000.

[13] D. Hrovat, "Survey of Advanced Suspension Developments and Related Optimal Control Applications," *Automatica*, vol. 33, no. 10, pp. 1781–1817, 1997.

[14] C. Gohrle, A. Schindler, A. Wagner, and O. Sawodny, "Design and Vehicle Implementation of Preview Active Suspension Controllers," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 3, pp. 1135–1142, 2014.

[15] H. E. Tseng and D. Hrovat, "State of the art survey: active and semi-active suspension control," *Vehicle System Dynamics*, vol. 53, no. 7, pp. 1034–1062, 2015.

[16] M. J. Griffin, "Discomfort from feeling vehicle vibration," *Vehicle System Dynamics*, vol. 45, no. 7–8, pp. 679–698, 2007.

[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] T. Zhao, L. Yang, Y. Xie, M. Ding, M. Tomizuka, and Y. Wei, "RoadBEV: Road surface reconstruction in bird's eye view," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 19 088–19 099, 2024.

[20] G. Savaia, Y. Sohn, S. Formentin, G. Panzani, M. Corno, and S. M. Savaresi, "Experimental automatic calibration of a semi-active suspension controller via Bayesian optimization," *Control Engineering Practice*, vol. 112, p. 104826, 2021.

[21] V. Wiberg, E. Wallin, A. Fälldin, T. Semberg, M. Rossander, E. Wadbro, and M. Servin, "Sim-to-real transfer of active suspension control using deep reinforcement learning," *Robotics and Autonomous Systems*, vol. 179, p. 104731, 2024.

[22] H. Lim, J. W. Lee, S.-B. Choi, and J. B. Choi, "VISKY: Virtual inertia skyhook control for semi-active suspension systems using magnetorheological dampers," arXiv:2503.12891v2 [eess.SY], 2025.

11.10 Exercises

11.1 Concept check · 2 min

The road is a sinusoid at $\omega_1 = \sqrt{k_t/m_u}$, about 11.25 Hz for the default car, and the tire stays on the road. Neglect the tire damping. What can the strut force do to the body acceleration there?

(a) Drive it toward zero with a large enough skyhook gain, since the skyhook term acts on the body alone

(b) Nothing: it is fixed at $k_t Z_r/M_s$ whatever the strut does, so only losing tire contact moves it

(c) Halve it, because at ω_1 the travel is fixed at 1.11 times the road amplitude and the strut absorbs the rest

(d) Change it only through the tire load, which at ω_1 is fixed whatever the strut does

11.2 Concept check · 1 min

The semi-active strut receives a request F_{req} with the same sign as the estimated relative velocity $\hat{v}_{rel}$, so $F_{req}\hat{v}_{rel} > 0$. What does its valve do?

(a) It goes to c_{max}, so that the damper delivers as much of the request as it can

(b) It reverses the damper force for one scan, and the residual damper absorbs the difference

(c) It goes to c_{min}, and the delivered force parts from the requested one on the force trend

(d) It holds the previous valve setting until the sign of the request changes again

11.3 Concept check · 2 min

The load-leveling loop of the active strut runs at $K_c = 5\,\mathrm{N/mm}$ and $\tau_I = 0.5\,\mathrm{s}$, an integral-dominant tuning. Why is the proportional gain kept this small?

(a) A larger K_c would leave a ride-height offset under the extra load that the integral could not remove

(b) A larger K_c would amplify the 0.5 mm ride-height noise enough to saturate the actuator

(c) A larger K_c would excite wheel hop at 11.9 Hz through the height filter and the strut

(d) Behind the 1 s height filter the proportional term lags nearly 90° at the body mode and adds negative damping

11.4 Theory and calculation · 7 min

Use the default car of Table 11.2: $m_s = 300\,\mathrm{kg}$ with the 60 kg design load aboard, $m_u = 40\,\mathrm{kg}$, $k_s = 25\,\mathrm{kN/m}$ and $k_t = 200\,\mathrm{kN/m}$. Neglect all damping.

(a) Find M_s and the body and wheel-hop frequencies from the approximations of Section 11.4. *Checkpoint:* 360 kg, 1.25 Hz and 11.94 Hz.

(b) Find the first invariant frequency $\omega_1/2\pi$ and the body acceleration per millimeter of road there. What amplitude and rms does the 7 mm washboard give at that frequency, whatever the strut? *Checkpoint:* 11.25 Hz, 0.556 m/s^2 per mm, 3.89 m/s^2 amplitude and 2.75 m/s^2 rms.

(c) Find $\omega_2/2\pi$ and the travel per millimeter of road there. At what speeds does the 1.25 m washboard reach ω_1 and wheel hop? *Checkpoint:* 3.56 Hz, 1.11, 14.1 m/s and 14.9 m/s.

11.5 Theory and calculation · 6 min

Case 2 meets a pothole of depth $D = 50$ mm whose entry edge is a half-cosine of length $e = 0.15$ m, Equation (11.5). Use $g = 9.81\,\mathrm{m/s^2}$ and the default car.

(a) Find the largest downward wheel acceleration the edge asks for, $(D/2)(\pi v/e)^2$, at 12 and 3 m/s. *Checkpoint:* 1579 m/s^2 and 98.7 m/s^2.

(b) Once the tire unloads, find the downward acceleration the spring preload and the wheel's weight give, $(F_{s0} + m_u g)/m_u$, and the value with the full 2500 N actuator pushing the wheel down as well. *Checkpoint:* 98.1 m/s^2 and 160.6 m/s^2.

(c) Find the highest speed at which the wheel can follow the edge in each case. *Checkpoint:* 2.99 m/s and 3.83 m/s.

11.6 Theory and calculation · 7 min

An engineer checks the timing of the active strut against the defaults of Table 11.2. Take $\omega_b = 2\pi(1.25\,\mathrm{Hz})$ and wheel hop at 11.94 Hz.

(a) The 1 s height filter turns the leveling gain into a negative damping of about $K_c/(\omega_b^2 \tau_h)$. Find it at 5 and 50 N/mm and compare it with the 500 N s/m residual damper. *Checkpoint:* 81 N s/m and 811 N s/m.

(b) Find the highest frequency at which a full 2500 N sinusoid can be followed under the 400 kN/s slew limit, and under 50 kN/s. *Checkpoint:* 25.5 Hz and 3.18 Hz.

(c) Find the phase the 10 ms actuator lag and the 2 ms sensor delay together cost at wheel hop. *Checkpoint:* 36.9° plus 8.6°, about 46°.

11.7 App challenge · 8 min

Open case 1, "1 · Isolated bump", at the default seed and the default 0.5× speed (a 10 s run takes 20 s).

(a) Run as shipped and pause when the Ride time tile reads 10.0 s. Read the Body acceleration rms, Tire load rms and Actuator energy tiles. *Checkpoint:* 0.885 m/s^2 (passive 1.197), 8.1 % (passive 11.0), 1 J, each within 0.005 or one unit of the last digit.

(b) Press Reset, set Road preview to On, and run again to 10.0 s. Read the same tiles and Peak travel. *Checkpoint:* 0.345 m/s^2, 4.3 %, 138 J, 61 of 80 mm.

(c) Explain why the actuator work rises more than a hundredfold while the body acceleration falls by more than half, and why preview uses more travel. *Checkpoint:* The actuator now compresses the strut over each hump.

11.8 App challenge · 8 min

Open case 5, "5 · Heavier passenger load", at the default seed and the default 0.5× speed. The extra 140 kg lands at 0.2 s.

(a) Run as shipped and pause when Ride time reads 12.0 s. Read the Bump-stop hits and Peak travel tiles. Export the CSV and read `ctrl_lev_N` in the row `time_s` = 12.00. *Checkpoint:* 0 (passive 7), 76 of 80 mm, 1370 N within 2 N.

(b) Find the static force that carries 140 kg and its share of the 2500 N rating, and compare with (a). *Checkpoint:* 1373 N, 55 %.

(c) Press Reset and, before pressing Run, set Leveling gain to 0 and Leveling integral time to 20 s. Run to 12.0 s and read the same two tiles. *Checkpoint:* 3 (passive 7), 89 of 80 mm.

11.9 App challenge · 9 min

Open the challenge, "Challenge · three surfaces", at the default seed and the default 0.5× speed. Each run stops itself 0.5 s after the route ends, at 12.8 s of simulated time.

(a) Run as shipped. Read Comfort vs passive, Wheel off the road and Challenge target. *Checkpoint:* −35.9 %, 0.122 s and 6 times, not met.

(b) Set Comfort ↔ road holding to 80 %, press Reset and run the route again. Read the same tiles. *Checkpoint:* −17.4 %, no liftoff, not met.

(c) Set Road preview to On, press Reset and run once more. Read the same tiles and Actuator energy, and say why preview is what makes the setting pass. *Checkpoint:* −28.5 %, no liftoff, met, 191 J.

11.10 Application area · 4 min

A worn passive damper has lost 60 % of its damping. In case 6 the first hump peaks at 3.83 m/s^2 against 5.90 m/s^2 for the healthy twin, and the body acceleration over 10 s is 1.234 m/s^2 against 1.197 m/s^2. In the bounce test the healthy body falls from 51.8 to 9.9 mm in one cycle and the worn one from 62.2 to 32.9 mm.

(a) From the bounce test, find the ratio of successive peaks and the damping ratio $\zeta = \delta/\sqrt{4\pi^2+\delta^2}$ with $\delta = \ln(x_1/x_2)$ for each damper. *Checkpoint:* 0.19 and 0.53, ζ 0.25 and 0.10.

(b) Why would a driver or a comfort rating fail to flag this damper, and which measurement would? *Checkpoint:* The comfort number moves by 3 %. Measure the decay or the lowest tire load.

11.11 Application area · 3 min

A production car reads the road ahead with a camera. In case 8 a clean reading gives 0.76 m/s^2 rms against 0.90 m/s^2 without preview. After spray makes the reading noisy the preview gives 1.22 m/s^2, as bad as the passive twin, against 1.01 m/s^2 with preview off, and the preview force rises from 242 N to 740 N rms.

(a) What should a supervisor do when the reading degrades, and which signal available to the controller could tell it so without knowing the true road? *Checkpoint:* Fall back to feedback or smooth more. Watch the preview force.

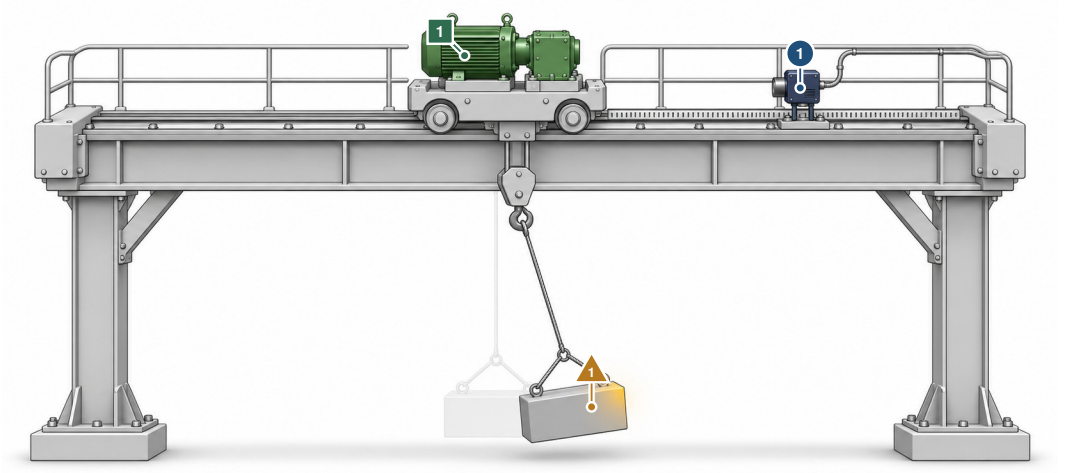
(b) Which camera error does the simulator leave out, and why would a wider smoothing window not cure it? *Checkpoint:* Biased misreadings such as shadows, water and painted lines.

11.12 Review the agent · 5 min

An AI assistant tunes the challenge. It writes: "In case 9 the MPC halves the body acceleration of skyhook with preview, 0.155 m/s^2 against 0.315 m/s^2, with no liftoff. Its model also keeps a band on the tire load. So keep 80 % comfort and preview On, select MPC as the Active law, and the challenge will pass with a far larger margin." Find the flaw, and say how you would check the claim on the simulator.

CHAPTER 12

Input Shaping on an Overhead Crane Where One Actuator Serves Two Outputs



The controller moves 1 trolley travel drive

The controller reads 1 trolley position

What disturbs it 1 hoist pay-out, payload mass and wind on the load

One drive moves the trolley, only its position is read, and the timing of each move decides whether the hanging load arrives swinging or still. Illustration, not to scale.

IN PLAIN TERMS **Moving a Crane Without Setting the Load Swinging**

The main idea. A crane trolley rolls along a beam with a load hanging from a cable below it. One motor must stop the trolley over its target and leave the load still. Nothing measures the swing.

The challenge. A quick move sets the load swinging like a pendulum, and the swing barely dies away. Splitting each move into a few timed pieces lets the later pieces cancel the swing the first one started. The timing must match the length of the cable.

Why it matters. Every second spent waiting for a load to stop swinging is lost again on every lift the crane makes.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact in the Introduction (Section 12.1), the history in Section 12.2, the fact and the try-its in Section 12.6, 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 12.6), Technology in Practice (Section 12.7), Limitations (Section 12.8) and the open challenge

The control question One drive on a gantry crane trolley has to park the trolley and leave the hanging load still, and no controller reads the swing. How much sway does a shaped command remove, what does it lose when its assumed cable length or the drive rating is wrong, and how does a predictive controller on the same model compare?

Learning objectives

- Derive the trolley and payload balances, and explain why a sudden push accelerates the trolley at F/M.
- Compute the impulse amplitudes and times of the ZV and ZVD shapers from an assumed cable length and damping ratio.
- Estimate the residual sway of a mistuned shaper from the ratio of true to design frequency, including the coupled mode of the closed loop.
- Show from runs why shaping fails on a saturated drive and on sway that is already present.
- Commission a crane move in the order damping, gain, shaping.

Prerequisites Sections 1.4.1 (the hanging cart), 1.4.4 (eigenvalues), 1.4.8 (input shaping) and 1.6.2 (velocity form). Section 1.6.3 is needed only for the predictive controller. Chapter 7 for a pendulum moved by its cart.

The app https://apmonitor.com/apps/overhead-crane/

See also Volume 3, Chapter 12 for two-impulse and three-impulse shapers on a ribbon mode under a speed loop. Volume 9, Chapter 4 for an offset-free MPC whose Kalman filter carries a disturbance state, and Volume 9, Chapter 12 for a point-to-point move planned by optimal control.

12.1 Introduction

A control problem is normally posed with one actuator behind every quantity that has to be held somewhere. A gantry crane does not offer that arrangement. The trolley carries one drive, the operator requires both that the trolley stop at a target and that the load beneath it be still when it arrives, and the swing angle has no actuator attached to it at all. The swing is excited or canceled only through the history of the trolley acceleration, which is the same signal the position loop is using for its own purpose. The plant carries four mechanical states, one input and two outputs of interest. Chapter 11 meets the same arrangement in a strut that pushes the body and the wheel with one force.

The system is not an inverted pendulum. The load hangs below the trolley and requires no balancing, unlike the upright pole of Chapter 7, which also responds only to the acceleration of its cart. It is the hanging cart of Section 1.4.1. The difficulty is that the load is a resonator with almost no damping, at a frequency $\omega_n = \sqrt{g/\ell}$ that a hoist changes every time it runs, and a point-to-point move whose duration is a few multiples of the swing period excites it strongly. On the default configuration of the application, a 6 m move that parks the trolley inside ± 0.10 m at 9.88 s leaves the payload swinging with an amplitude of 7.87°, and the swing is still 2.81° at $t = 20$ s. The position loop meets its own specification. It has no access to the quantity that is still moving.

Two remedies exist and they are structurally different. The first adds a sensor and closes a second loop around the angle, which is what a nonlinear coupling controller or a state-feedback anti-sway system does. The second leaves the loop untouched and changes what the loop is asked for, convolving the reference with a short train of impulses chosen so that the vibration the later impulses start cancels the vibration the earlier ones left. Command shaping is feedforward in the strict sense. It costs nothing in stability margin, it can be retrofitted to a crane whose drive controller is a sealed box, and it carries a specific and testable weakness, because a canceler tuned to the wrong frequency does not cancel. The absence of an angle measurement is a property of the instrumentation rather than of the machine, and it is what makes command shaping the appropriate tool on this apparatus, because a mistuned shaper announces itself only through the residual sway it leaves behind.

The experiments of this chapter isolate a specific set of process dynamics and control principles. The plant is underactuated, so the count of independent actuators against the count of regulated outputs decides what feedback can reach, and the residual sway is the part of the response no gain

touches. Command shaping separates feedforward action on the reference from feedback action on the error, because the impulse train changes the setpoint the loop follows and leaves every controller coefficient alone. Resonance, damping ratio and the sensitivity of a canceler to a frequency error become measurable rather than asserted. Actuator saturation is a failure class distinct from poor tuning, because a drive resting on its limit no longer follows the command it was given. The physical fundamentals are a horizontal momentum balance on the trolley and a moment balance on the suspended load, coupled through the cable tension, together with the time-scale separation between a position loop that has to be slow and a payload resonance that is not. Those are the same conservation principles and the same loop structures that govern process applications built on material and energy balances [1–3], in a setting where the geometry is directly observable. The behavior matters industrially because sway is paid for in the seconds an operator spends waiting for a load to settle, on every cycle of a machine whose cycle time sets the throughput of the facility around it.

> ✦ **DID YOU KNOW? Waiting for the Load at a Container Quay**
>
> On ship-to-shore cranes without anti-sway, the operator waits on average about 15 s for the load to settle at each end of a move. At about 19 containers an hour per crane, anti-sway was estimated to add about two [4].

Textbook accounts of input shaping give the impulse amplitudes and a sensitivity curve and then move on to an application example [5]. A printed sensitivity curve does not quantify what the same shaper loses when four meters of cable are paid out, what a longer shaper recovers, or what neither of them can repair once the drive is on its limit. This chapter puts numbers on what command shaping is worth on a simulated overhead crane as the cable length, the payload mass, the drive rating and the origin of the sway are changed one at a time. Input shaping and the crane model are both established, and the numbers are the subject.

Section 12.2 traces the problem from the crane operator's timed two-pulse command to the robust shapers. Section 12.3 carries every relation and every constant needed to rebuild the plant: its state equations, parameter values, hard limits and sample periods. Section 12.4 shows that the frequency a shaper on this machine ought to be tuned to is not $\sqrt{g/\ell}$ at all but the coupled mode of trolley and payload with the position loop closed. A fifth-order linearization reproduces that mode to within 0.1 % of the period timed in the simulator at three separate configurations, names the assumed cable length that minimizes residual sway to the 0.25 m resolution of a measured sweep, and accounts for the position loop damping the payload four times faster than the cable pivot manages on its own. The same section derives the two shapers and their sensitivity to a frequency error. Section 12.5 gives the discrete algorithm the controller executes and a model predictive controller given the shaper's cable length. Section 12.6 scores thirteen configurations on the numbers the application computes while it runs, and sets the predictive controller against the shapers in Section 12.6.5. Section 12.7 describes the sensors, drives and controllers of real cranes, Section 12.8 states the limitations and Section 12.9 concludes.

12.2 Background

Crane operators solved this problem by hand long before it was written down. A skilled driver of a bridge crane accelerates, coasts, and then decelerates at a moment chosen so that the load arrives over the target with the cable vertical, which is a two-pulse command timed against the pendulum period and is exactly the maneuver an automatic shaper performs. The cost of doing it by eye is that the operator must watch the load rather than the destination, and that every new load, every new cable length and every new driver produces a different result.

The control formulation arrived from a different direction, from Posicast control to the robust shapers, as the history segment below tells. Direct comparison of the robust shapers on common ground establishes what each one costs in delay [6].

> ⧖ HISTORY **From Posicast to Robust Shapers, 1957 to 1994**
>
> Posicast control, published in 1957, removed the ringing of a lightly damped feedback system by exciting several transient oscillations at closely spaced times, with magnitudes and phases chosen so that their sum is zero. Split into two spaced excitations, a step settled in one half cycle, and in an electrical feedback system the extra hardware was one or two artificial transmission lines [7]. Preshaped command inputs, published in 1990, completed a move without residual vibration for a time penalty on the order of one period of the first mode. They were robust to parameter uncertainty, could be applied open or closed loop, and cut the residual vibration of a simulator of the Space Shuttle's robot arm by a factor of 25 [8]. A vector-diagram method of 1994 needed only estimates of the frequency and the damping, and given a bound on the acceptable residual it chose a shaper insensitive to frequency errors [9].

Cranes became the standard test case for the method. Modeling and control of cranes has its own review literature, covering gantry, bridge, boom and tower configurations, the planar and three-dimensional models, and the split between open-loop and closed-loop approaches [10, 11]. Hoisting during a traverse is the difficult case, because the natural frequency sweeps while the shaper runs, and measurements on an industrial gantry quantify what a fixed shaper loses and what recomputing the impulse train from a live cable length recovers [12]. Volume 3, Chapter 12 applies two-impulse and three-impulse shapers to a structural mode of a ribbon under a climber. Combining a shaper on the command path with a positioning loop, so that a crane both parks accurately and arrives quiet, has been demonstrated on production bridge and gantry machines [13]. The closed-loop alternative develops in parallel, with nonlinear coupling laws that feed the swing angle back into the trolley force and prove asymptotic regulation of both coordinates [14], and the general problem of controlling a mechanical system with fewer actuators than degrees of freedom has become a field of its own [15].

The economic motive is throughput. A container terminal is a queue of ship-to-shore and yard cranes whose cycle time sets the berth productivity of the whole facility, and the operations research literature on terminal design treats crane cycle time as the primary resource [16]. Sway lengthens that cycle, because an operator who cannot rely on the load being still must slow down, wait, or make a correcting move. Controlled experiments with human operators driving an instrumented bridge crane through an obstacle course, with input shaping switched on and off, report faster and safer completion with shaping enabled [17]. The improvement is not confined to the machine, because the operator no longer has to compensate for the swing.

12.3 Crane Model

The simulated machine is a single trolley on a horizontal rail with one point-mass payload hanging beneath it on one cable, confined to the vertical plane that contains the rail. Figure 12.1 collects the geometry, the five states, the drive force, the three disturbances and the one instrument.

12.3.1 Signal Classification and Ranges

Two variables are declared as controlled variables and the shipped controller regulates only one of them. The trolley position is held against an operator setpoint by the single feedback loop of that controller. The swing angle is declared with a setpoint of 0° and a band of 0.5° so that it is trended and scored beside the position loop, and the shipped controller never reads it. The same drive force is reported under both, because one actuator serves both. Table 12.1 sorts each signal by the part it plays and gives the interval over which the application allows it to move.

12.3.2 Equations of Motion

The trolley obeys a horizontal momentum balance in which the payload pushes back through the cable, and the payload obeys a moment balance about the suspension point, already divided through by $m\ell$. They are given in Equation (12.1).

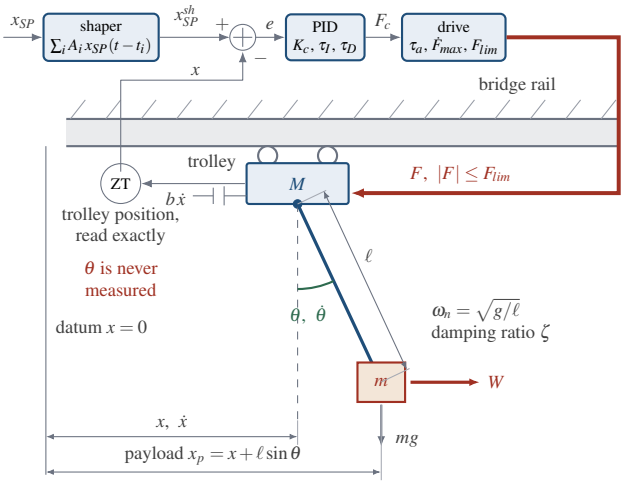


Figure 12.1 Overhead crane geometry with the command path drawn above it. Each symbol appearing in the two balances, in the impulse train and in the control law is labeled. The trolley position is the only instrument the shaped loop reads, and no path whatever leads from the swing angle back to the drive through that loop.

Table 12.1 Signals, Their Roles and Their Ranges

Symbol	Role	Unit	Range or default
x	CV, position loop	m	setpoint 0 to 24, band ± 0.10
θ	CV, MPC objective only	deg	setpoint 0, band ± 0.5
F	MV, position loop	N	$\lvert F\rvert \le F_{lim}$
x	state, trolley position	m	free
$\dot{x}$	state, trolley velocity	m/s	free
θ	state, swing angle	rad	$\lvert\theta\rvert \le \theta_{lim}$
$\dot{\theta}$	state, swing rate	rad/s	free
F	state, force developed	N	$\lvert F\rvert \le F_{lim}$
$\Delta\ell$	DV, hoist paid out	m	-2.5 to $+8$, default 0
Δm	DV, payload mass change	%	-60 to $+150$, default 0
W	DV, wind on the payload	N	-250 to $+250$, default 0
$\Delta\dot{\theta}$	DV, bump event	rad/s	$+0.09$
$\Delta\dot{\theta}$	DV, gust event	rad/s	-0.14
$\Delta\ell$	DV, pay-out event	m	$+2$, capped at $+8$
f_{man}	operator actuator	N	-4000 to $+4000$
θ_m	measurement, MPC option	deg	off by default
ℓ_a	shaper design length	m	1.5 to 12, default 4.00
ζ_a	shaper design damping	—	0 to 0.100, default 0.015

$$\begin{aligned}(M+m)\ddot{x} + m\ell\ddot{\theta}\cos\theta - m\ell\dot{\theta}^2\sin\theta + b\dot{x} &= F + W,\\ \ell\ddot{\theta} + \ddot{x}\cos\theta + g\sin\theta + 2\zeta\omega_n\ell\dot{\theta} &= \frac{W}{m}\cos\theta\end{aligned} \tag{12.1}$$

where M is the trolley mass, m the payload mass, ℓ the cable length in use, b the lumped rail and gearbox friction coefficient, ζ the sway damping ratio, g the gravitational acceleration, F the force the drive has developed and W a horizontal wind force on the payload. The payload hangs at $(x + \ell\sin\theta,\ -\ell\cos\theta)$ relative to the trolley, so a positive θ means the load is ahead of the trolley in the $+x$ direction, and a forward push therefore drives θ negative. The wind appears in both balances

because the payload can move both by the trolley translating and by the cable swinging, and its generalized forces are $Q_x = W$ and $Q_\theta = W\ell\cos\theta$. The last term of the second relation is written as $2\zeta\omega_n\ell\dot\theta$ so that ζ is the payload damping ratio directly, which is the number the shaper formulas need.

The coupling is not linearized. The two accelerations are collected on the left and solved as the two-by-two system of Equation (12.2) at every right-hand-side evaluation.

$$\begin{bmatrix} M+m & m\ell c \\ c & \ell \end{bmatrix} \begin{bmatrix} \ddot x \\ \ddot\theta \end{bmatrix} = \begin{bmatrix} F + W + m\ell\dot\theta^2 s - b\dot x \\ -gs - 2\zeta\omega_n\ell\dot\theta + (W/m)c \end{bmatrix} \tag{12.2}$$

where $s = \sin\theta$ and $c = \cos\theta$ have been abbreviated to keep the display inside the column.

The determinant is $\ell(M + m\sin^2\theta)$, which is strictly positive for any physical cable length, so the solve is unconditionally well posed. The solution at $\theta = 0$ states the character of the plant. The trolley acceleration is F/M, not $F/(M+m)$. A load on the end of a cable cannot resist a sudden push and contributes nothing to the trolley inertia until it has started to swing, and it is only at low frequency, where the payload follows the trolley, that the combined mass $M+m$ is the right one.

Pause. Try Exercise 12.2 before reading on.

The drive is a force source with a first-order lag, and the force it is commanded to produce is rate limited twice, once inside the controller and again on the applied command, as shown in Equation (12.3).

$$\begin{aligned} \tau_a\dot F &= F_a - F, \qquad |\dot F_a| \le \dot F_{max}, \\ F_{lim} &= \min\bigl(F_{max},\,(M+m)\,a_{max}\bigr) \end{aligned} \tag{12.3}$$

where τ_a is the drive time constant, F_a the applied command and $\dot F_{max}$ the slew limit. At the default parameters $(M+m)a_{max} = 1350\times 4.0 = 5400\,\mathrm{N}$ exceeds the motor rating, so $F_{lim} = F_{max} = 4000\,\mathrm{N}$ and the acceleration limit is inactive until a scenario derates the drive or a heavy payload is hung on it.

The natural frequency and period of the suspended load, and the damped frequency the shaper timings use, are given by Equation (12.4).

$$\omega_n = \sqrt{\frac{g}{\ell}}, \qquad T_n = 2\pi\sqrt{\frac{\ell}{g}}, \qquad \omega_d = \omega_n\sqrt{1-\zeta^2} \tag{12.4}$$

Neither mass appears. At the default 4.00 m cable this is $\omega_n = 1.5660\,\mathrm{rad/s}$ and $T_n = 4.012\,\mathrm{s}$, and because $\zeta = 0.015$ the damped and undamped frequencies differ by 0.011 %. The damping enters the shaper almost entirely through the impulse amplitudes and hardly at all through the impulse timings.

The instantaneous sway amplitude reported by the application is the phase-plane radius of Equation (12.5).

$$A = \sqrt{\theta^2 + \left(\frac{\dot\theta}{\omega_n}\right)^2} \tag{12.5}$$

For free oscillation $\theta = A\sin(\omega_n t + \phi)$ this is exact at every instant, so the amplitude is readable the moment a move ends rather than after a further full period, which is what makes a residual-sway number meaningful on a run that lasts only a few periods.

12.3.3 Parameters, Limits and Numerics

Every constant the simulation uses anywhere appears in Table 12.2, grouped as the plant, the drive, the loop with its shaper and the predictive controller, the limits and alarms, and the numerics, except the fixed decision period, weights, prices and noise variances of the predictive controller, which

Table 12.2 Every Constant in the Crane Simulation and Its Adjustable Range

Symbol	Meaning	Value	Adjustable range
M	trolley mass	1200 kg	400 to 3000
m	payload mass, nominal	150 kg	25 to 800
ℓ	cable length, nominal	4.00 m	1.5 to 12
b	trolley damping	400 N s/m	0 to 1500
ζ	sway damping ratio	0.015	0 to 0.100
g	gravity	9.81 m/s^2	1.6 to 24.8
F_{max}	drive rating	4000 N	400 to 4000
a_{max}	acceleration limit	4.0 m/s^2	0.3 to 6.0
$\dot{F}_{max}$	force slew limit	12 000 N/s	1000 to 40000
τ_a	drive time constant	0.08 s	0 to 0.40
F_{push}	manual push force (hidden)	2400 N	200 to 4000
t_{push}	manual push duration (hidden)	0.8 s	0.1 to 3.0
K_c	position loop gain	620 N/m	40 to 1800
τ_I	integral time	1000 s	5 to 2000
τ_D	derivative time	2.3 s	0 to 6
τ_f	derivative filter	0.02 s	0 to 0.50
ℓ_a	shaper design length	4.00 m	1.5 to 12
ζ_a	shaper design damping	0.015	0 to 0.100
T_p	MPC prediction horizon	6 s	2 to 10
θ_{max}	MPC swing limit	3°	0.5 to 15
x_{max}	MPC travel limit	24 m	1 to 24
x_{tol}	parked tolerance	0.10 m	0.02 to 0.50
θ_{ok}	quiet-sway threshold	0.5°	0.1 to 3.0
θ_{alarm}	excessive-sway alarm	12°	3 to 40
θ_{lim}	cable slack limit (hidden)	1.4 rad	0.5 to 1.5
Δt	integration step (RK4)	0.005 s	fixed
Δt_c	control period	0.01 s	fixed
f_{log}	historian rate	50 Hz	fixed
N_h	setpoint history depth	4096 samples	fixed

The setpoint history holds 40.96 s at the control rate, against a longest shaper span of 6.95 s at the 12 m maximum cable. Random draws are seeded at 20260824, although no scenario in this application injects noise.

are given with its design in Section 12.5.4. Three of them are not exposed for adjustment in the application. One of the three, the swing limit, halts the integration of an unphysical angle rather than continuing to simulate a cable that has gone slack.

The engine of Section 1.6.1 integrates the five states on a fixed 5 ms grid, roughly 800 steps per swing period at the default cable length. Control runs at 100 Hz, the rate a crane drive programmable logic controller would use, on every second step, and the historian samples at 50 Hz. Scheduled events run first in a step, so a setpoint step scheduled at 1.5 s reaches the controller scan at 1.5 s. The post-step block holds θ at θ_{lim} with $\dot{\theta}$ set to zero and raises a swing-limit fault when the angle exceeds 1.4 rad, clearing again at 90 % of that value, sets the developed force equal to F_a when τ_a is set to zero, evaluates Equation (12.5), tests $|x - x_{SP}| \le x_{tol}$ against the *raw* setpoint rather than the shaped one, and updates the residual-sway statistic only while that test passes. Every scenario initializes the trolley parked at its setpoint with the drive idle and, except configuration 9, the cable vertical, so the crane starts in an exact equilibrium and every move in the application is a genuine setpoint change for the shaper to act on.

The cable length in use is $\ell = \mathrm{clamp}(\ell_0 + \Delta\ell, 0.4, 24)$ and the payload mass in use is $m = m_0(1 + \Delta m/100)$, so the hoist disturbance changes the resonance and the mass disturbance does not.

12.3.4 Assumptions

Point mass on a rigid, massless, always taut cable. A real hook block and sling is a double pendulum with a second and faster mode that no single-mode shaper touches, and control of the double-pendulum case needs a shaper designed for two frequencies at once. Cable stretch, load rotation and the inertia of the load about its own center are all absent.

Constant cable length during a move. Cranes hoist and traverse at the same time, which sweeps the natural frequency continuously and defeats any fixed impulse train. The hoist here is a disturbance that steps between moves rather than a second axis running concurrently, and the measured penalty for a fixed shaper under a real hoisting profile is larger than anything reported below [12].

Planar motion. The bridge axis at right angles to the trolley axis, the out-of-plane sway mode it excites and the interaction between the two are all omitted, as is any wind component across the rail.

Linear viscous sway damping. The loss term is $2\zeta\omega_n\ell\dot{\theta}$ with $\zeta = 0.015$, which gives an amplitude decay time of $1/(\zeta\omega_n) = 42.6\,\mathrm{s}$ at 4 m with the trolley held. Real losses are dominated by aerodynamic drag and grow faster than linearly with swing rate, so the model overstates how long a large swing persists and understates how long a small one does.

An idealized drive. The force path is a rating, a slew limit and a first-order lag. There is no gearbox backlash, no torque ripple, no stiction and no breakaway friction, and because the rail loss is a single viscous coefficient the trolley creeps for any nonzero force.

Perfect position and no angle measurement. The trolley position is read exactly, with no noise, bias, quantization or delay, and the swing angle is not instrumented at all unless the optional sensor of the predictive controller is switched on, which reads it just as exactly. The second half of that is a modeling choice rather than a limitation, and it is the point of the chapter. Adding an angle measurement changes the problem from command shaping to genuine anti-sway feedback [14].

The fixed-pivot formula is an approximation. Equation (12.4) is the frequency of a pendulum on a rigid pivot. The trolley is not rigid, which Section 12.4 quantifies.

12.4 Dynamic Analysis

12.4.1 Position Loop Bandwidth

With the payload following the trolley at low frequency and the derivative action supplying the only damping the loop has, the position loop is the second-order system of Equation (12.6).

$$\omega_c = \sqrt{\frac{K_c}{M+m}}, \qquad \zeta_c = \frac{b + K_c\tau_D}{2\sqrt{K_c(M+m)}} \tag{12.6}$$

At the default tuning this places the loop at $\omega_c = 0.678\,\mathrm{rad/s}$ with $\zeta_c = 1.00$, which is well damped and, more to the point, a factor of 2.3 slower than the payload resonance at 1.566 rad/s. That separation is deliberate. Placing the loop at twice the sway frequency would need $K_c = 4\omega_n^2(M+m) = 13.2\,\mathrm{kN/m}$ and a 6 m step would then ask the drive for 79 kN against a 4 kN rating. The gap between the achievable loop bandwidth and the payload resonance is the reason command shaping exists on this plant. The swing is not removed by raising the gain.

Integral action is effectively disabled at the default $\tau_I = 1000\,\mathrm{s}$, because the plant has no steady-state offset to remove. At rest on a level rail, zero force holds any position. Only a wind force on the payload leaves a sustained offset for the integral term to remove.

12.4.2 The Closed-Loop Sway Mode

Linearizing Equation (12.1) about $\theta = 0$, closing the position loop with gain K_c and velocity feedback $K_c\tau_D$, and retaining the drive lag gives a fifth-order system in $(x, \dot{x}, \theta, \dot{\theta}, F)$. Setting the drive lag aside for a moment, eliminating x between the two linearized balances gives the quartic of Equation (12.7).

$$\begin{aligned} M\ell s^4 + \big[(M+m)d + c\ell\big]s^3 + \big[(M+m)g + cd + K_c\ell\big]s^2 \\ + \big[cg + K_c d\big]s + K_c g = 0 \end{aligned} \tag{12.7}$$

where $c = b + K_c\tau_D$ is the total velocity feedback seen by the trolley and $d = 2\zeta\omega_n\ell$ is the pivot damping coefficient. Two of the four roots are the position loop and two are the sway mode. Three limiting cases bracket the sway root. A rigidly held trolley gives $\sqrt{g/\ell}$, or 4.012 s at 4 m. A completely free trolley gives the free-cart value $\sqrt{g(M+m)/(M\ell)}$, or 3.783 s, because the trolley recoils as the load swings and shortens the period. A trolley held by a finite spring lies between the two, and K_c decides where.

Restoring the drive lag turns the quartic into the fifth-order system of Equation (12.8) in the states $(x, \dot{x}, \theta, \dot{\theta}, F)$.

$$\begin{bmatrix} M+m & m\ell \\ 1 & \ell \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} F - b\dot{x} \\ -g\theta - d\dot{\theta} \end{bmatrix}, \qquad \tau_a \dot{F} = -K_c x - K_c\tau_D \dot{x} - F \tag{12.8}$$

The control law has been linearized about a stationary setpoint with the integral term dropped, which is exact at $\tau_I = 1000$ s over the timescale of one swing, and the derivative filter is dropped as well because $\tau_f = 0.02$ s is two decades faster than anything else in the loop. Solving the eigenproblem at the default tuning gives a sway mode with a period of 3.8835 s and a damping ratio of 0.0579. The simulator, driven by the standard 6 m move with the shaper off and timed over thirteen zero crossings of the free decay that follows $t = 12$ s, returns 3.884 s and an amplitude decay time of 10.38 s against the predicted 10.65 s. The same calculation at 8 m predicts 5.5929 s and 15.22 s against a measured 5.5974 s and 15.02 s, and at 300 kg it predicts 3.7445 s and 6.10 s against a measured 3.743 s and 5.90 s.

Two consequences follow. The first is that the true sway frequency is 3.3 % above the $\sqrt{g/\ell}$ that the shaper formulas use, which is small but, as Section 12.4.4 shows, is not small compared with the tolerance of a two-impulse shaper. The second is that the closed position loop damps the payload. The sway damping ratio rises from the plant's 0.015 to 0.058, a factor of 3.9, because a swinging load pushes the trolley with a horizontal force of about $mg\theta$, the loop resists that push, and the drive dissipates the energy. The effect is incidental to the shaper and is the only mechanism the shaped loop has for removing sway that is already present.

12.4.3 Impulse Amplitudes and Times

Section 1.4.8 derives both shapers for an undamped mode. This section keeps the sway damping, which the formulas of this crane carry. An impulse of amplitude A_i applied at time t_i to a second-order mode of frequency ω and damping ratio ζ leaves, at any common time $t_n \geq \max_i t_i$, a sinusoidal residue whose vector magnitude is $A_i e^{-\zeta\omega(t_n - t_i)}$ and whose phase is $\omega_d t_i$. A sequence of impulses leaves the vector sum, so the residual vibration relative to what a single unit impulse would leave is Equation (12.9).

$$V(\omega,\zeta) = e^{-\zeta\omega t_n}\sqrt{C^2 + S^2}, \qquad C = \textstyle\sum_i A_i e^{\zeta\omega t_i}\cos(\omega_d t_i), \quad S = \sum_i A_i e^{\zeta\omega t_i}\sin(\omega_d t_i) \tag{12.9}$$

For two impulses at $t_1 = 0$ and $t_2 = \Delta$, setting $S = 0$ requires $\sin(\omega_d\Delta) = 0$, whose smallest positive solution is $\Delta = \pi/\omega_d$, one damped half period. Setting $C = 0$ then requires $A_1 = A_2 e^{\zeta\omega\Delta}$, and imposing $A_1 + A_2 = 1$ so that the shaped command reaches the same final value completes the zero-vibration shaper of Equation (12.10).

$$\mathbf{A}_{ZV} = \frac{1}{1+K}\begin{bmatrix} 1 \\ K \end{bmatrix}, \qquad \mathbf{t}_{ZV} = \begin{bmatrix} 0 \\ \pi/\omega_d \end{bmatrix} \tag{12.10}$$

where $K = \exp(-\zeta\pi/\sqrt{1-\zeta^2})$ is the factor by which the free response decays over one damped half period. Requiring in addition that $\partial V/\partial\omega = 0$ at the design frequency adds one constraint and

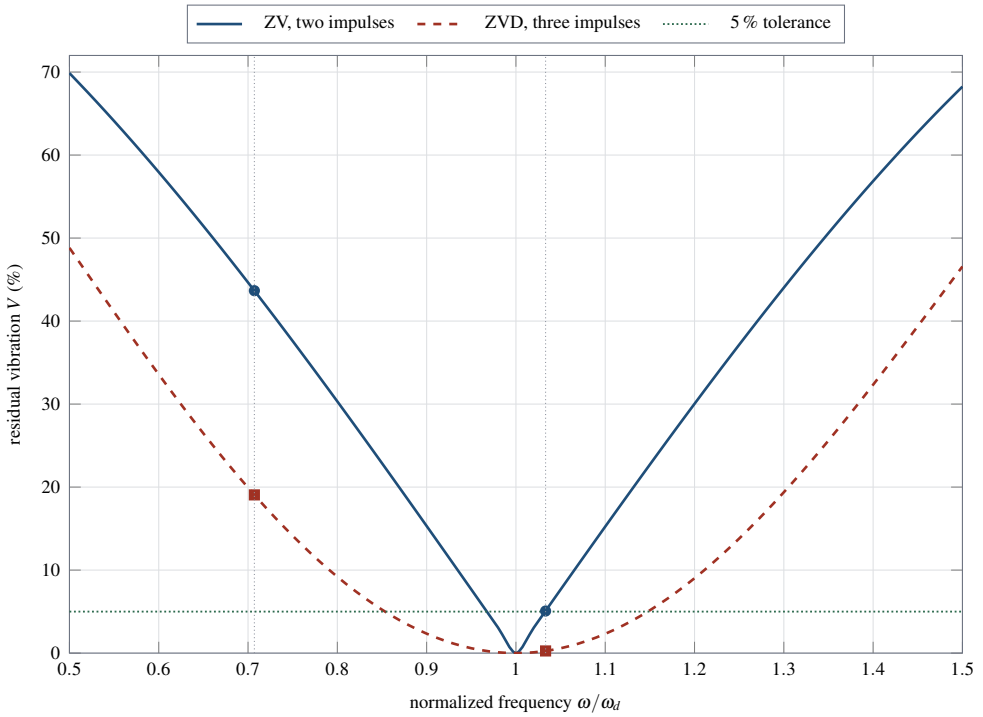


Figure 12.2 Residual vibration left by the two shapers as implemented, against plant frequency normalized to the design frequency, at $\zeta_a = 0.015$. Two impulses are solid with circles, three impulses dashed with squares, and the dotted horizontal is the 5 % tolerance. The dotted verticals are the two cases quantified in Section 12.6, the true coupled mode at $\omega/\omega_d = 1.0332$ when the shaper is given the nominal 4.00 m, and 0.707 for an 8 m cable with the shaper still set to 4 m.

therefore one impulse, and the solution is the zero-vibration-and-derivative shaper of Equation (12.11), which is the convolution of the two-impulse shaper with itself.

$$\mathbf{A}_{ZVD} = \frac{1}{(1+K)^2}\begin{bmatrix}1\\2K\\K^2\end{bmatrix}, \qquad \mathbf{t}_{ZVD} = \begin{bmatrix}0\\\pi/\omega_d\\2\pi/\omega_d\end{bmatrix} \tag{12.11}$$

The shaped command is the convolution of the raw setpoint history with the impulse train, as shown in Equation (12.12).

$$x_{SP}^{sh}(t) = \sum_i A_i\, x_{SP}(t - t_i), \qquad \sum_i A_i = 1 \tag{12.12}$$

Because the amplitudes sum to unity the trolley still arrives at the same place. It simply gets there one damped half period later for the two-impulse shaper and two for the three-impulse one, and that delay is the price of the cancellation rather than an implementation artifact. At the default $\zeta_a = 0.015$ and $\ell_a = 4.00\,\text{m}$ the impulse train is $[0.5118, 0.4882]$ at $[0, 2.006]$ seconds for the two-impulse shaper and $[0.2619, 0.4997, 0.2384]$ at $[0, 2.006, 4.013]$ seconds for the three-impulse one, because $K = 0.9540$. At an 8 m cable the half period grows to 2.837 s and the amplitudes do not change at all, because they depend only on ζ_a.

12.4.4 Residual Vibration Sensitivity

Evaluating Equation (12.9) with the impulse trains of Equations (12.10) and (12.11) held fixed at their design values while the plant frequency is swept produces the two curves of Figure 12.2. Both vanish at the design frequency. The two-impulse curve leaves it linearly and the three-impulse curve leaves it quadratically, which is the whole content of the extra derivative constraint. Holding the residual below 5 % of the unshaped value requires the plant frequency to lie within 0.968 to 1.033 of the design value for the two-impulse shaper, a total width of 0.065, and within 0.854 to 1.148 for the three-impulse one, a width of 0.294. Insensitivity improves by a factor of 4.5 for one extra damped half period of delay.

Two predictions follow immediately, and both are tested in Section 12.6. Entering the nominal 4.00 m into a shaper is not the best available choice, because the true coupled mode sits at 1.0332 of that design frequency, just outside the two-impulse tolerance band, where Figure 12.2 gives $V = 5.07\,\%$ against $V = 0.26\,\%$ for three impulses. The assumed length that puts the design frequency on the true mode is $\ell_a = g/\omega^2 = 3.75\,\mathrm{m}$, not 4.00 m. Doubling the cable to 8 m without retuning divides the plant frequency by $\sqrt{2}$, which lands at 0.707 where the curves give 43.7 % and 19.1 %. A two-impulse shaper at that frequency error removes 56 % of the residual the unshaped command would leave instead of all of it, and a three-impulse shaper removes 81 %.

12.5 Controller Design

12.5.1 The Discrete Algorithm

The whole control path runs to completion every $\Delta t_c = 0.01\,\mathrm{s}$ in both manual and automatic. The sequence is

1. push the current raw setpoint x_{SP} onto a delay line of 4096 samples stamped with the simulation time,
2. rebuild the impulse train from ℓ_a, ζ_a and g using Equation (12.10) or (12.11), or the single unit impulse at zero delay when the shaper is off,
3. form $x_{SP}^{sh} = \sum_i A_i\, x_{SP}(t - t_i)$ by sampling the delay line at each impulse time with linear interpolation,
4. evaluate F_{lim} from Equation (12.3) for the current payload mass and push the current tuning and limits into the loop,
5. in automatic, run the loop of Equations (12.14) and (12.15) on the error $e = x_{SP}^{sh} - x$ to get F_c,
6. in manual, form $F_c = \mathrm{clamp}(f_{man} + n_{key} F_{push}, \pm F_{lim})$ and call the loop in tracking mode against the same shaped reference,
7. slew limit the applied command, which gives $F_a \leftarrow \mathrm{slew}(F_c, F_a, \dot{F}_{max}, \Delta t_c)$,
8. raise the saturation flag when $|F_c| \geq F_{lim}$.

The delay line is seeded at $t = 0$ with the initial setpoint, and sampling before the oldest record returns that oldest value, so the pre-history the convolution assumes is genuinely constant and the first commanded step is shaped correctly. Because the initial setpoint is also the initial trolley position, the crane is in equilibrium while that pre-history is being written. Under the predictive controller of Section 12.5.4, step 5 takes F_c from the plan and the loop runs in tracking mode as in step 6.

12.5.2 Velocity Form Position Loop

The position loop computes an increment to the drive force each control interval rather than the force itself, which makes the algorithm the incremental, or velocity, form of the PID law [1, 18] of Section 1.6.2. The derivative acts on a filtered measurement rather than on the error, so a setpoint move produces no derivative kick. The error, the filtered position and the three increments are given by Equations (12.13) and (12.14).

$$e_k = x^{sh}_{SP,k} - x_k, \qquad \hat{x}_k = (1-a)\,x_k + a\,\hat{x}_{k-1} \tag{12.13}$$

$$\begin{aligned} \Delta u^P_k &= K_c\,(e_k - e_{k-1}) \\ \Delta u^I_k &= \frac{K_c}{\tau_I}\,e_k\,\Delta t_c \\ \Delta u^D_k &= -K_c\,\tau_D\left(\dot{\hat{x}}_k - \dot{\hat{x}}_{k-1}\right), \quad \dot{\hat{x}}_k = \frac{\hat{x}_k - \hat{x}_{k-1}}{\Delta t_c} \end{aligned} \tag{12.14}$$

where $a = e^{-\Delta t_c/\tau_f}$ is the exact zero-order-hold coefficient of the derivative filter, which is unconditionally stable at any ratio of step to time constant, and $\hat{x}$ is the filtered position the derivative estimate is taken from. The code forms each increment as the difference between the complete proportional or derivative term at two successive scans, which reduces to Equation (12.14) at constant gains and makes a gain change act at once. The increments are combined and the demand is passed through the force limit and the slew limit, as shown in Equation (12.15).

$$\begin{aligned} s_k &= s_{k-1} + \Delta u^P_k + \Delta u^D_k \\ c_k &= s_k + I_{k-1} \\ F_{c,k} &= \mathrm{slew}\big(\mathrm{clamp}(c_k, \pm F_{lim}),\, F_{c,k-1},\, \dot{F}_{max},\, \Delta t_c\big) \\ I_k &= I_{k-1} + \Delta u^I_k + \frac{\Delta t_c}{T_t}\left(F_{c,k} - c_k\right) \end{aligned} \tag{12.15}$$

Here s_k accumulates the proportional and derivative increments without limit (the S_k of Equation (1.13)), c_k is the force the algorithm demands (the v_k of Equation (1.14)), $F_{c,k}$ is that demand clamped to the force limit and slew limited, and I_k carries the integral contribution. The integral used at scan k was accumulated through scan $k-1$. Anti-windup is back-calculation with tracking time $T_t = \tau_I$, applied only when $K_I = K_c/\tau_I$ is nonzero, which holds at every τ_I the application offers. While the command is clipped, the last term drives I toward the value that reproduces the force actually emitted, so the integral term alone is corrected and the running sum s_k is never truncated [18]. Away from the limits the recursion is algebraically identical to the positional law with the same three gains, and Section 1.6.2 explains why the running sum is never clipped. At the default $\tau_I = 1000\,\mathrm{s}$ each scan removes only 10^{-5} of the clipped excess, which leaves the back-calculation nearly inert, but the integral it guards is equally small because $K_I = 0.62\,\mathrm{N/(m\,s)}$. In the saturated runs of Section 12.6 the integral term never exceeds 18.1 N, against force limits of 1200 N to 4000 N, so windup plays no part in any result reported here and the application offers no windup switch. The loop drives the trolley in the direction of a positive error, because more force moves the trolley in the $+x$ direction, and no action switch is offered. Three limits act in series on the way to the plant, the clamp and the slew limit inside Equation (12.15) and the slew limit applied again to F_a, after which the first-order drive of Equation (12.3) rounds off whatever is left. The effective actuator is therefore slower than any one of those three numbers suggests, and a 6 m move that commands a peak of 3428 N develops a peak of 3046 N.

No coefficient in Equation (12.14) changes when the shaper is switched on. The shaper writes x^{sh}_{SP} and nothing else, which is why the comparison between a shaped and an unshaped run in Section 12.6 is a controlled experiment with exactly one difference in it.

12.5.3 Transfer and Manual Operation

Transfer is bumpless in both directions, as in every framework app (Section 1.6.2). In manual the loop runs in tracking mode against the shaped reference, which sets I_k so that c_k equals the operator's own force, and in automatic the controller output is copied to the manual input after every scan.

The manual station puts the size of the problem in front of the operator before any theory is offered. The drive force is entered directly over $\pm 4000\,\mathrm{N}$, held at $\pm 2400\,\mathrm{N}$ for as long as an incremental command is applied, or delivered as a fixed 0.8 s push of the same size. Driving a 6 m move by hand and stopping the trolley on the target is straightforward. Stopping the load with it is not, which is the first demonstration that the second half of the task is a timing problem rather than a force problem.

12.5.4 Model Predictive Control on the Shaper's Model

A model predictive controller (MPC), selected only by configuration 12, decides the drive force rather than the reference. Each plan covers the next $T_p = 6\,\mathrm{s}$, only its start is applied, and a new plan follows (Section 1.6.3), the receding-horizon law whose constrained form and terminal costs are reviewed in [19], whose industrial use is surveyed in [20], and which has bounded the swing of a laboratory crane [21].

The controller's model is not identified. It is Equation (12.1) linearized at a vertical cable without the centripetal term, completed by the drive lag, a ramped command and a constant horizontal force w on the payload, as shown in Equation (12.16).

$$\begin{bmatrix} M_0 + m_0 & m_0 \ell_a \\ 1 & \ell_a \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} F + w - b_0 \dot{x} \\ -g\theta - d_a \dot{\theta} + w/m_0 \end{bmatrix}, \qquad \tau_{a0} \dot{F} = F_c - F, \qquad \dot{F}_c = \rho, \qquad \dot{w} = 0 \tag{12.16}$$

Here ρ is the force rate the controller decides, $d_a = 2\zeta_a \omega_a \ell_a$ with $\omega_a = \sqrt{g/\ell_a}$, and the subscript 0 marks a declared value of Table 12.2. The cable length, damping ratio and gravity are the shaper's, so a wrong cable length misleads both controllers alike, while the masses, rail damping and drive lag stay declared whatever the plant does. Driven open loop by the shaped loop's force, the model follows the simulated swing to 0.002° over 15 s of configuration 3 but errs by 18.0° on the 8 m cable with the model at 4 m.

A zero-order hold over the decision period $T_s = 50\,\mathrm{ms}$, every fifth control interval, gives steps k up to $N = T_p/T_s = 120$. The decisions are force rates over six move blocks of 1, 2, 4, 8, 16 and 89 steps, so the planned command is piecewise linear, which the slew limit passes unchanged, and between decisions F_c follows the first planned ramp. The objective is Equation (12.17).

$$J = \sum_{k=1}^{N-1} \left\| S^{-1}(y_k - y^*) \right\|^2 + \lambda \sum_{k=0}^{N-1} \frac{\rho_k^2}{s_\rho^2} + \tilde{z}_N^T P \tilde{z}_N + \sum_{i \in \{\theta, x\}} \left(q_i \varepsilon_i^2 + r_i \varepsilon_i \right) \tag{12.17}$$

Here $y = (x + \ell_a\theta,\, \dot{x},\, \theta,\, \dot{\theta}/\omega_a,\, F_c)$ and $y^* = (x_{SP} + \ell_a\theta_{ss},\, 0,\, \theta_{ss},\, 0,\, F_{ss})$ is rest at the setpoint with $\theta_{ss} = \hat{w}/(m_0 g)$ and $F_{ss} = -\hat{w}$, where $\hat{w}$ is the estimate of w. The scales are

$$S = \mathrm{diag}(0.6\,\mathrm{m},\, 0.6\,\mathrm{m/s},\, 6^\circ,\, 6^\circ,\, 4000\,\mathrm{N})$$

with $s_\rho = 12\,\mathrm{kN/s}$ and $\lambda = 300$, chosen once on configuration 12 as the least force travel among the 265 of 1008 settings of the first three scales and λ within 10 % of the shortest time to a quiet payload, all of which kept the swing at or below the 3.23° of three impulses. $\tilde{z}_N$ is the departure of the first six states from rest, and P solves the discrete Riccati equation for the same weights, the cost-to-go of the linear quadratic regulator (LQR). With every step free and no limit binding, the first planned rate is the LQR's to $2.3 \times 10^{-5}\,\mathrm{N/s}$ on moves of up to 339 N/s.

Table 12.3 The Thirteen Shipped Configurations and Their Subjects

Configuration	What it changes and what it shows
1 Fast move	6 m, shaper off. The residual sway a good position loop leaves behind.
2 Two impulses	Same move, same gains, shaped reference.
3 Three impulses	One more half period of delay against the 3.3 % nominal frequency error.
4 Long cable	Hoist +4 m. The $\sqrt{2}$ scaling of Equation (12.4), and a retuned shaper.
5 Wrong length	8 m cable, shaper still told 4 m. Figure 12.2 at 0.707.
6 Heavy payload	300 kg. Mass detunes the loop, length detunes the shaper.
7 Force saturation	10 m move on a 1200 N drive. Feedforward against a pinned actuator.
8 Derated drive	1400 N. Slowing the crane down as the operator's own anti-sway method.
9 Already swinging	$\theta_0 = 8°$, no move commanded. A shaper with nothing to shape.
10 Repeated moves	Four moves at 1.5, 14, 28 and 42 s including a 0.5 m trim. Convolution and overlap.
11 Wind	120 N on the payload with $\tau_l = 10$ s. A load disturbance, not a command.
12 Shaper against MPC	The move of 1 to 3 under the predictive controller, the loop preset to three impulses. Prediction against cancellation.
13 Challenge	8 m move from a deliberately poor tuning against three simultaneous targets.

The limits $|\rho_k| \le \dot{F}_{max}$ and $|F_c| \le F_{lim}$, the latter at every block end and so everywhere, are hard and read where the position loop reads them. The slacks $\varepsilon_i \ge 0$ soften two limits judged every 0.1 s of the plan, $|\theta_k - \theta_{ss}| \le \theta_{max}$ with $\theta_{max} = 3°$, $q_\theta = 10^8$ per rad^2 and $r_\theta = 10^6$ per radian, and $0 \le x_k \le x_{max}$ with $x_{max} = 24$ m, $q_x = 10^4$ per m^2 and $r_x = 10^3$ per meter, because a hard state limit can be infeasible. Each plan is a quadratic program in eight variables with 266 inequality rows and a ridge of 5.0×10^{-7}, solved in 2.9 dual active-set iterations on average.

A steady-state Kalman filter on $(x, \dot{x}, \theta, \dot{\theta}, F, w)$ at Δt_c, driven by the applied force, supplies the initial state from the trolley position, and from the swing angle as well when the optional sensor is on, with measurement variances $(1\,\text{mm})^2$ and $(0.05°)^2$ and process variances of 10^{-9}, 10^{-6}, 10^{-12}, 10^{-6}, 1 and 1 per interval in SI units. The swing reaches the position only through the recoil $m_0 g\theta/M_0$, 0.0214 m/s^2 per degree or 7.67 mm of free trolley motion per degree of swing, which leaves the filter observable at a condition number of 9.5×10^4, and its estimate stays within 0.038° of the swing in configuration 12. The state w is the integrating disturbance that makes the MPC offset free [22], as the fuel bias does for the boiler of Volume 9, Chapter 4.

A plan that is not optimal, whose first rate is not finite or exceeds $\dot{F}_{max}$, or whose first block ends beyond F_{lim} is rejected. The position loop, which tracks the applied force while the MPC drives, then takes the drive until the next accepted plan. Three rejections forced at seven instants of configuration 12 move the command by at most 12.4 N, against steps of up to 34.9 N between scans of the undisturbed run, and no shipped configuration rejects any of its 600 plans in 30 s. The plant itself remains the nonlinear Equations (12.1) to (12.3) at the true values, read through the exact position and, optionally, the exact angle. A plan takes 211 µs on average in Node.js and 0.3 to 0.6 ms in Chrome, at most 1.2 % of a core at real time.

12.6 Results

The thirteen configurations shipped with the application are named in Table 12.3 beside the mechanism each is built around. A configuration fixes only the disturbance, drive rating, shaper or tuning its own subject requires and inherits every other value from Table 12.2, which keeps the single change under examination identifiable. Every configuration except 9 commands its first move as a timed setpoint step at $t = 1.5$ s, after which the crane is at rest at a new target and the run continues until the payload has quieted or has demonstrated that it will not.

Table 12.4 collects the outcome of every configuration and of the variants discussed below. All values come from the running application over a common 30 s horizon, with every setting changed before the run starts. The parking time is the first logged instant after which the trolley stays inside

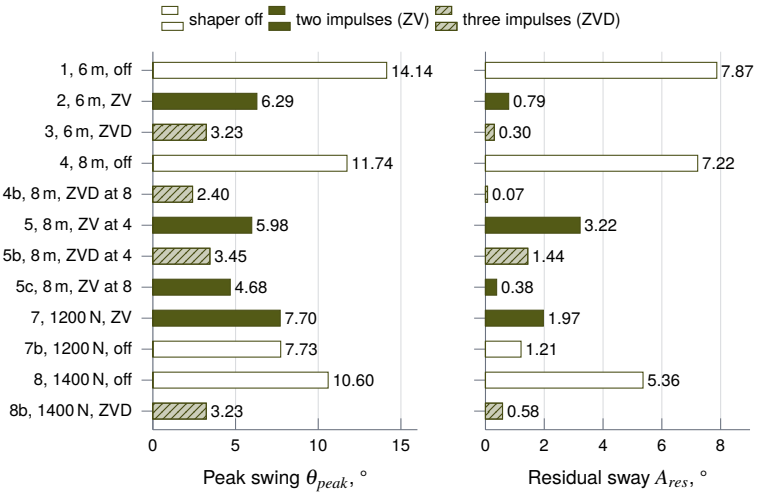


Figure 12.3 Peak swing and residual sway at parking from Table 12.4: the standard move (rows 1 to 3), the 8 m cable with the shaper told 4 or 8 m (rows 4 to 5c), the 1200 N drive on a 10 m move (rows 7 and 7b) and the 1400 N drive (rows 8 and 8b). White bars have the shaper off, solid bars two impulses and hatched bars three. Notice that a shaper told the true cable length leaves a fraction of a degree, a mistuned one leaves several times more, and on the 1200 N drive the shaped and unshaped peaks are the same.

± 0.10 m of the *raw* target for the rest of the run, A_{res} is the sway amplitude of Equation (12.5) at that instant, and A_{20} is the same amplitude at the fixed time $t = 20$ s, where the window over which the graded challenge is scored begins. All three are read from the 50 Hz record the application exports. The peak swing angle and the force-limited percentage are the application's own figures, the first updated at every 5 ms integration step, and the peak commanded force is taken over every controller scan. Figure 12.3 draws the peak swing and the residual sway of the main rows.

> **+ DID YOU KNOW? A Weaker Drive Swings the Load Less**
>
> From the simulator: derating the drive from 4000 N to 1400 N, with no control change at all, lowers the residual sway of the standard move from 7.87° to 5.36°, because the trolley can no longer accelerate as hard. Three impulses on the same derated drive leave 0.58° (Table 12.4, rows 1, 8 and 8b).

12.6.1 Shaping on the Nominal Crane

Configurations 1, 2 and 3 differ in one setting. The unshaped move peaks at 14.14° of swing, parks at 9.88 s and leaves 7.87° behind, and the payload is still moving 2.81° at $t = 20$ s. The two-impulse shaper cuts the peak to 6.29° and the residual to 0.79°, a factor of 10.0 on the residual, and the three-impulse shaper reaches 3.23° and 0.30°, a further factor of 2.7. The gains are identical in all three runs.

Table 12.4 Measured Outcome of Every Configuration Over a 30 s Horizon

Configuration	t_{park} (s)	θ_{peak} (deg)	A_{res} (deg)	A_{20} (deg)	$\lvert F_c \rvert_{peak}$ (N)	sat (%)
1 Fast move, off	9.88	14.14	7.87	2.81	3428	0.0
2 Two impulses	11.54	6.29	0.79	0.26	1935	0.0
3 Three impulses	12.72	3.23	0.30	0.07	1894	0.0
4 8 m, off	11.64	11.74	7.22	4.10	3428	0.0
4b 8 m, ZVD at 8	14.04	2.40	0.07	0.08	1828	0.0
5 8 m, ZV at 4	12.24	5.98	3.22	1.95	1914	0.0
5b 8 m, ZVD at 4	13.06	3.45	1.44	0.94	1882	0.0
5c 8 m, ZV at 8	12.14	4.68	0.38	0.32	1851	0.0
6 300 kg, ZV	11.14	6.00	0.96	0.15	2037	0.0
6b 300 kg, off	9.54	12.02	5.05	0.75	3429	0.0
7 1200 N, ZV	12.90	7.70	1.97	0.87	1200	10.1
7b 1200 N, off	12.86	7.73	1.21	0.49	1200	11.5
7c $K_c = 60$, off	26.32	3.56	0.70	—	600	0.0
7d $K_c = 60$, ZV	27.50	1.82	0.11	—	540	0.0
8 1400 N, off	10.40	10.60	5.36	2.12	1400	4.8
8b 1400 N, ZVD	12.74	3.23	0.58	0.22	1400	1.2
9 Already swinging, ZVD	—	8.02	—	1.29	124	0.0
11 Wind, ZVD, $\tau_l = 10$	—	8.18	—	4.28	1828	0.0
12 MPC	9.68	3.00	1.10	0.00	1043	0.0
13 Challenge, as loaded	26.10	13.29	2.95	4.32	2388	0.0
13b Challenge, tuned	14.32	4.26	0.15	0.16	2587	0.0
13c Challenge gains, off	12.88	18.55	7.59	3.85	4000	0.5

Rows 1 to 6, 8 and 12 are the standard 6 m move, rows 7 are 10 m and rows 13 are 8 m. Rows 6 and 7 are the configurations as shipped, with the two-impulse shaper on, and rows 6b and 7b switch it off. $\lvert F_c \rvert_{peak}$ is the peak commanded force, which the drive lag and the two slew limits reduce before it reaches the trolley. An em dash means the quantity does not apply, because the trolley never leaves its target in row 9, never settles inside the band in row 11, and has not yet parked at 20 s in rows 7c and 7d. Configuration 10 is discussed in the text rather than tabulated, because a sequence of four moves has no single parking time.

TRY IT IN THE APP **2 · ZV Input Shaping**

Load the scenario, press Run and watch the dashed shaped setpoint on the position trend. It climbs in two steps 2.006 s apart, and the swing the first step starts is canceled by the second. The peak sway reads 6.29° and the residual 0.79°, against 14.14° and 7.87° with the shaper off.

Two side effects are visible in the same table. The parking time grows from 9.88 s to 11.54 s and 12.72 s, which is 1.66 s and 2.84 s against shaper spans of 2.006 s and 4.013 s, the difference being that the loop was never going to track the raw step instantly either. The peak commanded force falls from 3428 N to 1935 N and 1894 N, because the loop is never presented with the whole 6 m error at once. Shaping buys actuator headroom as a side effect, and configuration 13 turns that side effect into the difference between passing and failing.

The three-impulse advantage on a crane that is not obviously mistuned is exactly the 3.3 % frequency error of Section 12.4. Figure 12.2 predicts 5.07 % residual for two impulses at that point and 0.26 % for three, a ratio of 20. The measured ratio of residuals is 2.7, which is far smaller, and the difference follows from the assumptions behind the prediction. Equation (12.9) describes a mode driven directly by the shaped signal through a linear time-invariant path. Here the mode is driven by the trolley acceleration produced by a saturating, slew-limited, lagged and trigonometrically nonlinear

closed loop, which puts a floor under the achievable residual. A sweep of the assumed length through $\ell_a = 2.5\,\text{m}$ to $8.0\,\text{m}$ locates that floor. The best A_{20} a two-impulse shaper reaches anywhere in the sweep is $0.210°$ and the best a three-impulse shaper reaches is $0.047°$, against $2.81°$ unshaped, so neither is reaching the zero that Equation (12.9) promises.

What the sweep does confirm is the position and the width of the two minima. The two-impulse residual is smallest at $\ell_a = 3.75\,\text{m}$, which is the value g/ω^2 predicted from the closed-loop mode of Section 12.4 and one grid step below the nominal $4.00\,\text{m}$, and it rises to $0.810°$ at $\ell_a = 5.0\,\text{m}$ and $2.138°$ at $\ell_a = 8.0\,\text{m}$. The three-impulse residual stays below $0.180°$ across the whole range $\ell_a = 3.0\,\text{m}$ to $5.0\,\text{m}$, where the two-impulse residual varies by a factor of 3.8.

12.6.2 Effect of Cable Length and Payload Mass

Paying out four meters of cable doubles ℓ and, by Equation (12.4), divides the frequency by $\sqrt{2}$. The predicted period rises from $4.012\,\text{s}$ to $5.674\,\text{s}$ and the measured closed-loop mode from $3.884\,\text{s}$ to $5.597\,\text{s}$, a measured ratio of 1.441 against the predicted 1.414. The payload mass does not appear in Equation (12.4) at all, and doubling it to $300\,\text{kg}$ moves the measured period only from $3.884\,\text{s}$ to $3.743\,\text{s}$, a shift of $3.6\,\%$ that comes entirely from the trolley recoiling harder against a heavier load. The two-impulse shaper survives that untouched, with the residual moving from $0.79°$ to $0.96°$ against $5.05°$ for the same move with no shaping.

The mistuned case is the sharpest test in the set. With a true $8\,\text{m}$ cable and a shaper still told $4\,\text{m}$, the second impulse lands $2.006\,\text{s}$ after the first when the payload needs $2.837\,\text{s}$, and the normalized frequency is 0.707. Four residuals on the same move quantify the loss. Doing nothing leaves $7.22°$, the mistuned two-impulse shaper leaves $3.22°$, the mistuned three-impulse shaper leaves $1.44°$, and a correctly tuned two-impulse shaper leaves $0.38°$. Expressed as fractions of the unshaped residual, the two mistuned shapers leave $44.6\,\%$ and $19.9\,\%$, against the $43.7\,\%$ and $19.1\,\%$ that Figure 12.2 predicts from Equation (12.9) alone. The sensitivity curve is quantitatively accurate where the frequency error is large, to within $5\,\%$ of the measured fractions. Retuning the three-impulse shaper to the true $8\,\text{m}$ recovers everything and more, at $2.40°$ of peak and $0.07°$ of residual.

The industrial literature reaches the same conclusion [12]. Robustness is not a substitute for knowing the cable length. It is insurance against the error that remains after the length has been measured as well as it can be, which is why a production system computes the impulse train from a live drum encoder and why the assumed length is a separate control in this application rather than being wired to the truth.

12.6.3 Conditions Under Which Shaping Fails

Three configurations show the method failing, and none of the three is repaired by retuning.

A $10\,\text{m}$ command at $K_c = 620\,\text{N/m}$ asks for $6200\,\text{N}$ from a drive rated at $1200\,\text{N}$. Shaping on and shaping off then give the same peak sway to within $0.4\,\%$, $7.70°$ against $7.73°$, and the shaped run ends with *more* residual sway, $1.97°$ against $1.21°$. The drive spends $10.1\,\%$ and $11.5\,\%$ of the run pinned at its limit, and while it is pinned the trolley is following the force limit rather than the reference, so reshaping the reference changes nothing that reaches the payload. Dropping the gain to $K_c = 60\,\text{N/m}$ keeps the peak command at $600\,\text{N}$, well inside the limit, and shaping immediately works again, cutting the peak from $3.56°$ to $1.82°$ and the residual from $0.70°$ to $0.11°$. Input shaping is a statement about a linear system, because convolution commutes with linear time-invariant dynamics and with nothing else. An actuator on its limit is not following the input at all, and a better shaper does not repair that. Making the commanded move something the actuator can execute does.

🕮 **Pause.** Try Exercise 12.3 before reading on.

> **TRY IT IN THE APP 7 · Drive Force Saturation**
>
> Run the scenario and watch the drive force sit on its 1200 N limit. The peak sway is 7.70° with two impulses and 7.73° without them, and the shaped run ends with more residual sway. Reload, set Proportional gain to 60 N/m and run again: the shaper now cuts the peak from 3.56° to 1.82°.

Starting the payload at 8° with no move commanded is the complementary failure. The three-impulse shaper is enabled and has no effect, because the setpoint never changes and a convolution with a constant returns that constant. Command shaping prevents sway that a command would otherwise have caused and has no mechanism whatever for removing sway that is already present. What does remove it is the position loop, through the accidental damping quantified in Section 12.4. The swing falls from 8.02° to 1.29° in 20 s, a log-linear fit over the decay giving an amplitude decay time of 10.4 s against the 42.6 s the cable pivot would give on its own, and the largest force the loop ever commands while doing it is 124 N.

Pause. Try Exercise 12.1 before reading on.

A steady 120 N wind on the payload is the third case, and it separates two mechanisms that are easy to confuse. At steady state the cable must tilt until its horizontal component balances the wind, so $\tan\theta = W/(mg)$ and the payload hangs at $\arctan(120/(150\times 9.81)) = 4.662°$ for as long as the wind blows. The application settles at 4.662°. No shaper and no position tuning can touch that, because the loop cannot reach θ. The trolley is a different matter. Holding station against the wind needs a steady 120 N of drive force, and proportional action alone can only produce it by sitting $W/K_c = 120/620 = 0.194$ m off target. Running the same case at the default $\tau_I = 1000$ s, which is slow enough to contribute almost nothing over a move, leaves the trolley 0.200 m past a 6.000 m target at 60 s with the drive holding −120.1 N, and the small excess over 0.194 m is integral action the loop has not yet unwound. The two problems have different mechanisms and neither remedy substitutes for the other.

12.6.4 Derated Drive, Repeated Moves and the Graded Challenge

Derating the drive from 4000 N to 1400 N lowers the peak sway from 14.14° to 10.60° and the residual from 7.87° to 5.36° with no control change at all, because sway is driven by trolley acceleration and the trolley can no longer accelerate as hard. That is the anti-sway method operators apply by hand, and it is the expensive one, because the reduction is paid for in move time on every cycle. The three-impulse shaper on the same derated drive gives 3.23° of peak and 0.58° of residual for one extra half period, and the 0.58° is worse than the 0.30° of the full-power case because the drive still spends 0.36 s on its limit, 1.2 % of the 30 s run, against 1.45 s with the shaper off.

Four moves in sequence show that the convolution is indifferent to the complexity of the setpoint history. Reading the sway amplitude at the last logged sample before the second, third and fourth commands and at the end of the 56 s run, the shaped run is at 0.013, 0.191, 0.273 and 0.061° while the unshaped run is at 4.25, 5.61, 7.21 and 2.17°. Peak sway over the whole sequence is 4.27° shaped and 21.75° unshaped. The shaper handles a 0.5 m trim and a 7 m traverse with the same relative treatment, and it handles a move that starts before the previous one has settled, because none of that is visible to a linear convolution. What accumulates in the unshaped run is sway from each move adding to the sway still left from the last one, which is what makes an unshaped crane demanding to drive.

The graded challenge asks for an 8 m move that holds the trolley within ±0.10 m of the target and the sway amplitude below 1.0° from $t = 20$ s to the end of a 30 s run, with no time at all on the force limit, starting from $K_c = 300$ N/m, $\tau_D = 0.5$ s and the shaper off. As loaded it fails the first two targets, overshooting to 9.90 m and not parking until 26.10 s, so the worst position error after 20 s is 0.14 m and the sway amplitude is still 4.32°. The order in which the three changes are made is what the challenge teaches. Raising τ_D to 2.5 s cuts the overshoot from 24 % to 0.8 %, because derivative action is the only damping Equation (12.6) offers, and raising K_c to 660 N/m makes the

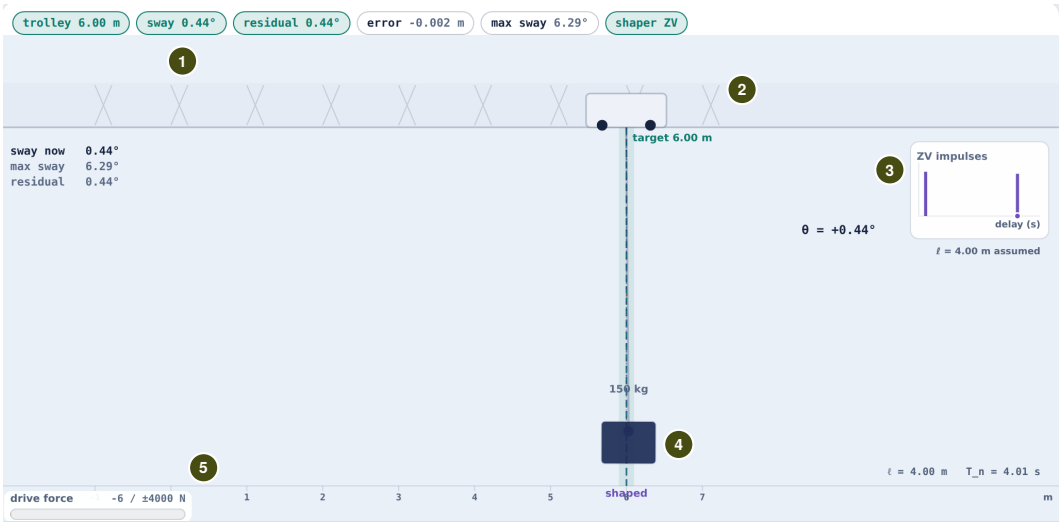


Figure 12.4 The two-impulse shaper on the standard 6 m move, paused at $t = 15$ s, cropped to the plant view. (1) The badges give the swing now, 0.44°, and its peak during the move. (2) The trolley stands at its target. (3) The impulse inset shows the two impulses of the shaped command, 0.512 and 0.488 separated by 2.006 s (Equation (12.10)), at the assumed cable length. (4) The load hangs nearly still above the shaped setpoint marker, leaving 0.44° against the 14.14° peak of the same move with the shaper switched off (Table 12.4). (5) The drive force is back near zero. The controller gains are unchanged.

move quick. Those gains alone, with the shaper still off, give 18.55° of peak sway, 7.59° of residual and a drive pinned at 4000 N for 0.16 s, 0.5 % of the run, so the sway and saturation targets are still missed although the trolley parks. Adding the three-impulse shaper makes the first commanded step $8 \times 0.262 = 2.10$ m instead of 8 m, which drops the peak command to 2587 N and takes the drive off its limit entirely, and the run parks at 14.32 s with 4.26° of peak sway and 0.15° of residual. Over the scored window the worst position error is then 0.02 m and the largest sway amplitude 0.16°, inside the targets by 0.08 m and 0.84°. Shaping returns nothing on a poorly damped loop or on a saturated one, so it is added last rather than first.

12.6.5 Shaper Against Predictive Control

Table 12.5 sets the shaped loop at its shipped tuning beside the MPC at configuration 12's settings, in 30 s runs with one exact trolley position, one cable length and no preview for both. The quiet time t_{quiet} is the first instant after which the trolley stays parked and the swing amplitude at or below θ_{ok}.

On the standard move the MPC parks at 9.68 s and is quiet at 10.56 s, where three impulses do both at 12.72 s, with a peak swing at its own 3.00° limit against 3.23° and 55 % of the peak command, 1043 N against 1894 N. It parks with 1.10° on the load against 0.30° and removes that inside the band within 0.88 s. The shaper's command is three pulses of 963, 1883 and 955 N whose swings cancel, while the plan pushes in two lobes of 1016 and 1042 N and holds the swing on its limit for 1.44 s, as shown in Figure 12.5. Holding the trolley back meanwhile costs position IAE against two impulses, 25.61 m s against 23.94 m s, although three impulses reach 29.67 m s.

On the 8 m cable with both controllers told 4 m, the MPC is quiet at 18.70 s against 29.28 s for three impulses, and two impulses not within the run, but it loses on peak swing, 4.44° against 3.45°. Its estimate, inferred from the recoil through a 4 m pendulum, errs by up to 1.39°, about the 1.44° by which it passes its limit, and its prediction is that of the shorter cable. The sensor removes the first

Table 12.5 Shaped Loop Against MPC Given One Cable Length, Runs of 30 s

Case	Law	t_{park} (s)	t_{quiet} (s)	A_{res} (deg)	θ_{peak} (deg)	$\lvert F_c\rvert_{peak}$ (N)	Travel (N)	t_{lim} (s)
6 m	ZV	11.54	12.58	0.79	6.29	1935	6491	0
	ZVD	12.72	12.72	0.30	3.23	1894	6366	0
	MPC	9.68	10.56	1.10	3.00	1043	4101	0
8 m, at 4	ZVD	13.06	29.28	1.44	3.45	1882	6691	0
	MPC	10.64	18.70	1.49	4.44	1019	4995	0
	MPC+s	10.64	15.92	1.37	3.74	1027	3962	0
8 m, at 8	ZVD	14.04	14.04	0.07	2.40	1828	6427	0
	MPC	10.26	11.04	1.09	3.00	1315	4841	0
300 kg	ZV	11.14	12.48	0.96	6.00	2037	6686	0
	MPC	13.54	13.90	0.72	3.31	1111	9169	0
1200 N	ZV	12.90	25.48	1.97	7.70	1200	3996	3.04
	MPC	11.56	12.38	1.22	3.00	1200	5009	0.05
8° start	ZVD	—	29.86	—	8.02	124	1491	0
	MPC	6.44	10.34	1.61	8.02	522	2959	0

The rows are configurations 2, 3 and 12, then 5 and 4 on the 8 m cable with $\ell_a = 4$ m, 4 with $\ell_a = 8$ m, then 6, 7 and 9. ZV and ZVD are the shaped loop with two and three impulses, MPC+s the MPC with the sway sensor, Travel the summed change of the commanded force and t_{lim} the time on the force limit. Every MPC run makes 600 plans and rejects none.

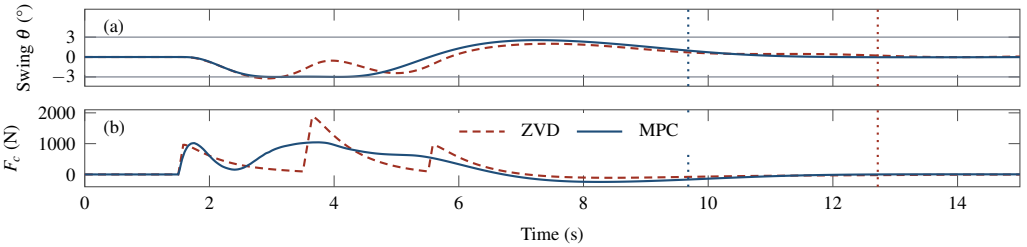


Figure 12.5 Configuration 12 from 0 to 15 s under the MPC and under the shaped loop with three impulses, both given the 4 m cable, the 6 m move commanded at 1.5 s. The MPC is solid and three impulses are dashed. (a) Swing angle against the ±3° limit of the plan (thin gray lines). (b) Commanded force. Dotted lines mark the parking instants, 9.68 s and 12.72 s.

error and not the second. It cuts the estimate error from 1.39° to 0.041° but the peak swing only from 4.44° to 3.74°. The 18.70 s belongs to the shipped settings. Over 20 combinations of decision period, block set and horizon the quiet time ranges from 14.3 to 25.4 s, while the peak stays between 3.9° and 5.4°, above the shaper's in every one. Told 8 m, both work, and the MPC is quiet 3.0 s sooner.

The MPC loses the doubled payload outright to two impulses, quiet at 13.90 s against 12.48 s with 9169 N of force travel against 6686 N, although it peaks at 3.31° against 6.00°. The recoil is twice what the model expects, w absorbs only part of it, and the estimate errs by 2.14°, while the shaper needs only a frequency that the mass moves by 3.6 %.

On the 1200 N drive the MPC rides the force limit, binding in 46 of its 600 plans, and is quiet at 12.38 s against 25.48 s. On the 1400 N drive its 1043 N peak never reaches the limit. The 8° initial swing, above 0.5° until 29.86 s under the shaped loop, is quiet at 10.34 s from the trolley position alone and at 8.28 s with the sensor, which then spends 0.15 s on the 4000 N limit. In the 120 N wind the estimate of w reaches 120.0 N and the trolley sits at 6.000 m at 30 s, where the loop with $\tau_I = 10$ s is at 6.133 m. With only the controller changed, the MPC meets all three targets of the challenge, parked at 10.64 s with 1.22° on the load, quiet from 11.48 s and never on the limit. The challenge is left as shipped, with the shaped loop's commissioning sequence as its suggested answer.

12.6.6 The Running Application

The application as it runs is reproduced in Figure 12.4. Its manual and automatic operation, scenarios and data export are those of every framework app (Section 1.6.4).

12.7 Technology in Practice

What Measures It

The shaped loop of this chapter needs one measurement, the trolley position, and one number entered by hand, the assumed cable length. On ship-to-shore cranes the second is measured too, by encoders on the hoisting rope, and the payload mass by weighing cells [4]. Those two readings are what a shaper needs in order to follow the hoist instead of waiting to be retuned, which is the gain scheduling that Section 12.6 performs by hand when it retells the shaper the true 8 m. The swing itself is not read by the shaped loop at all. The predictive controller's optional sensor stands for the cameras, lasers, inclinometers and inertial units that closed anti-sway loops use [23].

> **TECHNOLOGY Sensor: Rope Length, Load Mass and the Swing**
>
> On ship-to-shore cranes the length of the hoisting rope is read by digital encoders and the payload mass by weighing cells, both to about 5 %, so the rope length is known before a move [4]. A closed anti-sway loop needs further sensors for the load itself. Cameras with image processing, lasers, inclinometers and inertial units on the hook have all been used, and the inertial unit on the hook is small, cheap and need not be moved when the load changes [23]. The chapter's shaped loop reads the trolley position alone, and its assumed cable length is entered by hand.

What Moves It

A trolley drive is an electric motor fed by a frequency inverter, set by the operator's joystick as a desired trolley velocity on a quay crane, and on such a crane the motor pulls the trolley through gears, a drum and rope [4]. The drive's own torque and acceleration limits [24] are what the chapter models as F_{max} and $\dot{F}_{max}$, with a 0.08 s lag for the rest of its response. Those limits matter to the method, because a shaper is a linear device and the trolley leaves the linear range the moment the drive reaches its rating.

> **TECHNOLOGY Final Element: An Inverter-Fed Travel Drive**
>
> On a quay crane studied for anti-sway, the operator's joystick sets a desired trolley velocity, a frequency inverter controls the motors, and the motors wind or unwind rope on a drum through gears to pull the trolley [4]. Such drives limit speed, current and torque and the rate at which speed may change, and a host programmable logic controller can rewrite these limits while the motor runs [24]. The chapter's drive is a force source with a 4000 N rating, a slew limit and a first-order lag.

Where the Controller Runs

Programmable logic controllers serve as the primary controller of smaller systems that control discrete processes, and as local controllers under a supervisor in larger ones [25]. The chapter's controller runs every 0.01 s, the rate it names for a crane drive programmable logic controller. Its structure is a delay line of setpoints and an impulse train in front of a velocity-form PID block, followed by the clamp and the slew limits, and in a real installation the drive's own limits follow as well. Figure 12.6 draws the loop on that hardware.

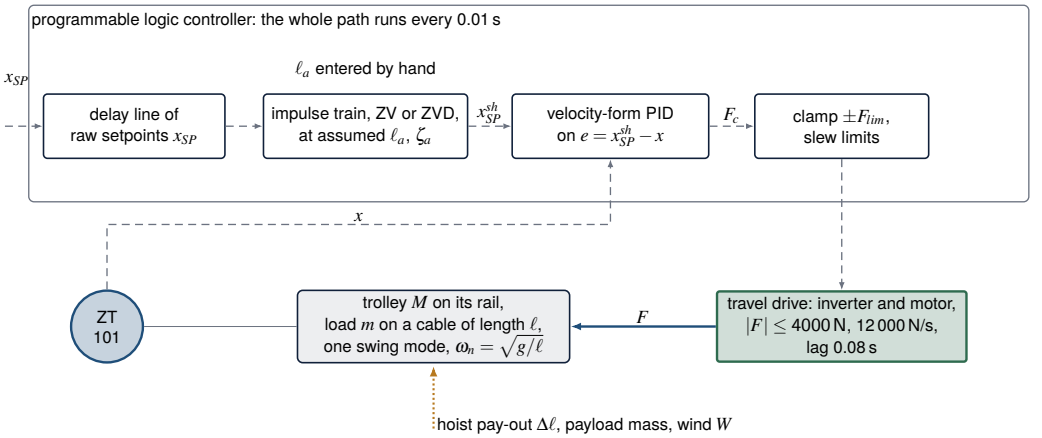


Figure 12.6 The crane loop on drive hardware. The trolley position ZT-101 is the only measurement. The raw setpoint passes a delay line and an impulse train evaluated at an assumed cable length entered by hand, the velocity-form PID acts on the shaped error, and the clamp and slew limits pass the command to the travel drive, which pushes the trolley. Hoisting, a change of payload and wind act on the crane. Notice that no path leads from the swing back to the drive, and that the impulse train learns of a hoisting change only if someone retells it the cable length.

> **IN PRACTICE** **What the Crane Rule Asks of the Operator**
>
> The United States rule for overhead and gantry cranes, 29 CFR 1910.179, requires that the hook be brought over the load so as to prevent swinging, that there be no sudden acceleration or deceleration of the moving load, and that the operator avoid carrying loads over people. Stops must be fitted at the limits of trolley travel, and bumpers must stop the crane at an average deceleration no greater than 3 ft/s^2 when it travels at 20 % of its rated load speed [26]. A shaped command is one way to move without the sudden acceleration the rule forbids.

What Is Modernizing It

Two developments are moving crane control beyond a fixed shaper. An inertial unit on the hook closed an anti-sway feedback loop on a five-ton industrial overhead crane and damped the first swing mode while the setpoint stood at zero, at the cost of slower tracking of the trolley velocity setpoint [23]. Image processing that tracks the payload has been used to measure how human operators drive an input-shaped bridge crane, and they finished obstacle courses faster and more safely with shaping on [17]. The predictive controller of Section 12.5.4 is a third path, and Section 12.6.5 measures what it gains and loses against the shapers when both are given the same cable length.

12.8 Limitations

The largest limitation is the single-mode payload. A real hook block on a sling is a double pendulum whose second mode is faster and lightly damped, and a shaper designed for one frequency leaves that mode untouched. Designing for two frequencies at once costs further delay and the trade has been worked out for tower and bridge cranes, but nothing in this application exhibits it.

The cable length is constant during every move. That assumption is the first to break in service, because cranes hoist and traverse together and the frequency sweeps while the impulse train is running. The application shows a paid-out cable as a step between moves, which understates the problem, and the industrial measurements are the appropriate reference for what a fixed shaper loses under a real hoisting profile [12].

OPEN CHALLENGE Two Pendulums and a Moving Hoist

A crane with a heavy hook and a long attachment below it swings in a second significant mode, and feedback designed on a single pendulum can then destabilize a mode it never modeled. On an industrial overhead crane, both modes had to be identified, and the case of hoisting while the trolley moves, needed in practice, was left for future work [23]. A quay-crane study simply assumed that the operator never moves trolley and hoist together [4]. Progress would be a shaper or anti-sway loop that follows a live rope length through both modes.

Everything is planar. A bridge crane has a second horizontal axis, its own sway mode, a skew mode when the two axes are excited together, and load rotation about the cable. None of that is here, and a shaper that works on one axis in isolation has to be checked against the other before it goes on a machine.

The measurement chain is idealized in both directions. The trolley position carries no noise, no quantization and no delay, which makes the derivative term of Equation (12.14) better behaved than a real encoder and filter would allow. The swing angle is absent, or exact behind the predictive controller's sensor, rather than noisy, so the application cannot show what an angle measurement of realistic quality would buy, which is the comparison a design study would require.

The predictive controller's model omits the sine of the angle, worth 0.316° on the 14.14° unshaped swing, and treats every change to the hoist, payload, trolley or drive as model error, which is why it passes its swing limit on the long cable and loses the doubled payload.

Two smaller items are worth stating. The residual sway reported in Table 12.4 is a sampled amplitude rather than an integral, so a run whose trolley enters and leaves the parked band during an overshoot produces a residual number that depends on when it is read, and the 30 s horizon is stated because the force-limited fraction is a fraction of elapsed run time and shrinks as the run continues, which is why the application reports the time on the limit in seconds beside it. Neither affects any comparison made above, because every comparison is between runs read the same way over the same horizon.

12.9 Conclusion

This chapter studied an overhead crane simulation on five states. Four of them are mechanical and reach their accelerations through a two-by-two solve that is never linearized, and the fifth is the force the drive has actually developed. The plant is underactuated, with one drive force reaching both the trolley position and a payload resonance whose frequency depends on the cable length and on neither mass. The shipped controller is a single velocity-form proportional-integral-derivative loop on trolley position, above which the reference is convolved with a two-impulse or three-impulse train whose amplitudes are $[1, K]/(1+K)$ and $[1, 2K, K^2]/(1+K)^2$ at multiples of the damped half period, with $K = \exp(-\zeta_a \pi/\sqrt{1-\zeta_a^2})$ evaluated at an assumed cable length rather than the true one.

A fifth-order linearization places the closed-loop sway mode at a period of 3.8835 s against a measured 3.884 s, which is 3.3 % away from the $\sqrt{g/\ell}$ the shaper formulas use and therefore right at the edge of the two-impulse tolerance band. Thirteen configurations quantify the outcome, from a 6 m move that leaves 7.87° of residual sway unshaped and 0.30° with three impulses, to a doubled cable that costs a mistuned two-impulse shaper 44.6 % of the unshaped residual against the 43.7 % the sensitivity curve predicts, to a saturated drive on which shaping on and off give the same 7.7° peak. A model predictive controller given the shaper's cable length quiets the standard move 2.16 s sooner than three impulses with 55 % of the peak force, but loses on a longer cable and a doubled payload. A fuller simulator would add a double-pendulum hook block, concurrent hoisting, the second horizontal axis, and an instrumented swing angle with realistic noise so that command shaping and feedback anti-sway can be measured against each other on the same move. The application measures what an open-loop canceler is worth once its design frequency is no longer quite the frequency the plant has.

KEY POINTS

- One drive moves the trolley, and the load swings only through how the trolley speeds up and slows down, so the position loop cannot reach the swing.
- A shaped command splits each move into impulses a damped half period apart, so the later ones cancel the swing the first one starts.
- A shaper told 4 m on an 8 m cable leaves 44.6 % of the unshaped residual with two impulses and 19.9 % with three.
- A drive on its limit does not follow the shaped reference, and a shaper cannot remove sway already present.

Where else this applies. The same two-impulse and three-impulse shapers sit above the climber's speed loop in Volume 3, Chapter 12, where they cancel a structural mode of the ribbon rather than a swinging load.

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

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

[4] E. Raubar and D. Vrančić, "Anti-sway system for ship-to-shore cranes," *Strojniški vestnik – Journal of Mechanical Engineering*, vol. 58, no. 5, pp. 338–344, 2012.

[5] 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.

[6] J. Vaughan, A. Yano, and W. Singhose, "Comparison of robust input shapers," *Journal of Sound and Vibration*, vol. 315, no. 4-5, pp. 797–815, 2008.

[7] O. J. M. Smith, "Posicast control of damped oscillatory systems," *Proceedings of the IRE*, vol. 45, no. 9, pp. 1249–1255, 1957.

[8] 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.

[9] W. Singhose, W. Seering, and N. Singer, "Residual vibration reduction using vector diagrams to generate shaped inputs," *Journal of Mechanical Design*, vol. 116, no. 2, pp. 654–659, 1994.

[10] E. M. Abdel-Rahman, A. H. Nayfeh, and Z. N. Masoud, "Dynamics and control of cranes: A review," *Journal of Vibration and Control*, vol. 9, no. 7, pp. 863–908, 2003.

[11] L. Ramli, Z. Mohamed, A. M. Abdullahi, H. I. Jaafar, and I. M. Lazim, "Control strategies for crane systems: A comprehensive review," *Mechanical Systems and Signal Processing*, vol. 95, pp. 1–23, 2017.

[12] W. Singhose, L. Porter, M. Kenison, and E. Kriikku, "Effects of hoisting on the input shaping control of gantry cranes," *Control Engineering Practice*, vol. 8, no. 10, pp. 1159–1165, 2000.

[13] K. L. Sorensen, W. Singhose, and S. Dickerson, "A controller enabling precise positioning and sway reduction in bridge and gantry cranes," *Control Engineering Practice*, vol. 15, no. 7, pp. 825–837, 2007.

[14] Y. Fang, W. E. Dixon, D. M. Dawson, and E. Zergeroglu, "Nonlinear coupling control laws for an underactuated overhead crane system," *IEEE/ASME Transactions on Mechatronics*, vol. 8, no. 3, pp. 418–423, 2003.

[15] Y. Liu and H. Yu, "A survey of underactuated mechanical systems," *IET Control Theory & Applications*, vol. 7, no. 7, pp. 921–935, 2013.

[16] I. F. A. Vis and R. de Koster, "Transshipment of containers at a container terminal: An overview," *European Journal of Operational Research*, vol. 147, no. 1, pp. 1–16, 2003.

[17] A. Khalid, J. Huey, W. Singhose, J. Lawrence, and D. Frakes, "Human operator performance testing using an input-shaped bridge crane," *Journal of Dynamic Systems, Measurement, and Control*, vol. 128, no. 4, pp. 835–841, 2006.

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

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

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

[21] H. Chen, Y. Fang, and N. Sun, "A swing constraint guaranteed MPC algorithm for underactuated overhead cranes," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 5, pp. 2543–2555, 2016.

[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] V. Helma and M. Goubej, "Active anti-sway crane control using partial state feedback from inertial sensor," in *2021 23rd International Conference on Process Control (PC)*. IEEE, 2021, pp. 137–142.

[24] 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.

[25] 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.

[26] Occupational Safety and Health Administration, "Overhead and gantry cranes, 29 CFR 1910.179," Code of Federal Regulations, Title 29, Chapter XVII, edition of July 1, 2024, U.S. Government Publishing Office, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/CFR-2024-title29-vol5/pdf/CFR-2024-title29-vol5-sec1910-179.pdf

12.10 Exercises

12.1 ⌕ Concept check · 2 min

The shipped controller of the crane is a velocity-form PID loop on trolley position with an input shaper above it. Scenario 9 starts the load swinging at 8° with the three-impulse shaper on and no move commanded. Which statement is right?

(a) The loop reads only the trolley position, so the shaper has nothing to act on and only the loop's coupling through the cable damps the swing

(b) The loop also reads the swing angle, so the three-impulse shaper cancels the initial swing within one damped half period

(c) The loop reads only the trolley position, but the shaper cancels the swing at its next impulse because its times follow the period

(d) The loop reads the cable tension, from which the shaper infers the swing and reshapes the setpoint to oppose it

12.2 ⌕ Concept check · 2 min

The first relation of Equation (12.1) is $(M+m)\ddot{x} + m\ell\ddot{\theta}\cos\theta - m\ell\dot{\theta}^2\sin\theta + b\dot{x} = F + W$. Which balance is it, and what trolley acceleration does a sudden drive force F produce with the load hanging still ($\theta = 0$, $\dot{\theta} = 0$, $\dot{x} = 0$, $W = 0$)?

(a) The moment balance on the payload about the pivot, and $F/(M+m)$

(b) The horizontal momentum balance on the trolley, and F/M

(c) The horizontal momentum balance on the trolley, and $F/(M+m)$

(d) The first-order lag of the drive force, and F/M

12.3 ⌕ Concept check · 2 min

Scenario 7 commands a 10 m move at $K_c = 620\,\mathrm{N/m}$ on a drive rated at 1200 N. The peak sway is 7.70° with the two-impulse shaper and 7.73° without it. Why does the shaper do nothing here?

(a) The assumed cable length suits a 6 m move, so the impulses land out of phase on a 10 m move

(b) The two-impulse shaper is too sensitive to the frequency, and the three-impulse shaper would recover the cancellation

(c) The drive sits on its limit, so the trolley follows the force limit rather than the shaped reference

(d) The derivative filter delays the shaped command by more than one damped half period

12.4 Theory and calculation · 6 min

The shaper is set to its defaults, an assumed cable length $\ell_a = 4.00\,\mathrm{m}$ and damping ratio $\zeta_a = 0.015$, with $g = 9.81\,\mathrm{m/s^2}$.

(a) Find ω_n, T_n, ω_d and the decay factor K of Equation (12.10). *Checkpoint:* $\omega_n = 1.5660\,\mathrm{rad/s}$, $T_n = 4.012\,\mathrm{s}$, $K = 0.9540$.

(b) Find the amplitudes and times of the two-impulse and the three-impulse shapers. *Checkpoint:* 0.5118 and 0.4882 at 0 and 2.006 s; 0.2619, 0.4997 and 0.2384 at 0, 2.006 and 4.013 s.

(c) The assumed length is set to 8.00 m. Find the new half period and the delay of the three-impulse shaper, and say which amplitudes change. *Checkpoint:* 2.837 s and 5.67 s; no amplitude changes.

12.5 Theory and calculation · 6 min

The position loop is approximately the second-order system of Equation (12.6), with $M + m = 1350\,\mathrm{kg}$ and $b = 400\,\mathrm{N\,s/m}$.

(a) Find ω_c and ζ_c at the default tuning, $K_c = 620\,\mathrm{N/m}$ and $\tau_D = 2.3\,\mathrm{s}$, and the ratio of the sway frequency 1.566 rad/s to ω_c. *Checkpoint:* 0.678 rad/s, 1.00, 2.3.

(b) The challenge loads $K_c = 300\,\mathrm{N/m}$ and $\tau_D = 0.5\,\mathrm{s}$, and suggests 660 N/m and 2.5 s. Find ζ_c for both, and the overshoot that a second-order loop predicts for the first. *Checkpoint:* 0.43 and 1.09; about 22 %.

(c) Find the gain that would place ω_c at twice the sway frequency, and the force it would ask for on a 6 m step. *Checkpoint:* 13.2 kN/m and 79 kN.

12.6 Theory and calculation · 5 min

For an undamped mode, Section 1.4.8 gives the residual of a two-impulse shaper as $|\cos(\pi r/2)|$ of the unshaped residual and that of a three-impulse shaper as $\cos^2(\pi r/2)$, where r is the ratio of the true to the design frequency.

(a) The cable is 8 m and the shaper is told 4 m. Find r and both residual fractions, and compare them with the 43.7 % and 19.1 % of Equation (12.9) with damping. *Checkpoint:* $r = 0.707$; 44.4 % and 19.7 %.

(b) Find the band of r that keeps each residual at or below 5 %, and the ratio of the band widths. *Checkpoint:* 0.968 to 1.032 and 0.856 to 1.144; a ratio of 4.5.

12.7 App challenge · 8 min

Load scenario 1, *Fast move, shaping off*, at the default seed and change nothing. Run to $t = 30\,\mathrm{s}$ and pause. Then do the same in scenario 2, *ZV input shaping*, and scenario 3, *ZVD input shaping*. After each run, read the Peak sway and Shaper delay tiles and export the CSV.

(a) In scenario 1, read the Peak sway tile and the column `sway_amplitude_deg` at $t = 20\,\mathrm{s}$ (tolerance 0.02°). *Checkpoint:* 14.14° and 2.81°.

(b) Read the same two values and the Shaper delay tile in scenarios 2 and 3. *Checkpoint:* 6.29°, 0.26°, 2.01 s; 3.23°, 0.07°, 4.01 s.

(c) The shaper is given the true 4.00 m cable in both runs. Explain why three impulses leave about a quarter of the amplitude of two at 20 s. *Checkpoint:* The true mode sits 3.3 % above the design frequency.

12.8 App challenge · 9 min

Load scenario 7, *Drive force saturation (10 m on a 1200 N drive)*, at the default seed. It starts with the two-impulse shaper on and $K_c = 620\,\mathrm{N/m}$. Each run lasts 30 s; read the Peak sway and Force limited tiles at the end (tolerance 0.02° and 0.05 s).

(a) Run the scenario as loaded. *Checkpoint:* 7.70°; 3.04 s (10.1 %).

(b) Press Reset, set Input shaper to OFF before Run, and run again. *Checkpoint:* 7.73°; 3.45 s (11.5 %).

(c) Reload the scenario, set Proportional gain to 60 N/m before Run and run with the shaper on. Then press Reset, set Input shaper to OFF and run again. Explain the four results. *Checkpoint:* 1.82° and 3.56°, with 0.00 s on the limit.

12.9 App challenge · 9 min

Load the challenge, *Challenge, park an 8 m move quietly*, at the default seed. It is judged at $t = 30$ s: from 20 s on the trolley must stay within 0.10 m of 8 m with a sway amplitude under 1.0°, and the drive must spend no time on its limit. A failed run is reported in the progress checklist with its worst position error, worst sway amplitude and force-limited share. Make every change before Run, press Reset between runs and leave the suggested tuning hidden.

(a) Run as loaded to 30 s. Which targets fail, and by how much? *Checkpoint:* 0.14 m and 4.32°; the force target is met.

(b) Set Derivative time to 2.5 s and Proportional gain to 660 N/m, leave the shaper OFF, and run again. Read the checklist and the Peak sway and Force limited tiles. *Checkpoint:* 0.05 m, 3.85°, 0.5 %; 18.55° and 0.16 s.

(c) Set Input shaper to ZVD and run again. Read the Peak sway and Force limited tiles, and the largest `sway_amplitude_deg` from 20 s on in the exported file (tolerance 0.02°). *Checkpoint:* 4.26°, 0.00 s, 0.16°; the challenge is met.

12.10 Application area · 4 min

A container terminal counts crane moves per hour, and an operator does not lower a load until its swing is below about 0.5°. The unshaped 6 m move of scenario 1 parks at 9.88 s with 7.87° of sway, and the closed loop decays the amplitude with a time constant of 10.38 s.

(a) Estimate when the load becomes quiet, and compare with the three-impulse shaper, which parks at 12.72 s with 0.30°. *Checkpoint:* About 38.5 s; about 26 s longer.

(b) Derating the drive to 1400 N cuts the unshaped residual to 5.36°. Say why an operator who slows the crane this way pays more than the shaper costs. *Checkpoint:* Move time on every cycle, and the load still has to be waited for.

12.11 Application area · 3 min

An outdoor gantry crane carries the default 150 kg load in a steady 120 N wind along the rail, as in scenario 11.

(a) Find the angle at which the load hangs, and say whether a shaper or a retuned position loop can bring it back to vertical. *Checkpoint:* 4.662°; neither can.

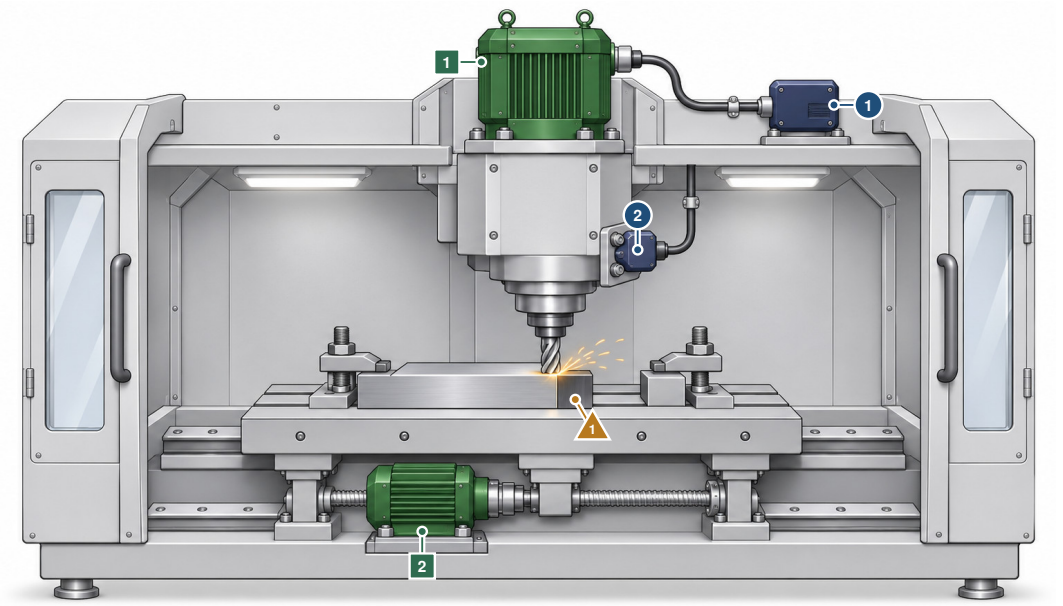
(b) The position loop runs proportional-derivative only. Find the steady offset of the trolley, and say what removes it. *Checkpoint:* About 0.194 m; integral action.

12.12 Review the agent · 5 min

An AI assistant is asked to commission the shaper on the default crane and replies: “The swing frequency of a hanging load is $\sqrt{g/\ell}$ and holds neither mass, so set Shaper cable length to exactly the true 4.00 m. Any other setting can only add residual sway. Use two impulses, since three add a needless 2 s.” Find the flaw, and say how to check it in the app.

CHAPTER 13

Stability Lobes and Adaptive Feed Control in a CNC Milling Simulator



The controller moves 1 the spindle drive, which sets the tool's speed
2 the feed drive, which sets how fast the tool moves through the metal

The controller reads 1 the spindle load, read from the drive 2 the vibration sensor on the spindle housing

What disturbs it 1 a harder patch in the metal, or a worn tool

Notice that the feed drive and the spindle drive act on the same cut from different places, one easing the load and the other moving the tool's speed out of chatter. Illustration, not to scale.

IN PLAIN TERMS Two Limits on One Cutting Tool

The main idea. A spinning cutter that clears a pocket in a block of metal can be stopped by two different limits. Its motor can be overloaded, mostly in the corners, and the tool can start to shake and leave a wavy, marked surface.

The challenge. Moving the tool more slowly through the metal eases the motor, but it does not stop the shaking. Only a change of the spinning speed ends it, so the controller must tell the two limits apart while it cuts.

Why it matters. A shop that slows down to cure the shaking loses time and still gets a surface that must be reworked.

THREE WAYS TO READ THIS CHAPTER

⊘	**Explore**	12 min	the first page, the fact in the Introduction (Section 13.1), the two history segments in Section 13.2, the try-its and the fact in Section 13.5, and the key points
🕮	**Study**	45 min	the whole chapter with its three pause points and the twelve exercises at its end
⏢	**Practice**	15 min	the first page, the results (Section 13.5), Technology in Practice (Section 13.6), Limitations (Section 13.7) and the open challenge

The control question An end mill clearing a pocket overloads its spindle in the corners and chatters above a critical depth, and both look like a cut that labors. Which limit does the feed override remove, which one only the spindle speed removes, and how does a controller tell them apart on line?

Learning objectives

- Derive the stability limit of a regenerative cut from its delay equation, and explain why the feed does not appear in it.
- Compute the spindle load from the removal rate, and tell when the spindle power rather than chatter limits a cut.
- Apply the whole-wave speed rule to a measured chatter frequency, including the bounded step.
- Compare, on the app, what an adaptive feed and a speed change each do about a corner, a hard patch, a worn tool and a late load signal.
- Judge a proposed remedy for chatter against the stability lobes.

Prerequisites Sections 1.4.9 (a delay that makes a cycle, bounded by a nonlinearity), 1.4.8 (a lightly damped mode) and 1.4.3 (a gain that moves with the operating point), with Sections 1.6.2 and 1.6.1 for the feed loop and the engine.

The app https://apmonitor.com/apps/cnc-milling/

See also Volume 5, Chapter 8 for a delay that makes a loop of passive parts chatter, Volume 5, Chapter 3 for a tuning that fails once its measurement arrives late, and Volume 5, Chapter 2 for control through a transport delay.

13.1 Introduction

An end mill clearing a pocket is stopped by two limits that an operator meets in the same way, as a spindle that labors and a cut that sounds wrong. The first is the load. The mean cutting power is the specific cutting force times the volumetric removal rate, and into every inside corner the arc of contact grows until the tool is cutting a full slot, so a feed chosen for the straights overloads the tool in the corners. The second is regenerative chatter. Each tooth cuts the surface the previous tooth left one tooth period earlier, a vibration written into that surface returns as a varying chip thickness, and above a critical depth of cut the returning wave grows on every pass [1, 2]. Chatter is among the principal limits on the productivity of machining, because it leaves a surface that must be reworked and it damages the cutting edges and the spindle [3, 4].

+ DID YOU KNOW? Most of the Slab Leaves as Chips

High-speed, high-power spindles were developed to remove more than 95 % of the material from aluminum slabs, leaving monolithic, lightweight parts for the aerospace industry [4]. The aluminum pocket of this chapter is one small step of that kind of work, and chatter is what limits how fast it can be taken.

The two limits call for different actuators, and a block diagram of either one hides that fact. The load is a level, proportional to the feed, and a feed override answers it. The critical depth is the stability boundary of a delay differential equation, set by the frequency response of the tool, the cutting stiffness and the tooth period, and the feed does not enter it at all [2, 5]. Slowing the feed thins the chip and so shrinks a chatter vibration, whose amplitude is held finite by the tool leaving the cut [6], but it does not end the chatter. The simulator is designed to overturn a belief common on the shop floor, that backing off the feed cures any cut that misbehaves. In the application, an adaptive feed loop that holds the load of a worn tool within 1.6 % of its target still leaves 72 mm of chatter marks on a 92 mm pass, while a spindle speed 5.4 % lower ends the chatter of a cut 11 % over its limit within 0.21 s of the command.

The chip regeneration belongs to a general class, the repetitive or multipass process, in which each pass of a tool responds to the profile the previous pass left, and whose classic examples are metal rolling and long-wall coal cutting [7]. A loop of that kind carries a delay equal to the time between

passes, its characteristic equation has infinitely many roots, and its stability can change with the delay while every gain stays fixed [8]. The feed loop carries a second general lesson. Machining control is usually organized in three levels, the servo level that moves the axes, the process level that regulates a cutting quantity such as force or power, and a supervisory level above both [9]. At the process level the load per unit of feed doubles between a straight and a corner, so a fixed-gain force loop tuned for one is mistuned for the other, which is why force control in milling became a problem in adaptive control [10]. The same shape appears wherever a process gain moves with an operating condition that the controller does not measure, as in the neutralizer of Volume 2, Chapter 6, whose gain changes by orders of magnitude across one valve stroke.

The two subjects are normally taught apart. Chatter appears as the stability analysis of a delay differential equation whose result is a lobe diagram, and machining force control as a regulation problem in which the cut is a static gain [4, 9]. However, neither presentation supplies an artifact in which one pocket is cut with the feed loop and the speed decision working together, so that the feed loop can be watched holding the load at a corner while the lobes decide the finish.

This chapter works with a CNC milling simulator in which the load limit and the stability limit act on one pocket, each with its own controller, so that the question of which actuator answers which limit is settled by measurement. The average-tooth stability analysis, the contact-loss nonlinearity, the adaptive force loop, the energy-ratio chatter indicator and the spindle speed rule are established methods, all from the literature reviewed in Section 13.2. The simulator advances a regenerative delay model with contact loss whose simulated stability boundary lies within 0.1 % of the analytic lobes, along a toolpath whose engagement is measured by removing material from a 0.2 mm grid, and both controllers are specified completely enough to rebuild. Eight cases each move a single mechanism, and a graded challenge can be passed only with both loops working.

Section 13.2 traces chatter theory, spindle speed selection and adaptive machining. Section 13.3 builds the tool, chip and machine model, derives the stability lobes and describes the numerical solution and its check against the lobes. Section 13.4 gives the spindle drive, the adaptive feed loop, the chatter detector and the speed rule. Section 13.5 reports the eight cases and the graded challenge, Section 13.6 sets the two loops on machine-tool hardware, Section 13.7 states the limitations and Section 13.8 draws the results together. The exercises in Section 13.9 work through the model and the cases.

13.2 Chatter Theory, Speed Selection and Adaptive Machining

The first stability laws, Tlusty's account of the theory of chatter and the stability analysis of machine tools among them [1], are traced in the segment that follows this paragraph. The regenerative loop is linear only while the tool stays in the material. Once the vibration exceeds the chip thickness the tool leaves the cut for part of each cycle, and that basic nonlinearity is what holds chatter at a finite amplitude [6].

> ⧖ HISTORY **From a Machinist's Puzzle to a Stability Law, 1906 to 1965**
>
> An article of 1906 called chatter the "most obscure and delicate of all problems facing the machinist" [4]. The first scientific stability laws came independently in the 1950s. One gave a critical depth of cut set by the machine's dynamic stiffness and the material's cutting force coefficient, whatever the spindle speed. The other added the regenerative delay of the spindle speed and found the stable pockets, or lobes [4]. Merritt then cast the regeneration as a closed loop with delayed feedback [4] and plotted the borderline of stability as a stability chart. His simpler criterion keeps the directional cutting stiffness below one half of the minimum directional dynamic stiffness of the structure, which assures chatter-free cutting at all spindle speeds. Machine testing with continual improvement of the structure was, he wrote, the best way to combat chatter [2].

Milling was harder to analyze, because the number and the direction of the cutting teeth rotate with the spindle and the delay equation acquires periodic coefficients. Averaging the directional factors over a tooth period gave an analytical prediction of the stability lobes of milling [5], the approach this application follows. The average fails in highly interrupted cutting at small radial immersion, where a theory supported by simulation and experiment predicts a doubling in the number of optimally stable speeds [11]. The extra pockets of intermittent cutting and the process damping contributed by the flank face of the tool are part of the modern account, reviewed in [4].

The high-speed machining of the early 1990s, with spindles fast enough to run near the tooth-passing frequencies at which the lobes open into stable pockets, made the stability chart a production tool rather than a laboratory result [12]. Finding the pocket automatically followed. Smith and Tlusty regulated the spindle speed from the measured chatter frequency, so that successive teeth meet the surface wave in phase [13], and a microphone was shown to be a competitive and often superior sensor for detecting chatter, against dynamometers, displacement probes and accelerometers [14]. Later detectors remove the forced vibration at the tooth-passing frequency and its harmonics and declare chatter when the energy of a remaining component grows relative to the forced part [15]. Speed selection is now one of many suppression techniques, passive and active, collected in [3].

Adaptive control of metal cutting dates from the same years as the stability chart, as the segment after this paragraph tells. Force regulation needs adaptation because its process gain moves with the cut. The feed loop of this application is a deliberately simple member of that family, a proportional-integral controller whose gain is divided by an identified load per unit of feed, and its filtered and optionally delayed load signal reproduces the trade-off that such a filter imposes.

> ⧖ HISTORY **Adaptive Control Reaches the Cut, 1964 to 1993**
>
> In 1964 numerical control closed its feedback loops only around the positions of the tool and the workpiece. An adaptive controller reported that year extended the measurement to the interface of tool and work and adjusted the process inputs to optimize a criterion such as cost, under constraints on the surface finish and the machine's horsepower [16]. A model reference adaptive force controller for milling later performed more satisfactorily than fixed-gain controllers but was difficult to implement and tune. Runout of the cutter put noise on the force signal, and the filter that removed it added dynamics that reduced the performance [10]. A review of 1993 ordered the research by the servo, process and supervisory levels of a control hierarchy and highlighted adaptive and optimal control [9].

13.3 Tool, Chip and Machine Model

The simulator clears a pocket 120 mm by 60 mm in 7075-T6 aluminum with a 12 mm three-flute carbide end mill. Seven states are integrated, the tool deflection y and its rate $\dot{y}$ along one bending mode, the spindle speed Ω, the position s along the toolpath, the feed velocity v_f, the tooth phase θ and the flank wear VB. The machine and its two loops, the engagement in plan view and the regenerative chip are shown in Figure 13.1, and the variables are classified in Table 13.1.

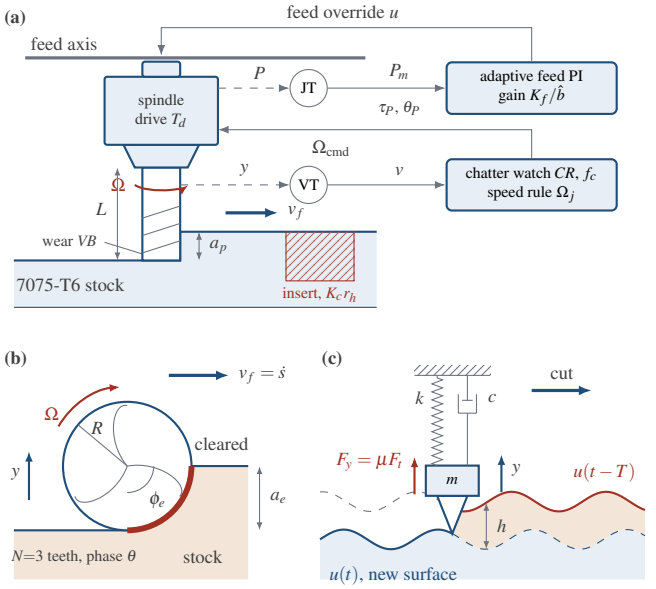


Figure 13.1 End milling with two loops. (a) The tool of overhang L cuts to depth a_p while the feed axis moves it at v_f, the spindle drive turns it at Ω with torque T_d, and flank wear VB grows on its edges. The load transmitter JT passes the cutting power P as P_m, after the load-signal filter τ_P and delay θ_P, to the adaptive feed loop that sets the override u, and the vibration transmitter VT passes the deflection y as v to the chatter watch that sets the speed command Ω_{cmd}. The hatched insert raises the specific cutting force K_c by the ratio r_h. (b) In plan view the tool of radius R and $N = 3$ teeth at phase θ removes a band of radial engagement a_e over the engagement angle ϕ_e, climb milling, and deflects by y away from the stock. (c) The tool mode m, k, c is driven by $F_y = \mu F_t$, and the chip h lies between the surface $u(t-T)$ the previous tooth left and the path the present tooth cuts.

Table 13.1 Variables and Their Roles in the Default Uniform Cut

Symbol	Role	Nominal	Range
$\bar{P}$	CV, spindle load	0.64 kW	SP 0.2–8 kW
CR	CV, chatter indicator	3 %	SP 0
u	MV, feed override	100 %	10–150 %
Ω_{cmd}	MV, speed command	20 000 rpm	3000–24 000
y, $\dot{y}$	states, deflection and rate	2.5 µm, 0	—
Ω	state, spindle speed	20 000 rpm	—
s, v_f	states, path and feed	760 mm, 80 mm/s	0–1044 mm
θ	state, tooth phase	0	0–2π
VB	state, flank wear	50 µm	start 0–400 µm
a_p	setting, axial depth	1.5 mm	0.5–12 mm
F	setting, programmed feed	4800 mm/min	600–9000
a_e	DV, engagement on the path	6 mm	0–12 mm
r_h	DV, insert force ratio	1	1–2.5
L	DV, tool overhang	40 mm	30–60 mm
θ_P	DV, load-signal delay	0 ms	0–200 ms
—	DV, tap on the tool	60 mm/s	event
P_m	measured load	0.64 kW	noise 20 W
v	measured vibration	2.5 µm	noise 0.2 µm

13.3.1 Pocket, Toolpath and Engagement

The pocket is cleared in one layer by a contour-parallel outward spiral, an entry slot 60 mm long along the middle followed by four rectangular loops at a 6 mm stepover, half the tool diameter. Each

loop starts and ends at the middle of its right side, the toolpath is 1044 mm long, and the loops run counterclockwise, which with a clockwise spindle is climb milling. The radial engagement $a_e(s)$ is measured once, before any run, by moving the tool disk along the path over a 0.2 mm material grid, counting the cells removed in each 0.2 mm of travel and smoothing the result with a 1 mm moving average. On a straight the tool cuts a 6 mm band, $a_e = R$, and over the last ten millimeters into each inside corner the arc of contact grows until it cuts a full slot, $a_e = 2R$. The engagement angle, the average number of teeth in the cut and the mean chip are then computed from it by Equation (13.1).

$$\phi_e = \arccos\!\left(1 - \frac{a_e}{R}\right), \quad N^* = \frac{N\phi_e}{2\pi}, \quad \bar{h} = f_z \frac{a_e}{R\,\phi_e} \tag{13.1}$$

for a tool of $N = 3$ teeth and radius $R = 6$ mm, with $f_z = 2\pi v_f/(N\Omega)$ the feed per tooth and $\bar{h}$ the chip thickness averaged over the engaged arc. A straight has $\phi_e = 90^\circ$ and $N^* = 0.75$, a slot $\phi_e = 180^\circ$ and $N^* = 1.5$, and both have $\bar{h} = 0.637 f_z$, so a corner doubles the teeth in the cut at the same chip. A cast-iron insert 20 mm by 18 mm set into the bottom middle of the pocket is crossed by the bottom passes of loops 2, 3 and 4 when its force ratio r_h exceeds 1, and the fraction χ of the removed cells that lie inside it scales the specific cutting force.

13.3.2 Tool Mode

The tool bends along one mode normal to the machined surface, with the deflection y positive away from the material, as shown in Equation (13.2).

$$m\ddot{y} + c\dot{y} + ky = F_y, \quad k = k_0\left(\frac{L_0}{L}\right)^3, \quad m = m_0\,\frac{L}{L_0} \tag{13.2}$$

where L is the overhang, $L_0 = 40$ mm the reference overhang, $k_0 = 12\,\mathrm{N/\mu m}$ the tip stiffness at L_0, $m_0 = k_0/(2\pi f_{n0})^2 = 0.475$ kg with $f_{n0} = 800$ Hz, and $c = 2\zeta\sqrt{km}$ with $\zeta = 0.03$. The stiffness falls with the cube of the overhang and, because the modal mass grows in proportion to it, the natural frequency falls with its square. At 50 mm the tool has 6.14 N/µm and 512 Hz.

13.3.3 Regenerative Chip and Contact Loss

With $u(t)$ the surface left at time t, measured in the same coordinate as y, the chip the tool takes is given by Equation (13.3).

$$\begin{aligned} h(t) &= h_s(t) + u(t-T) - y(t), \qquad T = \frac{2\pi}{N\Omega} \\ h_s(t) &= \bar{h}\,(1 + \beta_1\cos\theta + \beta_2\cos 2\theta) \end{aligned} \tag{13.3}$$

where T is the tooth period, the tooth phase obeys $\dot{\theta} = N\Omega$, and $\beta_1 = 0.15$ and $\beta_2 = 0.05$ give a two-harmonic ripple at the tooth-passing frequency that stands for the teeth entering and leaving the cut. The ripple places forced lines in the spectrum for a detector to reject and leaves the regenerative term untouched. The new surface follows Equation (13.4).

$$u(t) = \min\big(u(t-T) + h_s(t),\; y(t)\big) \tag{13.4}$$

While $h > 0$ the tooth cuts and leaves its own path as the new surface. When the vibration throws the tool clear, $h \le 0$, the tooth cuts nothing and the old surface is carried forward, which is the contact-loss nonlinearity that bounds chatter to a limit cycle [6], the mechanism of Section 1.4.9. In the uniform cut at 3 mm and 20 000 rpm that limit cycle is 60 µm rms, 1.2 times the 51 µm mean chip.

13.3.4 Cutting Force and Spindle Power

The tangential force follows the chip and the force along the mode is a fixed fraction of it, Equation (13.5).

$$F_t = \begin{cases} K a_p N^* h + K_{\text{rub}} VB\, a_p N^*, & h > 0 \\ 0, & h \le 0 \end{cases} \qquad (13.5)$$
$$K = K_c\,[1 + (r_h - 1)\chi]\,(1 + k_{VB} VB), \qquad F_y = \mu F_t$$

where a_p is the axial depth of cut, $K_c = 800\,\text{N/mm}^2$ the specific cutting force of the sharp tool in 7075-T6, $k_{VB} = 1.2\,\text{mm}^{-1}$ the rise of the cutting force with flank wear, $K_{\text{rub}} = 40\,\text{N/mm}^2$ a rubbing force per unit area of wear land and $\mu = 0.6$ the directional orientation factor of the average-tooth model [5]. The spindle load is the cutting power $P = F_t R\,\Omega$. Averaged over a tooth period on a steady cut, Equation (13.1) gives $N^* \bar{h} = N f_z a_e / (2\pi R)$, so the load is the specific cutting force times the removal rate, $\bar{P} = K a_p a_e v_f$, plus the rubbing term. On the 1.5 mm straight at 4800 mm/min the removal rate is 43.2 cm^3/min, a sharp tool would draw 0.58 kW, and the default 50 µm of flank wear raises the load to the 0.64 kW the application reports.

13.3.5 Stability Lobes

While the tool stays in the cut and the ripple is set aside, $u(t - T) = y(t - T)$, and Equations (13.2) to (13.5) reduce to a linear delay differential equation with the characteristic equation shown in Equation (13.6).

$$1 + \mu K a_p N^* \left(1 - \mathrm{e}^{-sT}\right) G(s) = 0 \qquad (13.6)$$

where $G(s) = 1/(ms^2 + cs + k)$ is the receptance of the tool mode. The exponential gives the equation infinitely many roots [8]. Setting $s = j\omega_c$ yields the depth at which a root crosses the imaginary axis and the chatter frequency there, Equation (13.7).

$$a_{\text{lim}} = \frac{-1}{2\mu K N^* \operatorname{Re} G(j\omega_c)}, \qquad \omega_c T = 3\pi + 2\psi + 2\pi k \qquad (13.7)$$

where $\psi \in (-\pi, 0)$ is the phase of $G(j\omega_c)$ and $k = 0, 1, 2, \ldots$ numbers the lobe. For each lobe the phase condition has one root above the natural frequency, which the application finds by bisection, and the boundary is the lowest lobe, the one-mode form of the average-tooth result [5]. The feed does not appear. The lowest depth over all speeds is $a_{\text{crit}} = 2k\zeta(1 + \zeta)/(\mu K N^*)$, which is Merritt's criterion of half the minimum dynamic stiffness [2], and it is 1.94 mm on the straight. The boundary for the straight at two overhangs is shown in Figure 13.2. At 20 000 rpm the limit is 2.10 mm with chatter predicted at 816 Hz, and at 16 300 rpm, where the 815 Hz tooth-passing frequency sits just above the natural frequency, it is 17.9 mm. The 50 mm tool moves the pockets to lower speeds, and its limit at 16 300 rpm is 1.19 mm.

Pause. Try Exercise 13.1 before reading on.

13.3.6 Spindle, Feed Axis and Wear

The spindle obeys Equation (13.8).

$$J\dot{\Omega} = T_d - F_t R - c_\Omega \Omega, \qquad |T_d| \le \min\left(T_{\max}, \frac{P_{\max}}{\Omega}\right) \qquad (13.8)$$

where $J = 0.004\,\text{kg m}^2$, $T_{\max} = 10\,\text{N m}$ and $P_{\max} = 5.5\,\text{kW}$, so above the base speed of 5252 rpm the drive is power limited, and $c_\Omega = 2 \times 10^{-5}\,\text{N m s}$ is the bearing drag. The feed velocity follows its reference through a servo lag, $\tau_F \dot{v}_f = v_{\text{ref}} - v_f$ with $\tau_F = 5\,\text{ms}$, and $\dot{s} = v_f$. Flank wear grows while the tool cuts, as given by Equation (13.9).

$$\frac{\mathrm{d}VB}{\mathrm{d}t} = w_0 \left(\frac{\Omega R}{v_0}\right)^2 \min\left(1, \frac{N^*}{0.75}\right) \qquad (13.9)$$

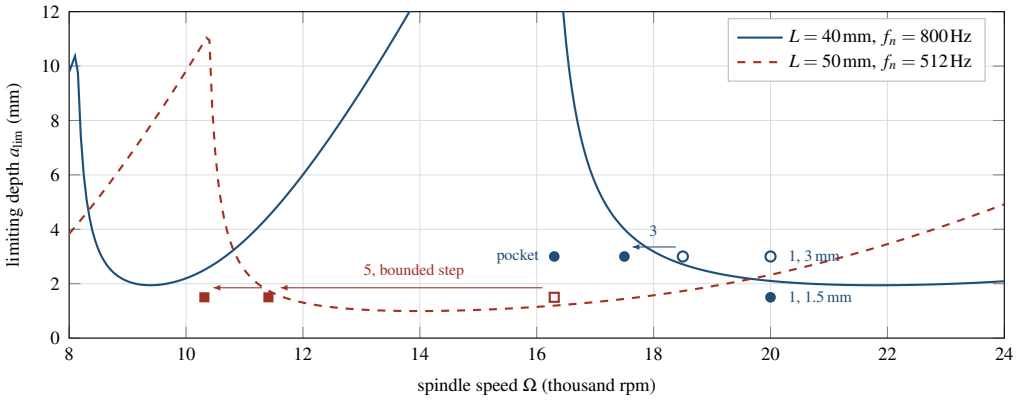


Figure 13.2 Stability lobes of a 50 % straight, $a_e = 6$ mm, from Equation (13.7) with the default 50 µm of flank wear. The solid line is the boundary of the 40 mm tool and the dashed line that of the 50 mm tool. Circles are operating points of the 40 mm tool and squares of the 50 mm tool, filled where the cut is stable and open where it chatters. Case 1 is stable at 1.5 mm and chatters at 3 mm, case 3 steps from 18 500 to 17 500 rpm at 3 mm, and case 5 takes a bounded step from 16 300 to 11 410 rpm and then moves to 10 314 rpm. The pocket near 16 300 rpm rises above the frame to 17.9 mm.

with $w_0 = 2\,\mu\text{m/s}$ at the reference surface speed $v_0 = 750\,\text{m/min}$, a rate the application accelerates about a thousandfold over a real tool so that wear shows within one pocket.

13.3.7 Measurements

Neither controller reads the state. The spindle load $\bar{P}$ is the cutting power through the drive's 10 ms average, and it is what the load trend and the scorecard report. The feed loop reads P_m, that average passed through a first-order load-signal filter of $\tau_P = 20$ ms, delayed by θ_P and corrupted by white noise of 20 W. The vibration sensor reads $v = y + n_v$ at every physics step, with n_v white noise of 0.2 µm, and the encoder gives the spindle speed exactly.

13.3.8 Numerical Solution

The engine's fourth-order Runge-Kutta integrator (Section 1.6.1) advances the equations in steps of 20 µs, 50 steps per tooth period at 20 000 rpm. The surface u is a ring buffer of 2048 entries written once per completed step, and every Runge-Kutta stage reads $u(t - T)$ by linear interpolation between the two entries that bracket it, with T computed from the spindle speed of that stage. At the top speed of 24 000 rpm a tooth period of the three-tooth tool spans more than 41 steps, so every stage query falls inside the stored history and nothing is extrapolated. The engagement, the cutting coefficient and the mean-chip factor are refreshed once per step and held through its four stages, because the tool moves about 2 µm in a step while the engagement changes over millimeters. The surface update (13.4), the wrap of the tooth phase and the vibration sensor run after each step, and every random draw is made outside the derivative function, so the right-hand side stays deterministic within a step. The controllers execute every 1 ms, the historian logs at 100 Hz, and the raw vibration is kept in its own 4096-sample buffer for the scope and the spectrum. A flying start places the tool at the case's starting point on the path at the commanded feed and speed, with the static deflection $\mu F_t/k$ of the steady cut, the surface history filled with that deflection and $\theta = 0$, and it seeds the drive's load averages, the load signal and $\hat{b}$ at their steady values. A start at the entry begins at rest with $v_f = 0$ and $y = 0$. Settings changed before the first step are part of the program and re-seed this initial state as Reset does (Section 1.6.1), so a setting made before the run and the same setting carried through Reset give one run.

Table 13.2 Model Parameters. Class Marks a Value the Application Lets the User Adjust (adj.), One It Keeps Hidden, and One Fixed in Its Code

Symbol	Meaning	Value	Class
N, R	teeth, tool radius	3, 6 mm	adj., fixed
L	tool overhang	40 mm	adj.
f_{n0}, k_0	mode at $L_0 = 40$ mm	800 Hz, 12 N/µm	adj.
ζ	damping ratio	0.03	adj.
K_c	specific cutting force, sharp	800 N/mm^2	adj.
μ	directional factor	0.6	hidden
β_1, β_2	tooth-pass ripple	0.15, 0.05	hidden
VB_0	flank wear at the start	0.05 mm	adj.
k_{VB}	force rise with wear	1.2 mm^{-1}	adj.
K_{rub}	rubbing force per wear area	40 N/mm^2	adj.
w_0	wear rate at 750 m/min	2 µm/s	adj.
J	spindle inertia	0.004 kg m^2	hidden
$P_{\max}$, $T_{\max}$	spindle ratings	5.5 kW, 10 N m	adj.
$\alpha_{\max}$	spindle acceleration	8000 rpm/s	adj.
K_ω	drive speed bandwidth	100 rad/s	hidden
c_Ω	bearing drag	2×10^{-5} N m s	hidden
$a_{\max}$	feed axis acceleration	3 m/s^2	adj.
τ_F	feed servo lag	5 ms	hidden
τ_P, θ_P	load-signal filter, delay	20 ms, 0 ms	adj.
σ_P, σ_v	load, vibration noise	20 W, 0.2 µm	adj.
—	load cap (challenge)	5.5 kW (2.2 kW)	adj.
—	waviness mark threshold	5 µm	adj.
—	drive load average	10 ms	fixed
—	drive torque estimate	2 ms	fixed
Δt, Δt_c	physics step, control period	20 µs, 1 ms	fixed

The delay treatment was checked against Equation (13.7). A tap on the tool on a clean straight, with the ripple, the noise and the wear removed, gives a free vibration whose logarithmic envelope grows or decays, and bisection on the depth places the change of sign at 2.103 mm against the analytic 2.101 mm at 20 000 rpm, 0.095 % apart, and at 6.011 mm against 6.007 mm on the $k = 1$ lobe at 12 000 rpm, 0.064 % apart. Halving the step to 10 µs changes the growth rate at $1.1\,a_{\text{lim}}$ by 0.79 % and the amplitude of a fully developed chatter limit cycle by 0.071 %. The parameters are listed in Table 13.2.

13.3.9 Assumptions

The force is the average-tooth, or zeroth-order, force. Real milling forces vary with the angle of every tooth, and below about a quarter of the diameter of radial engagement that variation adds period-doubling lobes and shifts the boundary [11]. The pocket is cut at half and full engagement, where the average is closer, although the thinning engagement at the end of each loop lies outside that range.

One bending mode along the chip-thickness direction represents the tool, with a fixed directional factor of 0.6. A real tool has two orthogonal modes, the holder and the spindle add their own, and the effective direction turns with the cut [4].

The cutting force is linear in the chip, with no size effect and no process damping. Process damping from the flank rubbing on the surface waves raises the limit sharply at low surface speed [4], so the lobes below about 5000 rpm are pessimistic.

The engagement is measured once from the planned toolpath. The vibration does not change what the next pass sees in plan view, and the axes take every corner at the commanded feed with no corner slowdown.

Wear is accelerated about a thousandfold and acts linearly, raising K by $k_{VB}VB$ and adding a rubbing force proportional to the wear land.

The load signal is the drive's estimate of the cutting power, and the vibration sensor reads the tool deflection with white noise and no dynamics of its own. A real accelerometer or microphone sees the tool through the spindle housing, and the workpiece and the machine frame are rigid here.

13.4 Two Controllers on One Machine

Three models should be kept apart. The simulation model is the delay equation with contact loss of Section 13.3. The feed loop's internal model is a static gain, the load per unit override, identified on line. The chatter watch has no model of the tool at all, and uses only the measured chatter frequency. The lobe diagram the application draws is computed from Equation (13.7) for display and is read by neither controller.

13.4.1 Spindle Drive and Feed Axis

The actuators are machine hardware and are not tuned by the user. The speed command is clamped between 3000 and 24 000 rpm, its reference ω_{ref} slews at $\alpha_{\max} = 8000\,\mathrm{rpm/s}$, and the drive torque is $T_d = J[K_\omega(\omega_{\mathrm{ref}} - \Omega) + \dot{\omega}_{\mathrm{ref}}] + \hat{T}_L + c_\Omega\Omega$, held inside the limit of Equation (13.8), where $K_\omega = 100\,\mathrm{rad/s}$ and $\hat{T}_L$ is a 2 ms average of the cutting torque. The feed command is $v_{\mathrm{cmd}} = \min(uv_{\mathrm{prog}}, \sqrt{2a_{\max}(s_{\mathrm{end}} - s)})$, where v_{prog} is the programmed feed F, so the axis stops exactly at the end of the path, and its reference slews at $a_{\max} = 3\,\mathrm{m/s^2}$. If the speed falls below half its reference the spindle is declared stalled and the feed stops.

13.4.2 Adaptive Feed on the Measured Load

To first order the feed loop sees a static gain from override to load, $b = Ka_p a_e v_{\mathrm{prog}}$, behind the feed-axis lag, the drive's 10 ms average, the 20 ms load-signal filter and any delay θ_P. On the default straight $b = 0.64\,\mathrm{kW}$ per unit override, and it doubles in a slot or a corner. Each execution, every $\Delta t_c = 1\,\mathrm{ms}$, first updates the identified gain, Equation (13.10).

$$\hat{b}_k = \mathrm{clamp}\Big(\hat{b}_{k-1} + \frac{\Delta t_c}{\tau_b}\Big(\frac{P_{m,k}}{\tilde{u}_k} - \hat{b}_{k-1}\Big),\ b_{\mathrm{lo}},\ b_{\mathrm{hi}}\Big) \tag{13.10}$$

where $\tilde{u}$ is the applied override passed through the same filter τ_P and the same delay θ_P as the load signal, though not through the drive's 10 ms average, $\tau_b = 0.1\,\mathrm{s}$, $b_{\mathrm{lo}} = 0.2\,\mathrm{kW}$ and $b_{\mathrm{hi}} = 20\,\mathrm{kW}$, and the update runs only while $P_m > 100\,\mathrm{W}$ and $\tilde{u} > 0.05$. The proportional gain is then $K_{P,k} = K_f/\hat{b}_k$, so the loop gain, the product of $K_{P,k}$ and the true b, stays near the dimensionless K_f on a straight, in a slot and at a corner. With adaptation switched off, $\hat{b}$ is replaced by the sharp-tool gain of a 50 % straight, $K_c a_p R\, v_{\mathrm{prog}}$.

The loop runs the velocity-form PID algorithm of Section 1.6.2 with a zero derivative time, written for this loop in Equations (13.11) to (13.14), where $\hat{u}$ is the override before the clamp and p the proportional term.

$$e_k = P_{\mathrm{SP}} - P_{m,k}, \qquad p_k = K_{P,k}\, e_k \tag{13.11}$$

$$S_k = S_{k-1} + (p_k - p_{k-1}) \tag{13.12}$$

$$\hat{u}_k = S_k + I_{k-1}, \qquad u_k = \mathrm{clamp}(\hat{u}_k, 0.1, 1.5) \tag{13.13}$$

$$I_k = I_{k-1} + \Big[\frac{K_{P,k}}{\tau_I}\, e_k + \frac{u_k - \hat{u}_k}{T_t}\Big]\Delta t_c \tag{13.14}$$

These are Equations (1.12) to (1.15) with S_k never clamped, so S_k equals p_k at the gain in force and a change of $\hat{b}$ reaches the whole proportional term at once. The back-calculation uses the tracking time $T_t = \tau_I$, and with no bias the integral carries the initial override from the reset. No slew limit is placed on the override, because the feed axis ramps at its own acceleration limit. The error is $P_{\mathrm{SP}} - P_m$, so a load above target lowers the override, the reverse-acting sense that a positive process gain requires. The defaults are $K_f = 0.6$ and $\tau_I = 50\,\mathrm{ms}$ with adaptation on.

Three conditions replace the update by tracking, in which the integral is set so that $S_k + I_k$ equals the held override exactly. Below $0.2\,P_{SP}$ the tool is taken to be cutting air, the override is held at $\min(u_{k-1}, 1)$, and the loop leaves the air cut without the jump to 150 % that an unattended integrator would bring to the next engagement. On the stopping ramp at the end of the path the axis, not the override, sets the feed, and the override is frozen. With the feed control set to Constant, or in MANUAL, the override is the operator's value. Switching to Adaptive is therefore bumpless. In AUTO each manual station follows its controller's output (Section 1.6.2), and the header MANUAL and AUTO switch moves both loops at once, so a loop handed to the operator starts from the value it last applied. Reset keeps the selected modes. Returning to AUTO restores the feed and spindle modes selected before the handover, including after Reset, while a freshly loaded case supplies its own modes. AUTO does not enable Adaptive feed or Chatter control unless that mode was selected.

13.4.3 Chatter Detection

The chatter watch runs at the physics rate. Forced vibration repeats every tooth and chatter does not, so the watch subtracts a synchronous average of the vibration kept in 32 bins of tooth phase and compares the energy of the remainder with the energy of the whole, as shown in Equation (13.15).

$$\begin{aligned} r_n &= v_n - \bar{v}_{b(n)}, \qquad b(n) = \lfloor 32\,\theta_n / 2\pi \rfloor \\ CR &= \min\left(1,\ \frac{\overline{r^2}}{\overline{(v-\bar{v})^2} + E_0}\right) \end{aligned} \tag{13.15}$$

where each bin averages its first 20 visits equally and then updates with a weight of 0.05, about 13 tooth periods of memory at 20 000 rpm, the overbars are first-order averages of 20 ms, $\bar{v}$ is a 50 ms running mean, and $E_0 = (1\,\mu\text{m})^2$ keeps a quiet cut from reading as chatter. The chatter frequency f_c is the count of upward zero crossings of r in the last 40 ms, less one, over the time between the first and the last, with a hysteresis of $\max(3\sigma_v, 0.3\,r_{rms})$ and at least six crossings. Chatter is declared when CR stays above the threshold $CR_{th} = 40\,\%$ for a dwell of 40 ms. The dwell may run only after 50 ms of learning, and a change of the speed command by more than 1 % clears the bins so that the new tooth-pass pattern is learned again. This is the energy-ratio idea of [15], with a synchronous average in place of a Kalman filter.

13.4.4 Speed Selection

When chatter is declared with f_c between 150 and 4000 Hz, the spindle is moved to a speed at which a whole number j of chatter waves fits between two teeth, Equation (13.16).

$$\Omega_j = \frac{60 f_c}{N j}, \qquad j \in \{\lfloor q \rfloor - 1,\ \lfloor q \rfloor,\ \lceil q \rceil,\ \lceil q \rceil + 1\} \tag{13.16}$$

with speeds in rpm, $j \geq 1$ and $q = 60 f_c / (N\Omega)$ the ratio of the chatter frequency to the tooth-passing frequency. At Ω_j the tooth period holds exactly j periods of the measured chatter, so the wave a tooth leaves is met in phase by the next and the chip variation it regenerates vanishes, the principle of spindle speed regulation [13]. Because f_c lies just above the natural frequency, Ω_j falls near the top of the stable pocket between lobes $j-1$ and j. Among the candidates inside the speed range the one nearest the present speed is chosen, and if it is more than the largest step, 30 % by default, the spindle takes the bounded step toward it instead. The supervisor then waits until the speed is within 1 % of its target, waits a further 0.15 s, resets the dwell and watches again.

🕮 **Pause.** Try Exercise 13.2 before reading on.

13.4.5 Finish and Scoring

Once per tooth pass the true deflection is recorded, and the waviness is the rms residual of the last 16 records about their least-squares line, so a ramp in the deflection where the engagement changes is

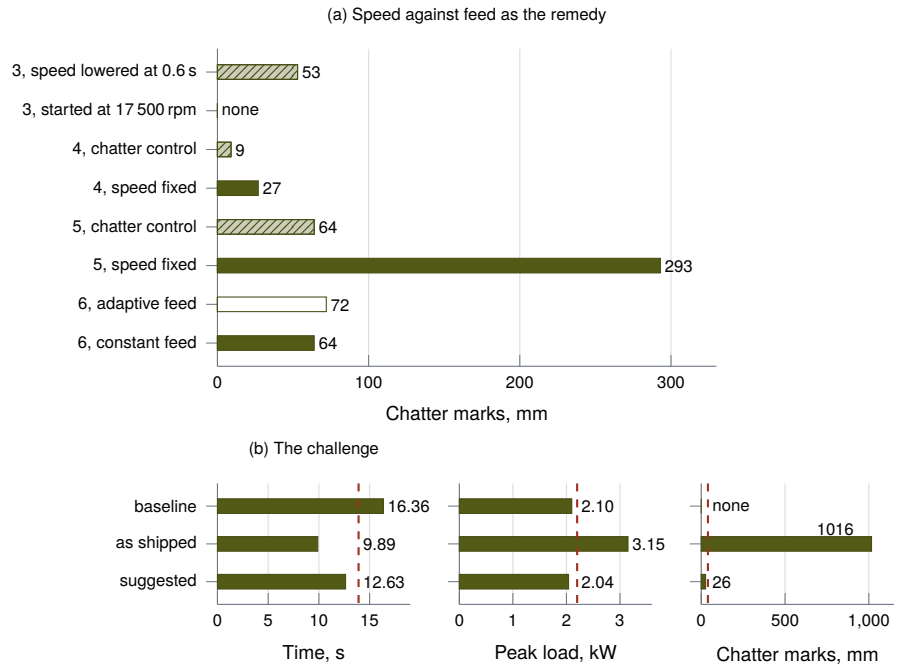


Figure 13.3 Rows of Table 13.4. (a) Chatter marks on cases 3 to 6: hatched bars where the spindle speed was changed or chosen, solid bars where nothing acted on the chatter, and the open bar where the adaptive feed slowed the cut. (b) The challenge, with its targets of 13.91 s, 2.2 kW and 40 mm as dashed lines. Notice that in cases 4 and 5 the speed change leaves far fewer marks than a fixed speed, while the slower feed of case 6 leaves more marks than a constant feed, and that only the suggested configuration meets all three targets.

not counted as waviness. A millimeter of path is marked as chattered when the waviness stays above 5 µm for 24 consecutive tooth passes, which the free ring-down after a corner cannot sustain, and only where material is being removed. The scorecard reports the machining time, the peak of the 10 ms load, the length of chatter marks, the removal rate, the peak rms vibration, the mean waviness, the flank wear and the number of speed changes. Beside them the framework's scorecard (Section 1.6.4) scores the load loop, and the chatter indicator as a loop with setpoint zero.

13.5 Results on the Pocket

Table 13.3 lists the cases. Each one cuts a single stretch of the pocket from a flying start, except case 8 and the challenge, which begin at the entry hole. The values in Table 13.4 were measured on the running engine at the default seed (Section 1.6.4), with each case as shipped unless the condition names a change made before the run. Figure 13.3 draws the chatter marks of cases 3 to 6 and the three challenge rows of the table.

13.5.1 The Load Scales and the Stability Does Not

Case 1 runs the long top side of the outer loop, a 92 mm straight at 50 % engagement, at 1.5 mm, 20 000 rpm and 4800 mm/min, a feed of 0.08 mm per tooth. The load is 0.64 kW, and the only vibration is forced, 0.5 µm rms at the 1000 Hz tooth-passing frequency, with the chatter indicator at 3 % and the waviness at most 0.33 µm. Raised to 3 mm the load doubles to about 1.3 kW, but the depth is now 43 % over the 2.10 mm limit. The waves grow until the tool loses contact for part of each cycle, and the cut settles into chatter at 823 Hz, just above the 800 Hz mode and off the tooth line,

Table 13.3 The Eight Cases and the Challenge

#	Stretch and settings	Teaching point
1	loop 4 top, 1.5 mm, 20 000 rpm, constant feed	load against limit
2	loop 3, 1.5 mm, adaptive feed at 0.65 kW	the corner overshoot
3	loop 4 top, 3 mm, 18 500 to 17 500 rpm at 0.6 s	the lobes
4	loop 4 bottom, insert $r_h = 1.6$, 3.5 mm, both loops	feed for load, speed for chatter
5	loop 4, $L = 50$ mm, 16 300 rpm, chatter control	lobes belong to the tool
6	loop 4 top, $VB_0 = 0.25$ mm, 1.8 mm, adaptive at 0.77 kW	load restored, finish not
7	loop 3, $\theta_P = 80$ ms, 3 mm, 16 300 rpm, $K_f = 1.2$, $\tau_I = 30$ ms	dead time in the feed loop
8	slot and loop 1, 8 mm, 16 300 rpm, constant feed	the power limit
C	whole pocket, 3 mm, insert, 2.2 kW cap, 20 000 rpm, adaptive at the cap	both loops together

Table 13.4 Results Measured on the Running Engine

Case and condition	Result
1, 1.5 mm	0.64 kW, 0.5 µm rms at 1000 Hz, *CR* 3 %, limit 2.10 mm
1, 3 mm	chatter 823 Hz, 60 µm waviness, 88 mm of 92 mm marked
1, 3 mm at 16 300 rpm	8.8 µm rms at 815 Hz, waviness 0.7 µm, no marks
2, constant / adaptive	corner peak 1.27 / 0.87 kW, time 3.38 / 3.65 s
2, $K_f = 1.2$, $\tau_I = 30$ ms	corner peaks 0.74–0.78 kW, run peak 0.84 kW at the step-out
3, as shipped	chatter 812 Hz, 69 µm, 53 mm marked, below 2 µm 0.21 s after the change
3, start at 17 500 rpm	at most 2.1 µm, no marks
4, as shipped	override 63 %, peak 1.97 kW, 811 Hz to 16 212 rpm, 9 mm marked
4, speed fixed / feed also constant	27 mm, 114 µm / peak 2.50 kW
5, as shipped	546 Hz to 11 410 rpm, 516 Hz to 10 314 rpm, 64 mm marked
5, speed fixed	293 mm marked
6, adaptive / constant	71 %, load within 1.6 %, 32 µm, 72 mm, 1.65 s / 1.09 kW, 42 µm, 64 mm, 1.19 s
7, θ_P 80 ms / detuned / 0 ms	peak 3.65 / 3.19 / 2.08 kW
8, constant / adaptive	6.78 kW, 14 675 rpm / 6.29 kW, 16 227 rpm at the first corner
C, baseline	16.36 s, 2.10 kW, no marks
C, as shipped	9.89 s, 3.15 kW, 1016 mm marked
C, suggested	12.63 s, 2.04 kW, 26 mm marked

with 60 µm of waviness and 88 mm of the pass marked. At 16 300 rpm the same 3 mm cut leaves no marks, because the limit there is 17.9 mm. The tool shakes harder, 8.8 µm rms, because the teeth now pass at 815 Hz, close to its natural frequency, but every tooth shakes it the same way, the indicator reads 0.4 %, and the waviness stays at 0.7 µm. The load scaled with the depth and the stability did not.

The deepest cut that leaves no marks can also be found by hand. Mapped in case 1 at a series of fixed speeds, it traces the lobe of Figure 13.2 to within 2 % between 17 000 and 22 000 rpm. From 14 000 to 16 000 rpm the pocket lies above the 12 mm depth the control allows, so no depth on the slider chatters there.

13.5.2 A Corner Arrives Before the Loop Sees It

Case 2 runs once around the third loop at 1.5 mm and ends with the 6 mm step across into the band of the outer loop. At constant feed the corners reach 1.27 kW against 0.64 kW on the straights. With the adaptive feed holding 0.65 kW, the override falls to between 56 and 58 % into each corner and the load is held, but the loop learns of a corner only through the drive's average and the load-signal filter, after the engagement has begun to grow, so each corner still peaks at 0.86 to 0.87 kW, 33 to 34 % over target. A gain of 1.2 with a 30 ms integral time pulls the override to 43 to 45 % at the corners and holds their peaks to 0.74 to 0.78 kW, 15 to 20 % over. Where the loop closes on its own start the engagement thins to air, and the override runs to 150 % until the air-cut hold returns it to about

100 %. The step across that follows is nearly a full slot met at that held override. It sets the lowest override of the run, 44 % with the stock tuning and 26 % with the faster one, and for the faster tuning also the highest load, 0.84 kW. At this target the adaptive run is slower, 3.65 s against 3.38 s, because it slows for every corner while the straights need only 100 %. The time is won with a higher target, as the challenge shows.

13.5.3 Five Percent of Spindle Speed

Case 3 cuts the top side at 3 mm and lowers the speed from 18 500 to 17 500 rpm at 0.6 s, a change the spindle makes in 0.125 s at its acceleration limit. The limit is 2.71 mm at the first speed, with chatter predicted at 810 Hz, and 4.00 mm at the second, as shown in Figure 13.2. At 18 500 rpm the cut is 11 % over its limit, the waviness passes 5 µm at 0.11 s, the first mark is recorded at 0.14 s once it has stayed above that level for 24 consecutive teeth, and chatter at 812 Hz grows to 69 µm of waviness. After the change the cut is 25 % under its limit, and the waviness is below 2 µm at 0.81 s, 0.21 s after the command, with 53 mm of the pass marked. Started at 17 500 rpm, the same cut leaves at most 2.1 µm and no marks. The load is 1.3 kW both ways.

> **TRY IT IN THE APP 3 · Stable and Unstable Speeds**
>
> Load the scenario, press Run and watch the spectrum and the vibration scope as the program lowers the speed from 18 500 to 17 500 rpm at 0.6 s. Chatter at 812 Hz grows to 69 µm of waviness, and 0.21 s after the change the waviness is below 2 µm. The load is 1.3 kW both ways.

13.5.4 Two Faults at the Insert

Case 4 crosses the cast-iron insert, 1.6 times harder, at 3.5 mm and 17 500 rpm with both loops on. The limit is 4.00 mm in the aluminum and 2.50 mm in the iron. At the insert the load jumps by 60 %, the feed loop pulls the override to 63 %, and the load peaks at 1.97 kW against the 1.5 kW target. The cut is then 40 % over its new limit. Eleven millimeters after the tool meets the iron, the chatter watch declares chatter at 811 Hz and moves the spindle to 16 212 rpm, where one chatter wave fits between two teeth, and the pass ends with 9 mm of marks. With the speed fixed the insert chatters to its end, with 27 mm of marks and a worst waviness of 114 µm, and with the feed constant as well the load reaches 2.50 kW. Each loop answers the fault it can see.

> **TRY IT IN THE APP 4 · A Harder Patch**
>
> Press Run and watch the load, the feed override and the chatter indicator as the tool meets the iron. The override falls to 63 % and the load peaks at 1.97 kW, then the chatter watch moves the spindle to 16 212 rpm and the pass ends with 9 mm of marks. Press Reset, set Spindle speed to Fixed and run again: the marks grow to 27 mm.

13.5.5 A Longer Tool and a Worn One

Case 5 fits a tool with 50 mm of overhang, 6.14 N/µm and 512 Hz, and starts it at 16 300 rpm, the best speed of the short tool, where the long tool's limit is 1.19 mm. At 1.5 mm it chatters at once, and at 0.09 s the watch reads 546 Hz, near its own mode. The nearest whole-wave speed, 10 925 rpm, is 33 % away, so the first move is the bounded step to 11 410 rpm. On the straight that speed is linearly stable, with a limit of 1.75 mm, and a run started there leaves no marks. The chatter grown during the ramp through unstable speeds is still decaying when the 0.15 s settling time ends, and the indicator still reads above 90 %, so at 0.88 s the watch reads 516 Hz and moves on to 10 314 rpm, where the limit is 10.9 mm. The loop ends with 64 mm of marks, against 293 mm at the fixed speed. The energy ratio cannot tell a decaying chatter from a growing one, which is why the second move happens.

Pause. Try Exercise 13.3 before reading on.

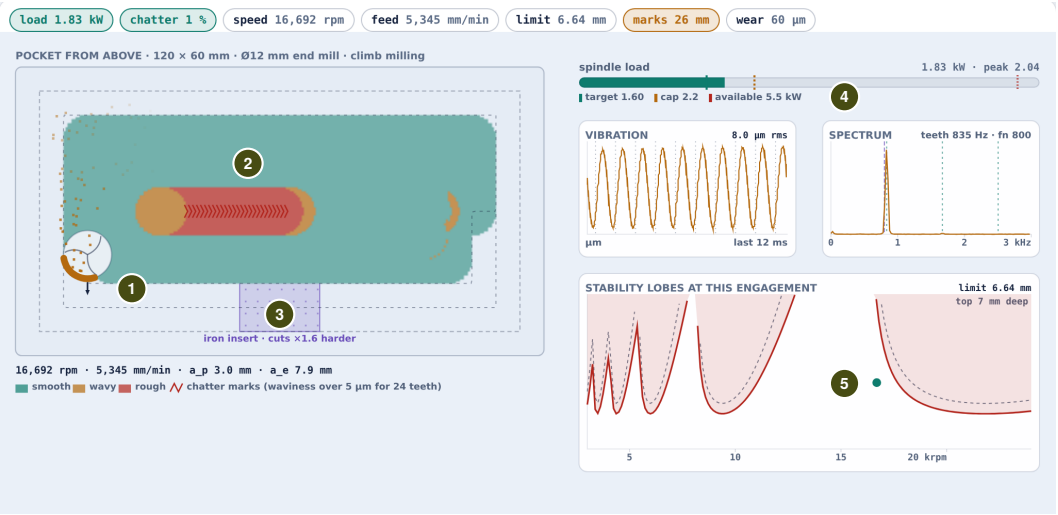


Figure 13.4 The CNC milling application at $t = 7.50\,\mathrm{s}$ of the challenge with the suggested configuration, a 1.6 kW load target, $K_f = 1.5$, $\tau_I = 30\,\mathrm{ms}$ and chatter control, cropped to the plant view. The pocket seen from above shows the start of the entry slot along the middle painted rough and crossed with zigzag chatter marks (2), the first two loops and most of the third cleared smooth around it, the tool on the left side of the third loop heading into its lower corner with its arc of contact drawn (1), and the cast-iron insert below (3). Beside it the load bar (4) reads 1.83 kW against the 1.60 kW target and the 2.2 kW cap with a 2.04 kW peak, the scope and spectrum show 8.0 µm rms of forced vibration at the 835 Hz tooth-passing frequency beside the 800 Hz natural frequency, and the lobes for the present 7.9 mm engagement place the operating point (5) at 16 692 rpm and 3 mm in a stable pocket below its 6.64 mm limit.

Case 6 runs a tool with 0.25 mm of flank wear on the top side at 1.8 mm and 20 000 rpm. The wear raises K by 30 % over a sharp edge, the load at constant feed peaks at 1.09 kW against the 0.77 kW of the default tool, and the limit falls from 2.10 to 1.71 mm. The adaptive feed settles near 71 % and holds the load within 1.6 % of 0.77 kW, and the cut chatters anyway, with 32 µm of waviness and 72 mm of marks, against 42 µm and 64 mm at constant feed. The slower feed made the waves about a quarter shallower, since the limit cycle scales with the chip, and made the pass 39 % longer, 1.65 s against 1.19 s, while the limit did not move.

> **DID YOU KNOW? A Steady Load Over a Chattering Cut**
>
> From the simulator: with a worn tool the adaptive feed holds the load within 1.6 % of its target, yet the pass leaves 72 mm of chatter marks against 64 mm at constant feed (Table 13.4). A load held on target says nothing about the finish.

13.5.6 Dead Time and the Power Limit

Case 7 delays the load signal by 80 ms on the third loop at 3 mm and 16 300 rpm, where the cut is stable, with a tuning that holds a 1.7 kW target to a 2.08 kW corner peak when the signal is prompt. With the delay the override swings between its 10 % and 150 % limits and the load peaks at 3.65 kW in the first corner. Detuned to a gain of 0.3 and an integral time of 200 ms, the loop stops swinging and the override stays above 87 %, but it is then too slow for the corners and they peak at 3.19 kW. Neither tuning comes near the 2.08 kW of the prompt signal, because dead time takes phase without taking gain. Volume 5, Chapter 3 measures the same loss in a composition loop, where a tuning made for a prompt measurement fails once the analyzer is slow.

Case 8 spends the depth that the pocket at 16 300 rpm allows. At 8 mm, below the slot's limit of 8.9 mm, the entry slot needs 6.78 kW from a 5.5 kW spindle. The drive holds its torque limit for 1.64 s, the rest comes out of the rotor, and the speed sags to 14 675 rpm, 10 % down, with every tooth then taking a thicker chip. With the adaptive feed at a 4.5 kW target the override falls to 51 % at the entry and holds near 65 % along the slot, the entry peaks at 5.41 kW, and through the slot the speed stays within 20 rpm of its command. The overload moves to the first inside corner of loop 1, which the loop meets at the override of about 134 % that the half-width straight needs, so the load reaches 6.29 kW there and the speed dips to 16 227 rpm. Once chatter is out of the way, power sets the removal rate and a load loop is still needed. The feed rate of a combine harvester in Volume 4, Chapter 4 meets the same kind of limit, an engine that slows everything it drives when the crop asks for more power than it has.

13.5.7 The Graded Challenge

The challenge clears the whole pocket at 3 mm with the insert under the bottom passes. The fixed-feed baseline, 80 % override at 16 300 rpm, finishes in 16.36 s with a 2.10 kW peak, no marks and a removal rate of 78 cm^3/min. The targets are a time of at most 13.91 s, 85 % of that baseline, a peak load of at most 2.2 kW and at most 40 mm of marks, each scored as the worst case of the whole run. As shipped, with the adaptive target at the cap and the spindle at 20 000 rpm, the pocket takes 9.89 s, but the corners reach 3.15 kW and chatter marks 1016 mm of the 1044 mm path, and the load spends 1.99 s above the cap.

The suggested configuration lowers the target to 1.6 kW, raises the gain to 1.5 with a 30 ms integral time and turns the chatter control on. The chatter watch reads 835 Hz in the slot at 0.09 s, the learning time plus the dwell, and moves the spindle to 16 692 rpm. The 26 mm of marks all lie in the first 28 mm of the slot, and nothing else on the pocket chatters. At each of the three steps from an air cut into a new loop the indicator crosses its threshold for only 13 to 15 ms, inside the dwell. The highest load, 2.04 kW, comes at the first of those steps, where the tool leaves the air cut that closes loop 1 for the band of loop 2, while the insert peaks at 2.00 kW and the corners at 1.92 kW. The pocket is finished in 12.63 s (Table 13.4), 23 % faster than the baseline at 101 cm^3/min, and Figure 13.4 shows the run at 7.50 s. A 1.5 kW target peaks at 1.96 kW but takes 13.50 s, and 1.7 kW finishes in 11.88 s but peaks at 2.14 kW in the insert, 2.8 % under the cap. The target trades time against the cap, and the speed is a separate decision.

Judged by the challenge's own rule, the shipped configuration fails two of the three targets, with a peak 43 % over the cap and 1016 mm marked, while the suggested one passes with 1.27 s, 0.16 kW and 14 mm to spare. Changing a model parameter or the insert voids an attempt.

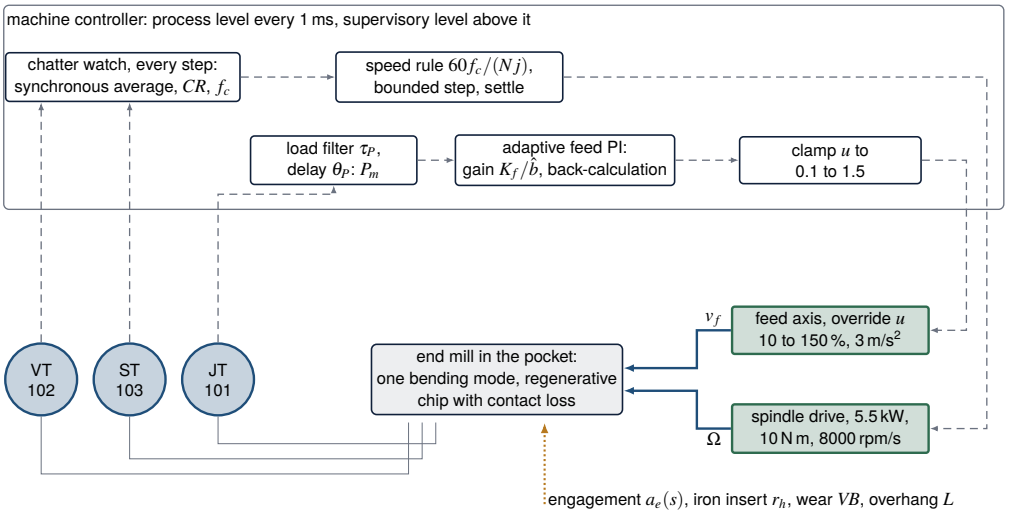


Figure 13.5 The two loops on machine-tool hardware. The load JT-101 passes through the load-signal filter and delay to the adaptive feed loop, whose clamped override drives the feed axis. The vibration VT-102 and the spindle speed from ST-103 feed the chatter watch, and the speed rule sends a new command to the spindle drive. Notice that the two loops share the cut but no signal: the feed loop never reads the vibration, and the speed rule never reads the load.

13.6 Technology in Practice

What Measures It

The two controllers of this chapter read two signals, a load and a vibration. In service the load usually comes from the spindle drive as motor current or power, and the vibration from a sensor on the machine or in the tool holder, or from a microphone [14, 17]. The chapter's load signal is the cutting power through the drive's 10 ms average, a 20 ms filter, an optional delay and noise, and its vibration signal is the tool deflection with white noise and no sensor dynamics.

> **TECHNOLOGY** **Sensor: Load and Vibration on the Machine**
>
> Reviews of machining monitoring rank cutting force and torque, motor current and power, vibration, temperature and acoustic emission as the most significant signals. Their sensors sit in the spindle and axis drives, on the frame, the table or the fixture, or in the tool holder, and they detect small changes better the closer they are to the cut [17]. The microphone named in Section 13.2 hears the whole room, and isolation, collection or intensity methods are needed to reject the background noise [14]. On a rotating tool holder the sensor's own rotation disturbed the acceleration signals, though their rms still marked chatter clearly [17]. The chapter's load signal stands for the drive's power estimate and its vibration signal for a sensor at the tool.

What Moves It

Two final elements share the cut. The feed override scales the programmed feed, and the feed axis follows it through its servo. The spindle drive turns the tool, and the chapter's drive is torque limited below its base speed and power limited above it. A speed change is made by the controller that sits above both and that, in a conventional numerical control, runs the program open loop with no process feedback [18].

Where the Controller Runs

Machining control is organized in servo, process and supervisory levels [9]. In an open machine controller architecture, the trajectory generator sends position and velocity setpoints at the interpolation interval to a servo module that computes the actuator commands at a higher rate. Constant-force machining is named there as a case in which the trajectory itself needs sensor feedback, and chatter control is one of the feedback functions such a controller was meant to add [18]. The chapter places its feed loop at the process level, running every 1 ms, and its chatter watch and speed rule at the supervisory level, with the watch reading the vibration at the 20 µs physics rate. Figure 13.5 draws both loops on that hardware.

> **TECHNOLOGY** **Final Element: The Spindle and the Feed Override**
>
> In a numerical controller the spindle normally runs independently of the axes, held at a constant velocity for machining and changed infrequently. The axes follow setpoints that a trajectory generator picks off the programmed path at each interpolation interval [18]. A change of spindle speed is therefore a supervisory decision, as it is in this chapter. Chatter accelerates the wear of the tools and of machine parts such as the spindle bearings [17], so a cut left chattering wears the final element itself. The chapter's spindle is power limited above its base speed, and its feed override scales the programmed feed between 10 and 150 %.

What Is Modernizing It

Three developments are changing the hardware around these loops. Tool holders now carry strain gauges, an accelerometer and a temperature sensor, powered by a generator that draws on the tool's own rotation and sending their data wirelessly. In one such holder the bending moment gave a reliable chatter feature [17]. Active chatter control applies a computed force in real time through actuators in the spindle, in the tool holder or on the workpiece, such as piezoelectric stacks and magnetic bearings, where passive tools with variable pitch or helix break up the regeneration by their geometry [19]. And the field's own review asks for physics-based stability prediction to be joined with on-line adaptive learning and tuning, so that detection and avoidance stay correct as the machine and the tool change [4]. The chapter's speed rule uses only the measured chatter frequency, the simplest member of that family.

> **IN PRACTICE** **Matching the Teeth to the Tool's Mode**
>
> Machine tools are tested dynamically, and their designers identify and stiffen the mode shapes that limit stability. A rule from high-speed milling matches the tooth-passing frequency, the spindle speed times the number of teeth, to the first bending mode of the spindle, so that the machine runs in the highest-speed lobe and removes metal fastest. The tool-point response is measured by modal testing, which needs equipment and expertise that a production shop may not have [4]. Case 1 of this chapter shows the rule: at 16 300 rpm the teeth pass at 815 Hz, beside the 800 Hz mode, and a 3 mm cut leaves no marks.

13.7 Limitations

The average-tooth force misses the extra lobes of light radial immersion [11], and the thin engagement at the end of each loop falls in that range.

One mode along a fixed direction stands in for the tool, the holder and the spindle, whose several modes shift and split the lobes of a real assembly [4].

There is no process damping and no size effect, so the lobes at low speed are pessimistic [4].

The engagement comes from the planned toolpath, the axes take corners with no slowdown, and wear is accelerated about a thousandfold.

The sensors are idealized. The load signal is the cutting power itself through a filter, a delay and noise, the vibration signal is the deflection with white noise, and there is no sound, so the microphone route to detection [14] is not represented.

The chatter watch has no growth test, so it reacts to the ring-down of chatter that its previous move has already cured, as in case 5. The speed rule uses only the measured frequency, and no supervisor chooses speeds from the lobes.

OPEN CHALLENGE **Knowing the Tool's Dynamics in Production**

A stability chart is only as good as the tool-point response, the cutting force coefficients and the process damping behind it. The response can change with the spindle speed, the cutting load and the position of the machine, a thin-walled part changes as metal is removed, and wear changes the rubbing that gives process damping. A review of the field calls for physics-based chatter prediction made off line to be combined with on-line adaptive learning and tuning [4]. Progress would be lobes that stay correct along the toolpath while the tool wears. The lobes of this chapter come from one fixed mode and are read by neither controller.

13.8 Conclusion

A pocket cut by a slender end mill is limited by a load that the feed can hold and a stability boundary that the feed cannot move, and the simulator of this chapter puts both on one toolpath with a controller for each. The regenerative chip is advanced as a delay differential equation with contact loss, and its simulated boundary lies 0.095 % from the analytic lobe at 20 000 rpm and 0.064 % at 12 000 rpm. In the uniform cut 1.5 mm draws 0.64 kW with 0.5 µm of forced vibration, while 3 mm doubles the load and chatters at 823 Hz with 60 µm of waviness against a 2.10 mm limit. The adaptive feed holds a corner to 34 % over its target where a constant feed reaches 95 %, and holds a worn tool's load within 1.6 % while the cut leaves 72 mm of marks. The chatter watch and the speed rule end chatter by moving the spindle to $60 f_c/(N j)$, and with both loops the pocket is cleared in 12.63 s against a 16.36 s baseline, under the 2.2 kW cap with 26 mm of marks. The model carries one tool mode, an average-tooth force and no process damping, so its lobes are pessimistic at low speed and incomplete at light immersion. A fuller model would add a second orthogonal mode, time-periodic force coefficients, process damping, a growth test in the chatter watch and a feedforward from the planned engagement. The simulator puts the two limits of a milled pocket, the one a feed can remove and the one only the spindle speed can, side by side where both can be measured.

KEY POINTS

- Slowing the feed lowers the load on the spindle, but it does not move the depth at which the cut starts to chatter.
- The critical depth is set by the tool's frequency response, the cutting stiffness and the tooth period, and the feed does not enter it.
- An adaptive feed loop that holds a worn tool's load within 1.6 % of its target still leaves 72 mm of chatter marks.
- A spindle speed 5.4 % lower ends the chatter of a cut 11 % over its limit within 0.21 s.

Where else this applies. Volume 4, Chapter 4 holds a harvester's feed rate against the power of its engine, the load half of this chapter's problem.

References

[1] J. Tlusty, "The theory of chatter and stability analysis," in *Machine Tool Structures*. Elsevier, 1970, pp. 133–177.

[2] H. E. Merritt, "Theory of self-excited machine-tool chatter: Contribution to machine-tool chatter research—1," *Journal of Engineering for Industry*, vol. 87, no. 4, pp. 447–454, 1965.

[3] J. Munoa, X. Beudaert, Z. Dombovari, Y. Altintas, E. Budak, C. Brecher, and G. Stepan, "Chatter suppression techniques in metal cutting," *CIRP Annals*, vol. 65, no. 2, pp. 785–808, 2016.

[4] Y. Altintas, G. Stepan, E. Budak, T. Schmitz, and Z. M. Kilic, "Chatter stability of machining operations," *Journal of Manufacturing Science and Engineering*, vol. 142, no. 11, p. 110801, 2020.

[5] Y. Altintaş and E. Budak, "Analytical prediction of stability lobes in milling," *CIRP Annals*, vol. 44, no. 1, pp. 357–362, 1995.

[6] J. Tlusty and F. Ismail, "Basic non-linearity in machining chatter," *CIRP Annals*, vol. 30, no. 1, pp. 299–304, 1981.

[7] E. Rogers, K. Galkowski, and D. H. Owens, "Examples and representations," in *Control Systems Theory and Applications for Linear Repetitive Processes*, ser. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2007, pp. 1–40.

[8] J.-P. Richard, "Time-delay systems: an overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003.

[9] A. G. Ulsoy and Y. Koren, "Control of machining processes," *Journal of Dynamic Systems, Measurement, and Control*, vol. 115, no. 2B, pp. 301–308, 1993.

[10] L. K. Lauderbaugh and A. G. Ulsoy, "Model reference adaptive force control in milling," *Journal of Engineering for Industry*, vol. 111, no. 1, pp. 13–21, 1989.

[11] M. A. Davies, J. R. Pratt, B. Dutterer, and T. J. Burns, "Stability prediction for low radial immersion milling," *Journal of Manufacturing Science and Engineering*, vol. 124, no. 2, pp. 217–225, 2002.

[12] J. Tlusty, "High-speed machining," *CIRP Annals*, vol. 42, no. 2, pp. 733–738, 1993.

[13] S. Smith and J. Tlusty, "Stabilizing chatter by automatic spindle speed regulation," *CIRP Annals*, vol. 41, no. 1, pp. 433–436, 1992.

[14] T. Delio, J. Tlusty, and S. Smith, "Use of audio signals for chatter detection and control," *Journal of Engineering for Industry*, vol. 114, no. 2, pp. 146–157, 1992.

[15] H. Caliskan, Z. M. Kilic, and Y. Altintas, "On-line energy-based milling chatter detection," *Journal of Manufacturing Science and Engineering*, vol. 140, no. 11, p. 111012, 2018.

[16] R. M. Centner and J. M. Idelsohn, "Adaptive controller for a metal cutting process," *IEEE Transactions on Applications and Industry*, vol. 83, no. 72, pp. 154–161, 1964.

[17] A. Schuster, A. Otto, H. Rentzsch, and S. Ihlenfeldt, "Multi-sensory tool holder for process force monitoring and chatter detection in milling," *Sensors*, vol. 24, no. 17, p. 5542, 2024.

[18] F. M. Proctor and J. Michaloski, "Enhanced machine controller architecture overview," National Institute of Standards and Technology, Gaithersburg, MD, NIST Interagency/Internal Report 5331, 1993.

[19] C. Wang, H. Ma, J. Chen, Z. Zhang, Q. Song, and Z. Liu, "Milling chatter control in low immersion condition with an active electromagnetic tool holder system," *Micromachines*, vol. 16, no. 3, p. 257, 2025.

13.9 Exercises

13.1 ⌕ Concept check · 2 min

A straight cut chatters at a fixed spindle speed with the feed override at 100 %. The operator lowers the override to 60 % and changes nothing else. What does the model of this chapter predict?

(a) The chatter goes on with shallower waves, because the feed does not enter the stability limit

(b) The chatter stops, because a thinner chip lowers the cutting stiffness below the stability limit

(c) The chatter stops, because a slower feed lengthens the tooth period and moves the cut into a pocket

(d) The chatter grows, because a slower feed raises the gain the adaptive loop identifies per unit feed

13.2 ⌕ Concept check · 1 min

The chatter watch moves the spindle to $\Omega_j = 60 f_c/(N j)$ for a measured chatter frequency f_c. Why does a speed chosen this way end the chatter?

(a) The tooth-passing frequency moves far from the natural frequency, so the teeth stop exciting the tool

(b) The tooth period holds a whole number of chatter waves, so each tooth meets the wave left before it in phase

(c) The surface speed falls, so the damping from the flank rubbing on the surface waves rises sharply

(d) The chip grows thinner in proportion to the speed, so the tool no longer leaves the cut on each cycle

13.3 Concept check · 2 min

In case 5 the chatter watch first takes the bounded step from 16,300 to 11 410 rpm, where the 1.5 mm cut of the long tool is stable, and then at 0.88 s moves on to 10 314 rpm. Why the second move?

(a) Flank wear grown during the ramp has lowered the limit at the first new speed below the depth

(b) The spindle overshot its new target and came back through a band of unstable speeds

(c) The chatter grown during the ramp is still ringing down, and the energy ratio reads it as chatter

(d) The first new speed put the tooth-passing frequency below the natural frequency, which always forces a move

13.4 Theory and calculation · 6 min

The tool mode of Equation (13.2) has $k_0 = 12\,\mathrm{N/\mu m}$ and $f_{n0} = 800\,\mathrm{Hz}$ at $L_0 = 40\,\mathrm{mm}$, with $\zeta = 0.03$. Take $\mu = 0.6$ and the cutting coefficient of the default tool, $K = K_c(1 + k_{VB}VB)$ with $K_c = 800\,\mathrm{N/mm^2}$, $k_{VB} = 1.2\,\mathrm{mm^{-1}}$ and $VB = 0.05\,\mathrm{mm}$.

(a) Find the modal mass m_0, and the stiffness and natural frequency of the tool at an overhang of 50 mm. *Checkpoint:* 0.475 kg, 6.14 N/μm and 512 Hz.

(b) Find the lowest limiting depth over all speeds, $a_{\mathrm{crit}} = 2k\zeta(1+\zeta)/(\mu K N^*)$, on a straight ($N^* = 0.75$) for both overhangs. *Checkpoint:* 1.94 mm and 0.995 mm.

(c) Say which of these cuts is stable at every spindle speed: case 1 at 1.5 mm with the 40 mm tool, case 5 at 1.5 mm with the 50 mm tool, and the 8 mm entry slot of case 8 ($N^* = 1.5$) with the 40 mm tool. *Checkpoint:* Only case 1. The slot's lowest limit is 0.97 mm.

13.5 Theory and calculation · 5 min

Use the steady load of Section 13.3, $\bar{P} = K a_p a_e v_f$ plus the rubbing term, in which the rubbing force $K_{\mathrm{rub}} VB\, a_p N^*$ with $K_{\mathrm{rub}} = 40\,\mathrm{N/mm^2}$ acts at the surface speed $R\Omega$, $R = 6\,\mathrm{mm}$. The default tool has $K = 848\,\mathrm{N/mm^2}$ and $VB = 0.05\,\mathrm{mm}$, and the feed is 4800 mm/min.

(a) Case 1 cuts a straight, $a_e = 6\,\mathrm{mm}$ and $N^* = 0.75$, at 1.5 mm and 20 000 rpm. Find the removal rate and the load. *Checkpoint:* 43.2 $\mathrm{cm^3}$/min and about 0.64 kW.

(b) Case 8 cuts the entry slot, $a_e = 12\,\mathrm{mm}$ and $N^* = 1.5$, at 8 mm and 16 300 rpm. Find the load and compare it with the 5.5 kW spindle. *Checkpoint:* About 6.76 kW.

(c) With $T_{\max} = 10\,\mathrm{N\,m}$, find the speed above which the drive is limited by power rather than torque. *Checkpoint:* 5252 rpm.

13.6 Theory and calculation · 7 min

Apply the speed rule of Equation (13.16) with $N = 3$, the speed range 3000 to 24 000 rpm and a largest step of 30 %. The candidate nearest the present speed is chosen.

(a) Case 4 declares chatter at 811 Hz with the spindle at 17 500 rpm. Find q, the candidate speeds and the new speed. *Checkpoint:* $q = 0.927$, candidates 16,220 and 8110 rpm, new speed 16 220 rpm.

(b) Case 5 declares chatter at 546 Hz with the long tool at 16 300 rpm. Find the nearest candidate and the speed the spindle actually takes. *Checkpoint:* 10 920 rpm is 33 % away, so the bounded step to 11 410 rpm.

(c) The watch then reads 516 Hz at 11 410 rpm. Find the next speed and the tooth-passing frequency there. *Checkpoint:* 10 320 rpm, with the teeth passing at 516 Hz.

13.7 App challenge · 8 min

Load scenario 1, Uniform cut, with the default seed. Each pass stops itself when it is complete, at 1.19 s, so any playback speed works. For each run read the Peak load, Chatter marks and Vibration (peak rms) tiles of the scorecard, and the limit line of the depth trend.

(a) Run the scenario as shipped, at 1.5 mm. *Checkpoint:* 0.64 kW, no marks, limit 2.10 mm.

(b) Press Reset, set Axial depth of cut to 3 mm before Run, and run the pass. *Checkpoint:* 1.32 kW, 88 mm, 60.1 µm.

(c) Press Reset, keep 3 mm, set Spindle speed to 16 300 rpm before Run, and run the pass. *Checkpoint:* 1.27 kW, no marks, 8.8 µm, limit 17.9 mm.

Accept loads within 0.02 kW, marks within 2 mm and vibration within 0.5 µm. Explain why run (c) shakes the tool far more than run (a) and still leaves no marks.

13.8 App challenge · 8 min

Load scenario 2, Corner engagement, with the default seed. Each run stops itself at the end of the loop. Read the Peak load and Machining time tiles.

(a) Run the scenario as shipped, with the adaptive feed at 0.65 kW, $K_f = 0.6$ and $\tau_I = 50$ ms. *Checkpoint:* 0.87 kW and 3.65 s.

(b) Press Reset, set Feed control to Constant before Run, and run again. *Checkpoint:* 1.27 kW and 3.38 s.

(c) Press the reload button, set Loop gain to 1.2 and Integral time to 30 ms before Run, and run again. Export the CSV file and find the lowest value of `feed_override_pct` with its `path_mm`. *Checkpoint:* 0.84 kW, 3.67 s, lowest override 25.9 % at about 730 mm.

Accept loads within 0.02 kW, times within 0.02 s and the override within 1 %. Say whether the faster tuning's peak and its lowest override are at a corner.

13.9 App challenge · 8 min

Load scenario 7, A late load signal, with the default seed. The load signal is delayed by 80 ms, and the loop runs $K_f = 1.2$ and $\tau_I = 30$ ms with a 1.7 kW target. Each run stops itself at the end of the loop. Read the Peak load tile, and the lowest feed override on the trend or in `feed_override_pct` of the CSV file.

(a) Run the scenario as shipped. *Checkpoint:* 3.65 kW, override down to 10 %.

(b) Press Reset, set Loop gain to 0.3 and Integral time to 200 ms before Run, and run again. *Checkpoint:* 3.19 kW, override never below about 87.6 %.

(c) Press the reload button, open Show Model, set the Load signal delay to 0 ms, and run the original tuning. *Checkpoint:* 2.08 kW.

Accept peaks within 0.02 kW and the override within 1 %. State which remedy recovers the corner peak and why detuning cannot.

13.10 Application area · 4 min

Case 6 runs a tool with 0.25 mm of flank wear on the case 1 straight at 1.8 mm and 20 000 rpm, with the adaptive feed holding the 0.77 kW that the default tool draws there.

(a) Find the factor by which the wear raises K over a sharp edge and over the default tool. With the rubbing term of Equation (13.5) at a surface speed of 12.6 m/s, estimate the load at constant feed. *Checkpoint:* 1.30 and 1.23, and about 1.07 kW.

(b) If the whole load scaled with the feed, what override would hold 0.77 kW? Name the logged signal a tool-condition monitor should watch while an adaptive feed is on. *Checkpoint:* About 72 %. The override, or the load per unit override.

13.11 Application area · 4 min

A shop changes from a 40 mm to a 50 mm tool overhang to reach a deeper pocket, and keeps the 16 300 rpm that worked before.

(a) Using $f_n \propto L^{-2}$, find the natural frequency of the long tool and the speed at which its tooth-passing frequency equals it, for $N = 3$. Compare with the 16 000 rpm of the short tool. *Checkpoint:* 512 Hz and 10 240 rpm.

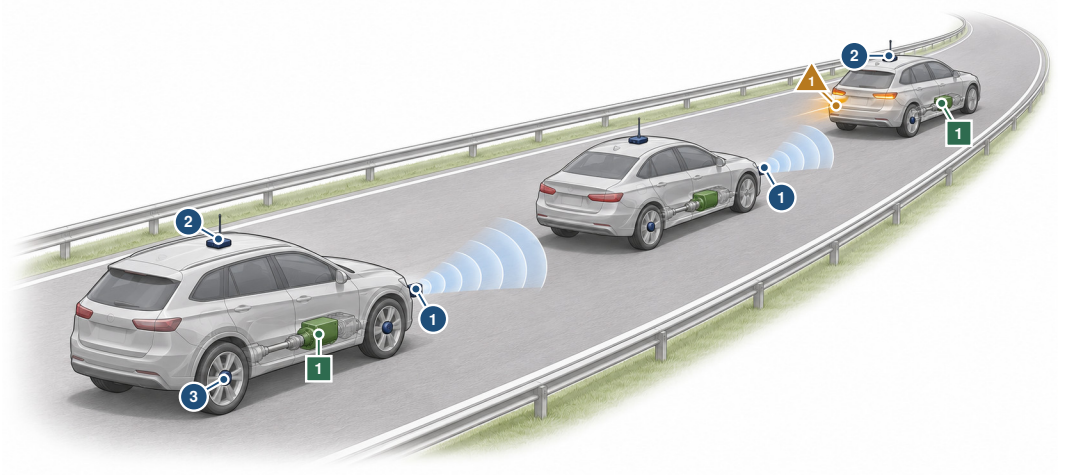
(b) Say what the shop should measure before trusting a speed, and name two features of a real tool assembly that this model leaves out. *Checkpoint:* A tap test of the tool in its holder. Several modes, two directions, process damping.

13.12 Review the agent · 5 min

An AI assistant reviews case 6, A worn tool, and replies as follows. “The worn tool chatters because it is overloaded: at constant feed it draws 1.09 kW against 0.77 kW for a good tool. Lower the Spindle load target from 0.77 to 0.50 kW. The adaptive feed will cut the chip by a third, the cutting force falls with it, and the 72 mm of chatter marks will disappear.” Find the flaw and say how you would check it on the app, and what would work instead.

CHAPTER 14

String Stability of Well-Tuned Followers in a Vehicle Platoon Simulator



The controller moves 1 each car's throttle and brakes, which follow its controller's command a little late

The controller reads 1 the radar that measures the gap to the car ahead and how fast it closes
2 the radio that hears the car ahead's acceleration 3 the car's own speed

What disturbs it 1 the lead car's brake tap

Notice that each car sees only the car directly ahead, by radar and by radio, so a change made by the lead reaches the last car only through every car between them. Illustration, not to scale.

IN PLAIN TERMS When a Line of Cars Passes a Wave Back

The main idea. A car that follows another on its own settles smoothly after the car ahead slows. In a line of ten such cars, a short tap on the first car's brake can grow into a much deeper slowdown at the back.

The challenge. Each car reacts only to the one ahead, and a little late. So a well-behaved car can still pass the wave back larger than it came. A longer gap or a radio link to the car ahead stops the growth, but one costs room on the road and the other can fail.

Why it matters. Growing waves like these are how traffic jams form with no crash or road work to cause them.

THREE WAYS TO READ THIS CHAPTER

Explore	12 min	the first page, the fact in the Introduction (Section 14.1), the two history segments in Section 14.2, the try-its and the fact in Section 14.6, and the key points
Study	60 min	the whole chapter with its three pause points and the twelve exercises at its end
Practice	20 min	the first page, the results (Section 14.6), Technology in Practice (Section 14.7), Limitations (Section 14.8) and the open challenges

The control question A car that follows another can settle smoothly after the car ahead changes speed and still pass the change back larger than it came. When does a line of well-damped followers let a brake tap die out, and what do the spacing policy, a radio link, human reaction delay and a predictive controller change?

Learning objectives

- Derive the car-to-car transfer function of a time-gap follower and the bound $h \geq \sqrt{2/k_p}$, and explain why constant spacing fails for every gain.
- Show that the damping of one follower does not decide string stability, using the roots and the $|\Gamma|$ peak of the same loop.
- Measure the amplification along the line in the app and compare it with the criterion, for radar ACC, cooperative ACC and drivers.
- Price string stability in lane flow, radio delay and packet loss, and explain what the fallback and the cut-in supervisor buy.
- Judge when a linearized criterion understates the growth of a large wave.

Prerequisites Sections 1.4.10 (string stability), 1.4.4 (eigenvalues), 1.4.7 (delay limits), 1.6.1 and 1.6.2 (engine and velocity form). Section 1.6.3 is needed only for the predictive controller. Volume 2, Chapter 2 for the cruise loop of one car.

The app https://apmonitor.com/apps/vehicle-platoon/

See also Volume 20, Chapter 2 for the speed dependence of an automated car's path-following loop, and Volume 20, Chapter 9 for a speed chosen from stopping distance. Volume 5, Chapter 8 for a loop closed through a seeded delay line with jitter and packet loss.

14.1 Introduction

An automated car that follows another regulates a gap whose setpoint moves with the car ahead, and its loop is first judged, like any other, by whether it settles. A line of such cars raises a question that no single loop answers, which is how much of the acceleration of the car ahead each car passes on to the car behind it. If that gain exceeds one at any frequency, a disturbance with energy there grows from car to car however well damped each loop is, and the requirement that it never does is called string stability [1, 2]. Section 1.4.10 states the criterion on a line of carts. Production adaptive cruise control (ACC) has made the question practical. Seven 2018 model-year systems identified from car-following experiments were all string unstable, and along a platoon of one model an initial disturbance of 6 mph grew to 25 mph [3].

This simulator makes the difference between a well-tuned loop and a well-behaved line measurable. At a time gap of 0.6 s each radar follower behind a car at constant speed has a closed-loop damping ratio of 0.50, yet the gain from car to car peaks at 1.306 at 1.12 rad/s, and a brake tap that takes the lead from 72 to 57.3 km/h takes car 10 down to 36.3 km/h. Stiffening the loop does not rescue the line. A spacing gain of 3 instead of 2 leaves the peak at 1.305, because the time gap must exceed $\sqrt{2/k_p}$, and over the whole range of the gain slider that bound never falls below 0.82 s. Holding a fixed distance instead of a time gap fails for every gain, as the vehicle-string literature showed for predecessor following on relative spacing alone [4], and the best of 90 gain pairs here still peaks at 1.050. A radio that delivers the predecessor's acceleration restores stability down to a 0.41 s gap at 50 ms of delay, but needs 0.74 s at 400 ms with one packet in five lost. None of this is visible in the poles of one follower, because the property belongs to the chain.

DID YOU KNOW? Production Time Gaps Span This Chapter's Cases

Models fitted to car-following data from seven production adaptive cruise controls gave effective time gaps from 0.544 s to 2.054 s across their shortest and longest settings, with lags, perception included, from 0.153 s to 0.994 s [5]. The 0.6 s and 1.5 s gaps of this chapter lie inside that range.

The structure is general. String stability was posed for countably infinite interconnections of nonlinear systems, with weak-coupling conditions that guarantee it [2], and $\mathscr{L}_p$ definitions extend it to nonlinear cascades [6]. Outside vehicles, the orders that each company in a supply chain places with its upstream neighbor vary more than its sales, and the distortion grows moving upstream, the bullwhip effect [7]. Volume 6, Chapter 11 measures that amplification on a dairy supply chain. A ring of such stages closes the chain on itself, and on a circuit of 22 cars a uniform flow breaks into a stop-and-go wave with no bottleneck at all [8]. In each case the stage-to-stage gain at every frequency, not the stability of any one stage, decides what the chain does.

However, the experimental record measures amplification on instrumented cars [3, 6, 9] and on test tracks [8, 10], so a student seldom has a plant on which the frequency-domain criterion and the measured response of the chain can be compared car by car while the policy, the gains, the radio or the drivers change one at a time. This chapter pairs the string-stability criterion, computed live from the linearization of the very law each car executes, with the amplification measured along the nonlinear chain of a browser platoon simulator. The spacing policies, the cooperative feedforward, the driver model and the criterion are established methods from the literature. In the simulator each controller sees the line only through a delayed radar, a lossy radio, its own wheel speed and accelerometer, or a driver's delayed eye, and every equation, constant, delay and limit appears below. Its ten cases and a challenge isolate the policy, the gains, the radio, a cut-in, a ring and a predictive controller in every follower.

Section 14.2 traces the problem from the first vehicle strings to cooperative platoons. Section 14.3 describes the line, its sensors, its radio and what is measured. Section 14.4 derives the criterion from the cars' own law, for radar ACC with a time gap, constant spacing, cooperation over a delayed radio and human drivers. Section 14.5 gives the lead's cruise control, the followers' laws, the supervisor and the distributed predictive controller. Section 14.6 reports the measured runs of the ten cases and the challenge, Section 14.7 sets the radar, the radio and the throttle and brakes beside their real counterparts, Section 14.8 states the limitations and Section 14.9 concludes.

14.2 From Vehicle Strings to Cooperative Platoons

Automatic following began as a problem of densely packed strings of high-speed vehicles, and the first history segment tells how string stability was named.

HISTORY Strings of Vehicles and Automated Highways, 1964 to 1980

Automatic following began as a problem of densely packed strings of high-speed vehicles. In 1966 the theory of optimal control gave a linear feedback that regulates the position and velocity of every vehicle in such a string, with analog computer simulations of a string of three [11]. A long-range program at The Ohio State University studied headway safety policy, longitudinal and lateral control and highway system operations from 1964 to 1980, and a 6.4 km automated vehicle control facility was almost complete when the work was terminated [12]. By 1974 string stability was named as an important performance criterion for longitudinal control. A vehicle that follows its own moving reference cell always has it, because no vehicle then reacts to another. Without such a reference, relative-motion systems could reach it only with both forward and rearward separation measurements [1].

The California PATH program was founded in 1986 as the first research program in North America on what is now called intelligent transportation systems [13]. In the decade that followed, a comparison of the two longitudinal policies gave a systematic design of constant-spacing controllers

that guarantee string stability and found that a constant headway, or time gap, assures it without communication between the vehicles, which suits it to traffic that mixes automated and manual vehicles and to autonomous intelligent cruise control [14]. String stability was then given a formal definition for countably infinite interconnections [2]. A team that included engineers of California PATH ran automated vehicles in platoons under an integrated longitudinal and lateral control system, string stable at very small inter-vehicle spacing from a stop to high-speed cruising and through split and join maneuvers under communication constraints, with experimental results from the NAHSC demonstration of automated highways in San Diego [15]. The limitation of constant spacing on relative information alone was later traced to a complementary sensitivity integral constraint [4].

Adaptive cruise control reached the market with the time gap as a setting the driver chooses. A microscopic simulation built on the time gaps that drivers from the general public used in a field experiment found that ACC was unlikely to change lane capacity by much, while cooperative ACC raised it greatly once its market penetration reached moderate to high percentages [16]. Whether production systems damp traffic waves was later measured directly on seven 2018 model-year vehicles widely available in the United States. Models identified from more than 1200 miles of car-following data showed all seven string unstable, and in a platoon of one model that offers ACC on all trim levels the growth of the disturbance ended with the last car's ACC disengaging [3].

Cooperative adaptive cruise control (CACC) adds the predecessor's acceleration received over a wireless link. A decentralized design with a link to the nearest preceding vehicle only was given a necessary and sufficient frequency-domain condition for string stability, which showed that the link permits small distances under a velocity-dependent spacing policy while a constant spacing cannot be guaranteed string stable, and it was validated on two equipped cars [9]. An $\mathscr{L}_p$ definition for nonlinear cascades was validated on a platoon of six passenger vehicles with CACC [6]. The link is also the weakness. Latency and packet loss reach the controller, and under persistent loss CACC degrades to ACC and needs a larger distance to stay string stable, which motivated a graceful degradation that estimates the predecessor's acceleration with onboard sensors [17].

The ring road showed that a line needs no leader to amplify, as the second history segment tells.

> ⧖ HISTORY **A Jam With No Bottleneck, 2008**
>
> An experiment reported in 2008 put 22 cars on a circuit of 230 m. The drivers were asked to cruise at about 30 km/h and to follow the car ahead safely. Small fluctuations in the gaps grew until several cars stopped for a moment, and a jam formed with no bottleneck at all. Cars inside the jam stood still while those outside moved at about 40 km/h, and the jam traveled backward at roughly 20 km/h, the common speed of a jam on real highways [8]. A later field experiment on a 260 m track with 21 or 22 vehicles found that such waves emerged consistently and that controlling the velocity of a single vehicle in the flow dampened them [10].

What the application adds to these experiments is a chain on which the frequency-domain criterion and the measurement that the field studies made can be read together, one car at a time.

14.3 The Line, Its Sensors and Its Radio

The simulated system is one straight lane of identical cars, drawn in Figure 14.1. The cars are numbered from the front, car 1 being the lead, and a line holds $N = 10$ by default and up to 24. Two further slots hold a car that cuts in from the next lane and a slower car that may pull in ahead of the lead, so the state vector has 26 slots of three states each, 78 in all, and a slot that is not on the road has zero derivatives. On the ring there is no lead, and car 1 follows car N a lap ahead. Table 14.1 classifies the signals as controlled (CV), manipulated (MV) and disturbance (DV) variables, states, measurements and outputs, with SP for a setpoint.

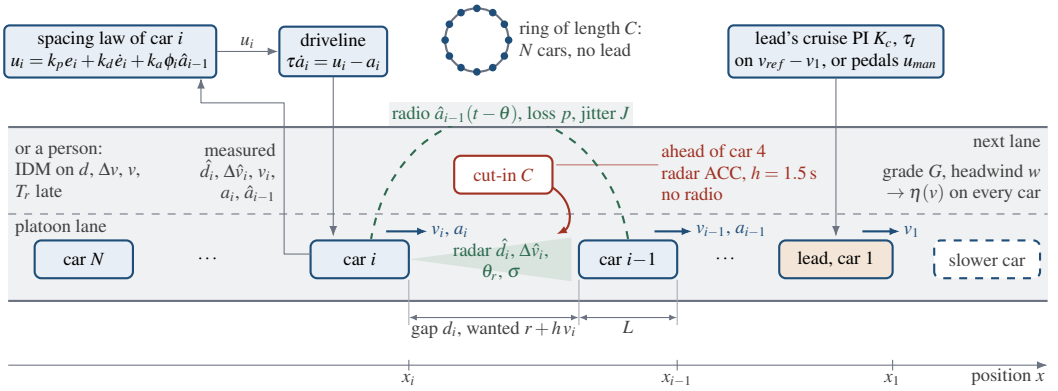


Figure 14.1 The line of cars, its sensors and its radio, not to scale. Car i has its front bumper at x_i, length L, speed v_i and delivered acceleration a_i, which follows the command u_i through the driveline lag τ. Its radar measures the gap d_i to the car ahead and the gap's rate of change as $\hat{d}_i$ and $\Delta\hat{v}_i$, its own speed and acceleration are measured exactly, and the radio delivers the acceleration of car $i-1$ after a delay θ with loss p and jitter J. The spacing law wants a gap of $r + h v_i$. The lead runs a cruise PI on its speed reference v_{ref}, which follows the set speed v_{set} through the chosen profile, or takes the operator's pedals, a car may cut in ahead of car 4, a slower car may pull in ahead of the lead, grade G and headwind w act on every car, and on a ring of length C the N cars follow one another with no lead.

Table 14.1 Signals of the Platoon by Role, With Their Defaults

Symbol	Role	Default	Range
v_1	CV of the lead's cruise loop	SP 72 km/h	10–130
d_i	CV of follower i	SP $r + h v_i$	≥ 0
u_1	MV, the lead's command	held 50 ms	-6 to 2 m/s^2
u_i	MV, a follower's command	held 50 ms	-6 to 2 m/s^2
u_{man}	MV in MANUAL, the lead's pedals	0	-6 to 2 m/s^2
x_i, v_i, a_i	states of 26 slots	laid out	—
G, w	DVs, grade and headwind	0, 0	± 6 %, ± 15 m/s
—	DVs, tap, hard stop, cut-in, radio burst, slower car	per case	events
$\hat{d}_i$, $\Delta\hat{v}_i$	measured, radar	50 ms late	σ 0.05
v_i, a_i	measured, wheel speed and accelerometer	exact	—
$\hat{a}_{i-1}$	measured, radio	50 ms late	p, J
γ, $\lvert\Gamma\rvert_{max}$	outputs, measured and predicted amplification	—	—
q, d_{min}	outputs, lane flow, smallest gap	—	—

14.3.1 The Cars

Every car is a point mass whose delivered acceleration follows its command with a first-order lag, as shown in Equation (14.1).

$$\dot{x}_i = v_i, \qquad \dot{v}_i = a_i + \eta(v_i), \qquad \tau \dot{a}_i = u_i - a_i \tag{14.1}$$

where u_i is the command that the car's controller or driver holds over each 50 ms control period, already clipped to $[a_{min}, a_{max}] = [-6, 2]$ m/s^2, $\tau = 0.4$ s is one lag for throttle and brake alike, and η is the disturbance on $\dot{v}$ that grade and headwind add, Equation (14.2). A car whose speed has reached zero stays stopped while $a_i + \eta < 0$, because brakes hold and nothing reverses.

$$\eta(v) = -\frac{g\,G}{100} - k_w\left[(v+w)\lvert v+w\rvert - v\lvert v\rvert\right] \tag{14.2}$$

where $g = 9.81\,\mathrm{m/s^2}$, G is the grade in percent with uphill positive, w is the headwind and $k_w = 2.6 \times 10^{-4}\,\mathrm{m^{-1}}$ is the drag coefficient $\frac{1}{2}\rho C_d A/m$ that turns the headwind into a deceleration. The bracket subtracts the drag of still air, so on a level road in still air no road load acts and a coasting car keeps its speed. The gap to the car ahead is Equation (14.3).

$$d_i = x_{p(i)} - x_i - L + C_i \tag{14.3}$$

where $p(i)$ is the car ahead of car i, $L = 4.5\,\mathrm{m}$, and C_i equals the ring length C for car 1 of a ring, whose car ahead is car N a lap further on, and zero otherwise. Nothing overtakes, so the order never changes. A gap that reaches zero is a collision. The two cars share their momentum as equal masses in a perfectly inelastic impact, are placed touching, brake at a_{min} to a standstill and stay out of the run, which pauses at the first one. The speed floor, the collision test, the cut-in car's lane change, the slower car leaving and every measured quantity are handled between integration steps, so the derivative function itself contains no switch.

14.3.2 Spacing Policies and the Following Law

A radar follower wants a gap that is either fixed or grows with its own speed, as shown in Equation (14.4).

$$d_i^* = \begin{cases} r + h\,v_{set} & \text{constant spacing} \\ r + h\,v_i & \text{constant time gap} \end{cases}, \qquad e_i = d_i - d_i^* - \delta_i \tag{14.4}$$

where $r = 3\,\mathrm{m}$ is the standstill gap, h the time gap, v_{set} the lead's set speed and $\delta_i \le 0$ the supervisor's allowance after a cut-in (Section 14.5.3). Constant spacing holds the gap that the time gap would give at the set speed, 33 m at 72 km/h for $h = 1.5\,\mathrm{s}$, whatever the car's own speed. Radar ACC and cooperative ACC run Equation (14.5) on the error.

$$u_i = k_p e_i + k_d \dot{e}_i + k_a \phi_i \hat{a}_{i-1}, \qquad \dot{e}_i = \Delta\hat{v}_i - h\,a_i \tag{14.5}$$

where $\Delta\hat{v}_i$ is the radar's measure of the rate at which the gap opens, $v_{p(i)} - v_i$, a_i is the car's own accelerometer reading of $\dot{v}_i$, $k_p = 2\,\mathrm{s^{-2}}$ and $k_d = 1\,\mathrm{s^{-1}}$ are the time-gap gains, $k_a = 1$ is the radio feedforward gain of cooperative ACC and zero for radar ACC, and $\phi_i \in [0, 1]$ is the supervisor's fade of the radio term. For constant spacing the $h\,a_i$ term is absent, because the wanted gap does not move with the car's speed. Under a time gap it is the derivative of the speed-dependent part of the wanted gap, so the law damps itself with the car's own acceleration. Constant spacing runs its own gains, 0.3 and 1.0, because the time-gap gains leave that loop with a damping ratio of 0.054 (Section 14.4). An ACC also respects a set speed of its own, $1.2\,v_{set}$, by taking the smaller of u_i and $\frac{1}{2}(1.2\,v_{set} - v_i)$.

14.3.3 People

A follower may instead be a person. Each driver follows the intelligent driver model [18] on the gap, the approach rate and their own speed, all seen a reaction delay T_r late, as shown in Equation (14.6).

$$\begin{aligned} u_i(t) &= a_m \left[1 - \left(\frac{v}{v_0} \right)^4 - \left(\frac{s^*}{s} \right)^2 \right]_{t - T_r} \\ s^* &= s_0 + v\,T_g + \frac{v\,\Delta v}{2\sqrt{a_m b}} \end{aligned} \tag{14.6}$$

where $s = d_i$, $v = v_i$, $\Delta v = v_i - v_{p(i)}$ is the approach rate, $v_0 = 120\,\mathrm{km/h}$ is the desired speed, $T_g = 1.6\,\mathrm{s}$ the desired time gap, $a_m = 0.73\,\mathrm{m/s^2}$ the maximum acceleration, $b = 1.67\,\mathrm{m/s^2}$ the comfortable deceleration and $s_0 = 2\,\mathrm{m}$ the jam distance, values the application takes from that paper. Each driver records the true gap, approach rate and speed at every 50 ms tick and acts on the record $T_r = 0.70\,\mathrm{s}$ old, fourteen ticks. The equilibrium gap $(s_0 + v T_g)/\sqrt{1 - (v/v_0)^4}$ is 36.44 m at 72 km/h and 15.36 m at 30 km/h, so 22 such drivers hold 30 km/h on a ring of $22(L + 15.36) = 437.0\,\mathrm{m}$, and on the 230 m circuit of the ring experiments [8] they would settle at 8.90 km/h. The reaction delay is a calibrated teaching value, because at 0.8 s the model collides (Section 14.6).

14.3.4 Radar and Radio

The radar and the radio are the same sampled link seen two ways, the application's channel model, which now lives in the framework as `shared/src/framework/core/channel.js`. Every 50 ms the sender stamps a sample with the time it was taken, and the link delivers or drops it as shown in Equation (14.7).

$$\begin{aligned} t_{arr,k} &= t_k + \theta + J\nu_k, \quad \text{dropped if } \mu_k < p \text{ or } t_k < t_{burst} \\ \hat{y}(t) &= y(t_{k^*}), \quad k^* = \arg\max_k \{t_k : t_{arr,k} \le t\} \end{aligned} \tag{14.7}$$

where θ is the fixed delay, J the largest jitter, p the loss probability, t_{burst} the end of a burst that drops every packet, and ν_k and μ_k two uniform deviates that every packet draws from its link's own seeded generator whatever the settings, so a link at 10 % loss and the same link at 20 % lose the same packets first. The receiver keeps the newest packet that has arrived and discards one overtaken by a newer one. A sample is used at the first 50 ms tick after it arrives, so the radar's 50 ms latency is served exactly and a latency of 120 ms becomes 150 ms, which the application's own check confirms. The radar's range and range rate carry Gaussian noise of 0.05 m and 0.05 m/s from a third generator of each car. The radio carries the predecessor's measured acceleration with a default delay of 50 ms. A link counts as stopped after 0.5 s, ten periods, without an arrival, and a burst drops every packet of every link for 6 s. The car that cuts in and the slower car ahead of the lead have no radio. Volume 5, Chapter 8 meets the same seeded delay, jitter and loss on the link of a remote manipulator.

14.3.5 Lead, Events and the Ring

The lead's set speed is 72 km/h unless a case names another. A profile moves its reference, holding steady cruise, running a stop and go that holds the set speed for 20 s, slows at 1.5 m/s^2 to a third of it, holds 20 s and returns at 0.8 m/s^2, repeating every 100 s, or adding a sine of ± 1 m/s at a chosen period. A brake tap overrides the cruise control with 3 m/s^2 of braking for 1.5 s. Scripted and learner taps hold 30 control periods, 1.5 s, with endpoint comparisons tolerant of floating-point roundoff. The commanded speed decrement is 4.5 m/s. A hard stop brakes at 5 m/s^2 to a standstill and holds 4 s. A cut-in places a car with no radio in the middle of the gap ahead of car 4, at car 4's speed and only if that gap exceeds a car length plus 1 m and holds 0.5 m in a stop at the line's present deceleration answered 0.45 s later at half the braking limit, running radar ACC at $h = 1.5$ s with $k_p = 1\,\mathrm{s}^{-2}$ and $k_d = 1\,\mathrm{s}^{-1}$. A slower car pulls in 3 m plus 1 s of travel ahead of the lead, holds 70 % of the lead's speed v_S with the command $\frac{1}{2}(v_S - v)$, and leaves the lane after 30 s. On the ring, $N = 22$ drivers start evenly spaced on $C = 437$ m at the speed their policy settles at for that spacing, each moved by at most 10 cm by a seeded uniform draw.

14.3.6 What Is Measured

The flow is computed from the platoon's own spacing, as shown in Equation (14.8).

$$q = 3600\,\frac{\bar{v}}{\bar{\ell}}, \qquad q_h = \frac{3600}{h + (L + r)/v} \tag{14.8}$$

where q is in vehicles per hour, $\bar{v}$ is the mean speed of the cars on the road, $\bar{\ell}$ their mean front-to-front spacing, and q_h the flow of a time-gap policy at speed v. At 72 km/h a 1.5 s gap gives 1920 veh/h and a 0.6 s gap 3692 veh/h. The lane-flow tile of the scorecard averages q over the last 30 s, sampled every 0.5 s. The amplification is measured as a ratio of root-mean-square accelerations, Equation (14.9).

$$\gamma_i = \sqrt{\frac{\int a_i^2 \,\mathrm{d}t}{\int a_{p(i)}^2 \,\mathrm{d}t}}, \qquad \gamma = \max_i \gamma_i \tag{14.9}$$

where both integrals run over the same window and a here is the kinematic acceleration $\dot{v}$. A pair counts once the car ahead has passed back at least 0.5 m^2/s^3, an RMS of 0.1 m/s^2 for 50 s, and only

Table 14.2 Model Constants and Where Each Is Set: V in the Show Model Dialog, A Among the Advanced Controls, H Hidden From the Student, C in the Source

Symbol	Meaning	Value	Class
Cars			
L	car length	4.5 m	V
a_{max}, a_{min}	acceleration and braking limits	2, –6 m/s^2	V
k_w	drag per unit mass, headwind only	2.6×10^{-4} m^{-1}	V
τ	driveline lag	0.4 s	A
—	cut-in car's speed relative to car 4	0	H
Sensors and radio			
θ_r, σ	radar latency, noise on range and rate	50 ms, 0.05	A
θ, p, J	radio delay, loss, jitter	50 ms, 0, 0	A
—	link timeout, burst length	0.5, 6 s	C
People, intelligent driver model			
v_0, T_g	desired speed and time gap	120 km/h, 1.6 s	A
a_m, b	maximum acceleration, comfortable braking	0.73, 1.67 m/s^2	A
s_0	jam distance	2 m	A
T_r	reaction delay	0.70 s	A
Line, ring and measurement			
N	cars in the line or on the ring	10, or 22	A
C	ring length	437 m	A
—	ring start jitter	±10 cm	C
—	energy threshold, quiet time	0.5 m^2/s^3, 15 s	C
Numerics			
Δt, T	integration step, control period	10, 50 ms	C
—	log rate, seed	10 Hz, 20261001	C

while the whole line has run its policy for 15 s, untouched by a cut-in's gap regrowth or a fallback's change of gap. A pair starts again when a car's predecessor changes, the pair behind the slower car is never counted, and on the ring neither is car 1's. The RMS ratio is the gain that the criterion of Section 14.4 bounds, an $\mathscr{L}_2$ notion of string stability [6], and Section 14.4.6 shows why a ratio of peaks is not used.

14.3.7 Constants, Start and Integration

Table 14.2 gathers the model's constants, and the settings a student can move appear in Table 14.3. Every case on the open road starts with the line at its policy's gap for the lead's set speed and every car at that speed with zero acceleration, an equilibrium of Equations (14.1) to (14.5) to within 4×10^{-16} by the application's check, and each sensor starts as though it had been tracking the car ahead, with samples already in flight. Classical fourth-order Runge-Kutta integrates the 78 states on a fixed 10 ms step, every controller and sensor runs at 20 Hz, and the historian records at 10 Hz, with controllers, scripted events and the historian on the engine's schedule of Section 1.6.1. The default seed is 20261001, and each car's sensors draw from generators of their own, so a case replays bit for bit and the losses of one link do not change when another car's settings do. Halving the step changes the measured amplification of the 0.6 s case and of the human case by a relative 1.3×10^{-6}. The application passes its 53 verification checks.

14.3.8 Assumptions

One lane, longitudinal motion only. A cut-in joins the line at once in the middle of the gap, and its lane change is only drawn. Grade and headwind act on every car at once, as a long uniform slope and a steady wind would.

Identical cars with one lag. Every car has the same 0.4 s lag for throttle and brake and the same hard limits. A real driveline brakes faster than it accelerates, and its lag changes with speed, gear and load.

No road load in still air. The drag of still air and rolling resistance are left out, so the lead's cruise control has no steady load to hold and a coasting car keeps its speed. Aerodynamic drag and rolling resistance would add a steady load that grows with speed on every car.

People who look one car ahead. The drivers are identical, noiseless and a pure delay late, and they watch only the car ahead. Real drivers anticipate and watch several cars ahead, and without that this model collides from 0.8 s of delay.

An ideal radar and an independent radio. The radar sees exactly the car ahead at any range. The radio's losses are independent, the clocks of the cars agree, and loss does not grow with the number of cars talking, as it would on a congested channel.

A criterion from one linearization. The criterion is computed at the line's mean follower speed and describes small waves. For the radar and radio laws, which are linear, it holds until a limit is reached, while for people a large wave grows faster than the linear peak suggests.

Teaching values. A collision is perfectly inelastic, and nothing about injury or damage is modeled. Every threshold, the challenge's among them, is a teaching value, and the application names ISO 15622 for adaptive cruise control and ISO 20035 for cooperative adaptive cruise control as the sources of real requirements.

14.4 The Criterion From the Cars' Own Law

The application never writes down a model of a follower. It linearizes the function that the cars execute, so the criterion and the chain cannot disagree about what the law is.

14.4.1 Linearization

A follower's command is one function of five measured quantities. Central differences of that function about the operating point give five gains, g_s on the gap, g_{dv} on its opening rate, g_v on the car's own speed, g_{ai} on its own acceleration and g_a on the predecessor's broadcast acceleration. With $A_i(s)$ the transform of a_i, the gap and the opening rate written through the two accelerations as $(A_{i-1}-A_i)/s^2$ and $(A_{i-1}-A_i)/s$, each input arriving late by its own delay and the command held over the control period, the ratio $\Gamma(s)=A_i(s)/A_{i-1}(s)$ is Equation (14.10).

$$\begin{aligned}\Gamma(s) &= \frac{N(s)}{D(s)}, \qquad H(s)=\frac{1-e^{-sT}}{sT}\\ N(s) &= H\left[e^{-\theta_d s}(g_s+g_{dv}s)+e^{-\theta_c s}g_a s^2\right]\\ D(s) &= \tau s^3+s^2+H\left[e^{-\theta_d s}(g_s+g_{dv}s)\right.\\ &\qquad \left.-e^{-\theta_v s}g_v s-g_{ai}s^2\right]\end{aligned} \tag{14.10}$$

where $T=50\,\text{ms}$, θ_d is the radar's latency rounded up to whole ticks or the driver's delay, θ_v the driver's delay on their own speed and zero for an ACC, and θ_c the radio's effective delay. A lost packet leaves the last good one in use for a geometric number of extra periods, with mean $p/(1-p)$, so θ_c is the mean of $T\lceil(\theta+Jv)/T\rceil$ over the jitter plus $Tp/(1-p)$, and 20 % loss adds 12.5 ms. For radar ACC the gains are the law's own, $g_s=k_p$, $g_{dv}=k_d$, $g_v=-k_p h$, $g_{ai}=-k_d h$ and $g_a=k_a$, and the units make Γ dimensionless. The line is string stable when $|\Gamma(j\omega)|\le 1$ at every frequency [2, 9], as in Section 1.4.10. Because a lasting change of speed is passed on whole, $\Gamma(0)=1$ for every policy, so a stable line has its largest $|\Gamma|$ at zero frequency and an unstable one has a peak above one somewhere. The application finds the peak on 161 logarithmically spaced frequencies from 0.003 to 30 rad/s, refined by 30 golden-section steps, and counts the line stable when the peak is at most $1+10^{-6}$.

The application's own check drives the lead with a sine of 0.2 m/s and requires the ratio passed from car 2 to car 3 to match Equation (14.10) within 1 % for three configurations. Repeated at the peak frequency of five unstable configurations and at 1 rad/s for two stable ones, the two agree within

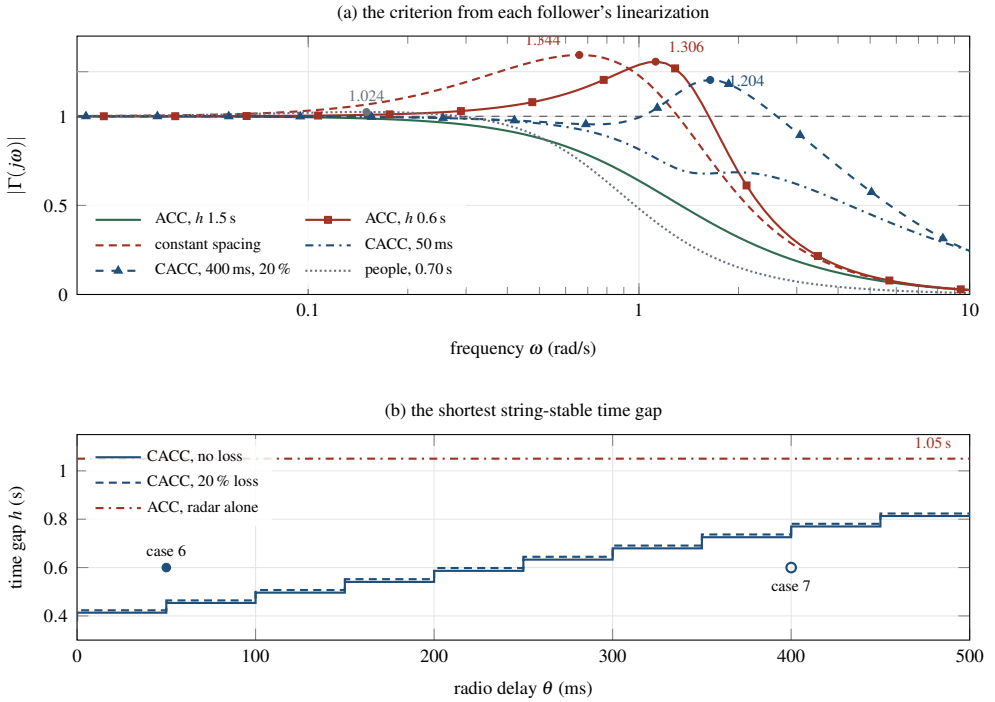


Figure 14.2 The string-stability criterion at 72 km/h, from the application's own linearization. (a) $|\Gamma(j\omega)|$ for radar ACC at time gaps of 1.5 and 0.6 s, constant spacing with its gains of 0.3 and 1.0, cooperative ACC at 0.6 s with a 50 ms radio and with 400 ms and 20 % loss, and drivers 0.70 s late, with the unstable peaks marked. Radar ACC at 1.5 s is solid, at 0.6 s solid with squares, constant spacing densely dashed, cooperative ACC with the fast radio dash-dotted and with the slow one dashed with triangles, and the drivers dotted. (b) The shortest time gap at which cooperative ACC with $k_a = 1$ is string stable, rising in steps of one 50 ms tick of radio delay, solid without loss and dashed at 20 % loss, against the 1.05 s that radar ACC needs, dash-dotted. Case 6, the filled circle, lies in the stable region above the steps and case 7, the open circle, below them.

0.19 %, and within 0.27 % for a stable line at 30 % packet loss. Figure 14.2(a) draws the criterion for six followers at 72 km/h.

14.4.2 Radar ACC With a Time Gap

With no delays and no hold, Equation (14.10) reduces for radar ACC to the transfer function of Equation (14.11).

$$\Gamma(s) = \frac{k_d s + k_p}{\tau s^3 + (1 + k_d h)s^2 + (k_d + k_p h)s + k_p} \tag{14.11}$$

The application's check confirms that its linearization reproduces it within a relative 4.1×10^{-11}. With no lag, $\tau = 0$, it is Equation (1.11) with h in place of t_g. With N and D now the numerator and denominator of Equation (14.11), string stability asks $|D|^2 - |N|^2 \geq 0$ at every frequency, and the difference is the polynomial of Equation (14.12).

$$\begin{aligned} |D|^2 - |N|^2 &= c_2\omega^2 + c_4\omega^4 + \tau^2\omega^6 \\ c_2 = k_p(k_p h^2 - 2), &\qquad c_4 = (1 + k_d h)^2 - 2\tau(k_d + k_p h) \end{aligned} \tag{14.12}$$

Near zero frequency $|\Gamma|^2 \approx 1 + \omega^2(2/k_p - h^2)$, so the line is stable only for $h \geq \sqrt{2/k_p}$, 1.00 s at $k_p = 2\,\mathrm{s}^{-2}$. The polynomial is non-negative at every frequency exactly when $c_2 \geq 0$ and either $c_4 \geq 0$

or $c_4^2 \le 4\tau^2 c_2$, and at the default gains this full condition gives the same 1.0000 s, because $c_4 = 1.6$ there. The radar's 50 ms and the hold move the boundary to 1.050 s. At k_p of 1 and 3 the bound is 1.414 s and 0.816 s, and 1.463 s and 0.868 s with the delays. The time gap helps twice. The wanted gap grows with the car's own speed, which supplies the $k_p h$ term of the s coefficient, and the rate of the error carries the car's own acceleration, which supplies the $k_d h$ term of the s^2 coefficient.

The denominator of Equation (14.11) is also the characteristic polynomial of one follower behind a car at constant speed, and it says nothing alarming. At $h = 1.5$ s its complex pair has a damping ratio of 0.968, and at 0.6 s the roots are $-0.684 \pm 1.197\,j$ and -2.632 s^{-1}, a damping ratio of 0.496 at a natural frequency of 1.378 rad/s. The same loop has its $|\Gamma|$ peak of 1.306 at 1.12 rad/s, near that natural frequency, and stays above one from zero up to 1.63 rad/s. The peak is 1.0035 at 1.0 s, 1.084 at 0.8 s and 1.306 at 0.6 s, as shown in Figure 14.2(a). Raising k_p to 3 moves the peak to 1.45 rad/s and leaves it at 1.305 while the damping ratio falls to 0.456, and raising k_d to 2 lowers the peak to 1.089 while the damping ratio rises to 0.773.

Pause. Try Exercise 14.1 before reading on.

14.4.3 Constant Spacing

Without the $h\,v_i$ term the h in the coefficients disappears, and the transfer function becomes $(k_d s + k_p)/(\tau s^3 + s^2 + k_d s + k_p)$, with $c_2 = -2k_p$. Near zero frequency $|\Gamma|^2 \approx 1 + 2\omega^2/k_p$, above one for every positive gain, the limitation that the literature traces to a complementary sensitivity integral [4]. Two integrators lie between a car's acceleration and its gap, and a fixed gap asks the car to match every speed change of the car ahead at once. The shipped gains of 0.3 and 1.0 give each car a damping ratio of 0.816 and the line a peak of 1.344 at 0.66 rad/s. The time-gap gains of 2 and 1 leave the loop damped at only 0.054, with a peak of 131.7. Over 90 gain pairs, k_p from 0.05 to 3 and k_d from 0 to 3, the smallest peak is 1.050 at $k_p = 0.05$ and $k_d = 1$, the lower end of the spacing-gain slider, a loop with a real pole at –0.053 s^{-1}, so slow that its gap drifts.

Pause. Try Exercise 14.2 before reading on.

14.4.4 Cooperation and Delay

The radio adds $-k_a\omega^2$ to the real part of the numerator and changes the low-frequency coefficient to $c_2 = k_p\big(k_p h^2 - 2(1-k_a)\big)$, so near zero frequency $|\Gamma|^2 \approx 1 + \omega^2\big(2(1-k_a)/k_p - h^2\big)$. At $k_a = 1$ the bracket is negative at any gap, and the bound moves to higher frequencies, where the lag and the delays decide it. With no delays at all the shortest stable gap is 0.342 s, the radar's latency and the hold raise it to 0.376 s with an instant radio, and the default 50 ms radio to 0.413 s. Each further 100 ms of radio delay costs about 0.09 s of gap, to 0.541 s at 200 ms, 0.725 s at 400 ms and 0.813 s at 500 ms, and 20 % loss adds 0.009 to 0.012 s more, 0.737 s at 400 ms. Every delay is served at a 50 ms tick, so the boundary rises in steps of one tick, as shown in Figure 14.2(b). The feedforward gain has an optimum near one. At 50 ms the boundary is 0.758 s at $k_a = 0.5$, 0.505 s at 0.8, 0.413 s at 1.0 and 0.464 s at 1.2.

14.4.5 Drivers Seen Through the Linearization

For a driver the gains are the slopes of Equation (14.6) at its equilibrium, all delayed by T_r. At 72 km/h they are $g_s = 0.0349\,\text{s}^{-2}$, $g_{dv} = 0.339\,\text{s}^{-1}$ and $g_v = -0.0787\,\text{s}^{-1}$, and the peak is 1.024 at 0.15 rad/s, a period of 42 s. A driver 0.70 s late is thus barely string unstable, and only for slow waves, while reaction delays of 0.5, 0.6, 0.75, 0.8 and 1.0 s give peaks of 1.017, 1.020, 1.027, 1.031 and 1.084. At 30 km/h, the speed of the ring, the slopes grow to $g_s = 0.0947\,\text{s}^{-2}$ and $g_v = -0.153\,\text{s}^{-1}$ and the peak to 1.198 at 0.43 rad/s. These are small-wave numbers. The braking term of Equation (14.6) grows with the square of s^*/s, and s^* grows with the approach rate, so a driver who is late and close brakes far harder than the linearization predicts.

Table 14.3 Controller Settings, Defaults and Ranges

Symbol	Setting	Default	Range
Lead			
v_{set}	set speed	72 km/h	10–130
K_c, τ_I	cruise PI gain and integral time	0.8 s^{-1}, 10 s	0.1–3, 1–60
—	speed profile	cruise	3 profiles
P	sine period	6 s	3–60
Followers			
—	policy	ACC, time gap	6 policies
h	time gap	1.5 s	0.3–2.5
k_p, k_d	spacing and rate gains	2, 1	0.05–3, 0–3
r	standstill gap	3 m	1–10
Radio and supervisor			
k_a	radio feedforward gain	1	0–1.5
h_{fb}	fallback time gap	1.5 s	0.6–3
ρ	gap regrowth after a cut-in	1 m/s	0–3
—	radio fade, widening, narrowing	0.5, 3, 6 s	fixed
—	MPC preview	plan	3 settings

Units of k_p and k_d are s^{-2} and s^{-1}. The six policies are people, ACC with constant spacing, ACC with a time gap, CACC, CACC with fallback and MPC (Section 14.5.4).

14.4.6 Which Ratio the Criterion Bounds

Equation (14.10) bounds the ratio of RMS values. By Parseval's theorem the energy of a_i is the integral of $|\Gamma|^2$ times the energy spectrum of $a_{p(i)}$, at most $|\Gamma|^2_{max}$ times the energy of $a_{p(i)}$. The ratio of peak accelerations is bounded instead by the integral of the absolute impulse response, which exceeds $|\Gamma(0)| = 1$ whenever the response changes sign, so a string-stable line can pass back larger peaks. With the lead of case 4 switched to the stop and go profile from 72 km/h, its scripted tap included, ACC at 1.2 s passes back peaks 1.0130 times larger at an RMS ratio of 0.9904 over 100 s, and cooperative ACC at 0.6 s 1.0164 at 0.9902. The tap's own peak hides the effect at 1.5 s (0.9869 at 0.9845), but a model run without the tap gives 1.0096 at 0.9867 there. The measured amplification of Equation (14.9) is therefore the $\mathscr{L}_2$ quantity, the one the criterion can be checked against [6].

Pause. Try Exercise 14.3 before reading on.

14.5 Cruise Control, Following Laws and the Supervisor

Every controller runs every 50 ms in a fixed order. Each car with a radio broadcasts its acceleration to the car behind it, each radar takes its sample, everything that has arrived is delivered, the lead's cruise control runs, every follower runs its law under the platoon supervisor, and the criterion of Equation (14.10) is recomputed when a setting it depends on changes, every 5 s, or when the speed of a line of people drifts by 0.25 m/s. Each command is held until the next tick, and no controller reads another car's state except through its sensors. The settings are listed in Table 14.3.

14.5.1 The Lead's Cruise Control

The lead's cruise control is the framework's velocity-form PID of Section 1.6.2, here a proportional-integral loop on the speed reference whose output is the acceleration command, as shown in Equation (14.13).

$$\begin{aligned} e_k &= v_{ref,k} - v_{1,k}, \qquad S_k = S_{k-1} + K_c e_k - P_{k-1} \\ \tilde{u}_k &= S_k + I_{k-1}, \qquad u_{1,k} = \min\left(a_{max}, \max(a_{min}, \tilde{u}_k)\right) \\ I_k &= I_{k-1} + \left[\frac{K_c}{\tau_I} e_k + \frac{u_{1,k} - \tilde{u}_k}{\tau_I}\right] T \end{aligned} \tag{14.13}$$

where $P_{k-1} = K_c e_{k-1}$ is the proportional term formed at the previous scan, $K_c = 0.8\,s^{-1}$, in m/s^2 per m/s, $\tau_I = 10$ s, the tracking time of the back-calculation equals τ_I, T is the 50 ms control period,

and there is no derivative action and no rate limit. These are Equations (1.13) to (1.15) without the derivative term, and Volume 2, Chapter 2 tunes the same loop on a single car. The unclamped running sum S_k telescopes to $K_c e_k$, the clip acts on the emitted output only, and the anti-windup acts on the integral only, so the law is the positional PI with back-calculation on its integral, at the limits as well. Fed the same errors through a run whose set-speed steps hold the command at a limit for 153 of 1800 scans, the two give identical commands. A tap or a hard stop replaces the command while it lasts. When it ends, the PI restarts from zero output and the reference restarts at the speed the car is doing, climbing back to the profile at 1 m/s^2, as a production cruise control resumes from coasting. In MANUAL the operator's pedal value u_{man} is the command and the integral is set each tick so that the PI's output equals it, so a switch of mode in either direction moves nothing. Behind a slower car the lead takes the smaller of the PI's command and that of a radar ACC at 1.5 s on the slower car.

14.5.2 The Followers

Each follower evaluates Equation (14.5), or Equation (14.6) for a person, on what its sensors delivered at this tick, clips the result to the acceleration limits and holds it. A radar car whose first sample of a new target has not yet arrived keeps its last command, and a car that has collided brakes at a_{min}. A person's law reads the record kept 14 ticks earlier in a history of 64 ticks, 3.2 s. The rate of the error comes from the radar's range rate and the accelerometer, so no signal is differentiated and no derivative filter is needed. The followers carry no integral term, so nothing winds up against the acceleration limits and no anti-windup is needed. The steady gap error is nevertheless zero behind a car at constant speed, because two integrations lie between a car's command and its gap. Every policy is direct acting on the gap. A gap longer than wanted or a faster car ahead makes it accelerate, and a shorter gap or a slower car ahead makes it brake, which the application's direction check confirms for all four.

14.5.3 What the Supervisor Asks For

The supervisor shapes what each follower is asked for. When a new car appears ahead, the follower accepts the gap it was left and grows it back at a limited rate, as shown in Equation (14.14).

$$\delta_i(t_0) = \min\left(0,\ \hat{d}_i - d_i^*\right), \qquad \dot{\delta}_i = \rho \ \text{ while } \delta_i < 0 \tag{14.14}$$

where t_0 is the tick at which the radar's first sample of the new car arrives and $\rho = 1\,\mathrm{m/s}$, so the car opens its gap gradually instead of braking for the whole shortfall at once, and the car that cut in does the same toward the car ahead of it. A cooperative follower watches the silence since its newest packet. Past 0.5 s the link counts as stopped, ϕ_i fades from one to zero over 0.5 s, and in the degraded mode the time gap in force moves toward the fallback gap $h_{fb} = 1.5\,\mathrm{s}$, as shown in Equation (14.15).

$$\dot{h}_{eff} = \begin{cases} \min\left(0.1,\ (h_{fb} - h_{eff})/3\,\mathrm{s}\right) & \text{link stopped} \\ \max\left(-0.05,\ (h - h_{eff})/6\,\mathrm{s}\right) & \text{packets back} \end{cases} \tag{14.15}$$

where the rates are seconds of time gap per second and h_{eff} replaces h in Equations (14.4) and (14.5). When packets return, ϕ_i fades back to one over 0.5 s and the gap narrows again. Both moves are first-order approaches with a rate limit, so neither ends in a corner that would leave the cars at the back braking out of a closing speed. A predecessor with no radio at all, such as the car that cut in, is treated like a stopped link, so the car behind it falls back to h_{fb}.

14.5.4 Distributed Model Predictive Control

A sixth policy, model predictive control (MPC), never the default, replaces Equation (14.5) in every follower by a receding-horizon optimization [19] (Section 1.6.3), the distributed arrangement in which each car plans its own command and passes the plan to the car behind [20, 21]. At every control period $T = 50\,\mathrm{ms}$ each follower plans its command over the next $N_p = 60$ steps, 3 s or more than seven driveline lags, held in eight blocks of $b_j = 1, 1, 2, 3, 5, 8, 15$ and 25 steps, applies the first

move, holds it for one period and plans again. Its model is the car of Equation (14.1) written in the frame of the gap, Equation (14.16).

$$\begin{aligned} &\dot{d} = \Delta v, \quad \dot{\Delta v} = a_p - (a+\beta), \quad \dot{v} = a+\beta, \\ &\bar{\tau}\dot{a} = u - a, \quad \dot{\beta} = 0 \end{aligned} \tag{14.16}$$

where Δv is the opening rate, a_p the acceleration of the car ahead, which enters as a preview over the horizon, β an input disturbance on $\dot{v}$ that stands for grade, headwind and any lag other than the declared one, and $\bar{\tau} = 0.4\,\mathrm{s}$ the declared lag, fixed when the application loads, so that a lag set on the plant is model mismatch. The model is written from the plant's equations, not identified, and sampled exactly by a zero-order hold over T, which a fourth-order Runge-Kutta integration reproduces to 4×10^{-16}. Each plan solves Equation (14.17).

$$\begin{aligned} \min_{u^b, s} \; & \sum_{k=1}^{N_p} \left(Q_e e_k^2 + Q_v \Delta v_k^2 \right) + P_e e_{N_p}^2 + P_v \Delta v_{N_p}^2 \\ & + S \sum_{j=0}^{7} \left(u_j^b - u_{j-1}^b \right)^2 + R \sum_{k=0}^{N_p - 1} \left(u_k + \hat{\beta} \right)^2 + \rho s^2 + \sigma s \\ \text{s.t.} \; & e_k = d_k - (r + h_{eff} v_k + \delta_i), \\ & d_k - \tfrac{1}{2} h_{eff} v_k \geq \tfrac{1}{2} r + \delta_i - s, \quad s \geq 0, \\ & a_{min} \leq u_j^b \leq a_{max}, \quad \left| u_j^b - u_{j-1}^b \right| \leq j_{max} T \, c_j \end{aligned} \tag{14.17}$$

where u_j^b is the command of block j, u_k the command in force at step k, u_{-1}^b the move in force at the decision, e_k the gap error of the time-gap policy with the supervisor's h_{eff} and δ_i, and s one slack for the whole horizon. The weights are $Q_e = Q_v = 1$, $P_e = P_v = 10$, $S = 0.5$ and $R = 0.2$ in SI units, with $\rho = 10^5$ and $\sigma = 10^4$ on the slack. The R term prices the acceleration $u + \hat{\beta}$ that the command would sustain, so holding speed against a grade costs nothing. The acceleration limits are hard and read as the clip of Section 14.5 reads them, the floor at half the time-gap gap is soft, and $j_{max} = 10\,\mathrm{m/s^3}$ is a comfort limit on the jerk, hard, with $c_0 = 1$ on the first move and $c_j = (b_{j-1} + b_j)/2$ between blocks, the change that a ramp at the limit makes between two block centers. The set-speed cap and the braking of a collided car stay in force above the plan.

The ramp form corrects the first build, which applied one limit of $0.5\,\mathrm{m/s^2}$ at every block boundary, as the shared solver does whatever a block's length, so a plan believed its command could change by at most $4\,\mathrm{m/s^2}$ in 3 s and braking hard committed it to braking for seconds. In a hard stop at 8 s on case 10's line car 2 then commanded only -3.63 to $-4.83\,\mathrm{m/s^2}$ from 8.8 to 10.4 s behind a lead at $-4.2\,\mathrm{m/s^2}$ to -5.0 (Section 14.6.6). With the ramp form it does as well as with no jerk limit, and the largest move on case 10's tap is $0.37\,\mathrm{m/s^2}$ per period.

A steady-state Kalman filter [22], its gain computed once, estimates the five states, augmented with the gap and its rate one period earlier because the radar's samples arrive one period late, from the radar's gap and rate, the wheel speed and the accelerometer, which reads $a + \beta$, with variances of 0.05^2, 0.05^2, 10^{-6} and 10^{-4} in SI units and a process noise covariance of $\mathrm{diag}(10^{-5}, 4 \times 10^{-3}, 10^{-6}, 10^{-3}, 10^{-4}, 10^{-9}, 10^{-9})$. The disturbance state makes the controller offset free. With a lag of 0.6 s, a 3 % grade and an 8 m/s headwind on the plant, car 2's $\hat{\beta}$ settles at $-0.3960\,\mathrm{m/s^2}$ against the plant's -0.3942, and the worst gap error is 3.3 cm against 3.2 cm at nominal conditions, the radar's noise.

With the plan preview each car broadcasts ten numbers, its eight block commands, its driveline state and $\hat{\beta}$, from which the car behind reconstructs the predicted acceleration, an extension of the radio protocol under this policy only. With the acceleration preview the one acceleration the radio carries today is held for 0.4 s and then taken as zero, the like-for-like comparison. With the preview off nothing from the radio is used, and the supervisor widens every gap to h_{fb} as for a stopped link.

The lead has no plan, so car 2 always previews one acceleration. The plan is a forecast, and its acceleration 1 s ahead is in error by 0.20 m/s^2 RMS on the tap of case 10, 0.60 m/s^2 under the 400 ms, 20 % radio of case 7 and 0.62 m/s^2 in the hard stop.

The controller uses a plan only when the solver returns the status optimal, the plan, predictions, cost and slack are finite, every planned command lies within the limits and the first planned command equals the move applied. Otherwise Equation (14.5) with $k_a = 1$ drives that car, with a bias from the last move that decays with a 1 s time constant, so the hand-over moves nothing, until a later plan is accepted. Under the plan k_p, k_d and k_a have no effect, while h, r, h_{fb} and ρ stay in force.

With no constraint active, the first move and the broadcast plan are affine in the estimate, the last command and the preview, with gains taken from the solver itself. Because a plan rather than one number passes from car to car, the criterion replaces Equation (14.10) by propagating each car's acceleration and plan along the line frequency by frequency, through the plant at the lag as set, the estimator and the delays, and reports the worst car. It matches simulated 6 s sines within 2 % by the application's check, and it holds while nothing binds, as on the tap of case 10, where the floor's slack stays at zero, but not in a hard stop, where the braking limit binds.

The weights were chosen once, on the 0.3 s line of case 6, as those with the least $|\Gamma|$ peak whose hard stop from cruise leaves at least 2 m. A rule on the tap alone would have picked $R = 3$ and $S = 2$, 0.943 on the tap of case 10, whose linearization peaks at 1.119 at 0.4 s and which keeps 0.11 m in the hard stop at 8 s. In Chrome 157 each period's nine solves take 2.2 % of a processor core at real time, 0.12 ms a solve, and 41 % at twenty times. At sixty times the page painted 14.5 frames a second, so a frame guard stops a frame's stepping after 25 ms, skipping no decision and leaving every logged row unchanged, and the page runs at 37.6 times real time and 33 frames a second.

14.5.5 Simulation, Control and Analysis Models

Equations (14.1) to (14.7) make up the simulation model, nonlinear through the clips, the set-speed cap, the stopped-car rule, the collisions and the driver model. Under the five shipped policies no controller holds a model of the line, since the spacing laws are model-free and the lead's PI knows nothing of the followers, while the predictive controller of Section 14.5.4 carries its own model of one car at the declared lag, Equation (14.16), and sees the plant only through the instruments of Equation (14.7). Equation (14.10), evaluated at the line's mean follower speed with the delays as the cars experience them, is an analysis model used only to display the criterion and its map, and no model is identified from data.

14.6 Eleven Runs Along the Line

Table 14.4 lists the cases. Every case begins from the defaults of Tables 14.2 and 14.3, ten cars at 72 km/h on radar ACC at 1.5 s, and departs from them as its row gives. The figures that follow were produced by the engine alone, with the default seed and a span of 90 s unless another is named. Each variant is the case with the setting changed at $t = 0$ before Run, and a time gap or policy chosen there lays the line out at the gap it wants. Reset keeps the student's settings and replays the case's scripted events, so the same setting carried through Reset gives a run identical to the last digit on the open road. In cases 2 to 7 the brake tap is the first disturbance the line meets. Until it comes, the drivers of case 2 hold their start exactly, and the radar cars of cases 3 to 7 stay within 0.16 km/h of it, moved only by the radar's noise. Table 14.5 collects the results, and Figure 14.3 sets the predicted peak beside the measured amplification and the lane flow for the main policies.

14.6.1 People and the Reaction Delay

In case 1 the operator holds the lead's pedals while the scenario taps the brake at 6 s, 3 m/s^2 for 1.5 s, which commands a 4.5 m/s speed decrement. In MANUAL nothing restores it, and with no road load the lead coasts at 55.8 km/h. Each driver sees the change 0.70 s late and answers harder than the one ahead, so the dip deepens down the line. Car 2 bottoms out at 53.3 km/h, car 5 at 42.8, car 8 at 14.3 and car 9 at 0.9. Car 10 reaches zero at 27.14 s, 21.14 s after the tap. The cars pass back up to 1.29

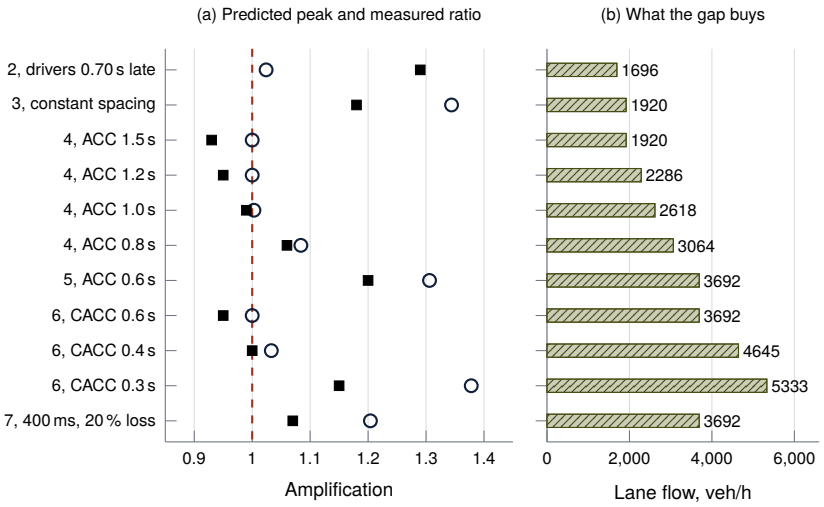


Figure 14.3 Rows of Table 14.5. (a) The criterion's peak $|\Gamma|_{max}$ (open circles) and the measured amplification γ (filled squares), with the string-stability bound of one dashed. (b) The lane flow over the last 30 s. Notice that each shorter time gap of radar ACC buys flow, reaching the bound at 1.0 s and passing it at 0.8 s, while the radio holds 0.6 s within it at the flow of case 5, and that the drivers measure 1.29 where the criterion predicts only 1.024.

Table 14.4 The Ten Cases and the Challenge

#	Change from the reference line	Teaching point
1	people 0.70 s late, lead in MANUAL, tap at 6 s	the accordion
2	people 0.70 s late, tap at 5 s	delay as a cliff
3	constant spacing at 33 m, gains 0.3 and 1.0, tap	no gain suffices
4	time gap 1.5 s, tap	stability and its price
5	time gap 0.6 s, tap	well damped, unstable
6	CACC at 0.6 s, 50 ms radio, tap	stability regained
7	CACC with fallback at 0.6 s, 400 ms, 20 % loss, tap, burst at 40 s	the network's price
8	CACC with fallback at 0.6 s, cut-in at 10 s	shaping the request
9	22 people on a 437 m ring at 30 km/h	the phantom jam
10	MPC at 0.4 s, broadcast plan, tap	planning against CACC
11	challenge, 30 km/h, 0.6 s, 10 % loss, cut-in, stop and go	graceful degradation

times the RMS acceleration they receive, at car 6, and the closest call is 2.09 m, car 10 behind car 9. Reset replays the tap, so a student's own pedal input adds to it. Braking at 0.5 m/s^2 from 10 to 19 s takes the lead down to 39.6 km/h, and car 10 still reaches zero at 27.14 s. Accelerating at 1.5 m/s^2 from 7.6 to 10.7 s restores about 72 km/h and keeps car 10 above 42.8 km/h.

Case 2 gives the lead its cruise control, which brings it back from a low of 57.3 km/h at 7.1 s. Car 10 bottoms out at 28.3 km/h at 26.7 s, and the measured ratio rises from 0.80 at car 2 to 1.29 at car 10. The reaction delay is a cliff. At 0.5 s the line measures 0.99 and car 10 stays above 60.5 km/h. At 0.6 s it measures 1.03, and at 0.75 s car 10 falls to 0.5 km/h. At 0.8 s car 10 runs into car 9 at 9.7 km/h after 26.62 s. At 1.0 s four collisions follow, the first car 7 into car 6 at 26.0 km/h after 20.95 s. Over the same range of delay the linear peak rises only from 1.017 to 1.084, while the measured ratio is already 1.29 at 0.70 s. The tap is not a small wave, and the driver model's braking is nonlinear in the approach rate.

Table 14.5 Measured Amplification, Car 10's Lowest Speed, Smallest Gap and Lane Flow. Runs Last 90 s Except Where Named

Case and configuration	$\|\Gamma\|_{max}$	γ (car)	v_{10} (km/h)	d_{min} (m)	q (veh/h)
1, as shipped	1.024	1.29 (6)	0.0	2.09	1462
2, T_r 0.70 s	1.024	1.29 (10)	28.3	20.55	1696
2, T_r 0.5 s	1.017	0.99 (10)	60.5	29.63	1757
2, T_r 0.8 s	1.031	1.35 (7)	0.0	collision	412
2, T_r 1.0 s	1.084	1.44 (5)	0.0	4 collisions	255
3, as shipped	1.344	1.18 (3)	6.3	21.88	1920
3, k_p 0.05, k_d 1, 120 s	1.050	1.02 (10)	56.5	29.04	1920
3, k_p 2, k_d 1	131.7	4.23 (2)	0.0	8.54	973
4, h 1.5 s	1.000	0.93 (10)	66.8	28.18	1920
4, h 1.2 s	1.000	0.95 (10)	64.4	22.64	2286
4, h 1.0 s	1.003	0.99 (10)	61.2	18.97	2618
4, h 0.8 s	1.084	1.06 (6)	54.5	14.82	3064
5, h 0.6 s	1.306	1.20 (3)	36.3	8.22	3692
5, k_p 3	1.305	1.21 (3)	37.9	8.95	3692
5, k_d 2	1.089	1.06 (6)	53.2	11.35	3692
6, CACC 0.6 s	1.000	0.95 (10)	63.7	12.81	3692
6, h 0.4 s	1.033	1.00 (5)	58.1	9.40	4645
6, h 0.3 s	1.378	1.15 (3)	46.2	6.87	5333
6, k_a 0.5	1.048	1.05 (8)	57.2	12.25	3692
7, as shipped, 120 s	1.204	1.07 (5)	37.6	12.12	3692
7, h 0.8 s, 120 s	1.000	0.93 (10)	43.9	16.37	3063
7, plain CACC, 120 s	1.204	1.07 (5)	52.2	12.12	3692
8, regrowth 1 m/s, 80 s	1.000	—	59.1	5.24	3117
8, regrowth 0, 80 s	1.000	—	36.5	5.24	3117
9, ring, 300 s	1.198	1.03 (6)	—	—	810
9, ring on ACC 1.5 s, 300 s	1.000	—	—	—	1494

$\|\Gamma\|_{max}$ at the start of the run, v_{10} the lowest speed of car 10, q over the last 30 s. A run with a collision pauses at the first one, and its row is read after Run is pressed again.

14.6.2 Spacing Policy and Gains

Case 3 holds a fixed 33 m gap with gains that damp each car at 0.816. After the tap car 10 bottoms out at 6.3 km/h at 17.6 s, a dip of 65.7 km/h behind the lead's 14.7, and the line measures 1.18 at car 3, every car above one. The best pair of the grid, which the spacing-gain slider reaches at its lower end, still measures 1.02 over 120 s, and the time-gap gains measure 4.23 at car 2, with six cars standing still at once. This is the honestly infeasible case, because no gain on the grid or off it brings the low-frequency term of Section 14.4 below one.

Case 4 asks for 3 m plus 1.5 s of travel, the same 33 m at 72 km/h. The tap now dies out on its way back. Car 10's low is 66.8 km/h at 20.4 s against the lead's 57.3, and the measured ratio rises from 0.68 at car 2 to 0.93 at car 10, every car passing back less than it receives. The price is flow, 1920 veh/h. Shortening the gap buys flow with amplification. At 1.2 s the line measures 0.95 at 2286 veh/h, at 1.0 s, with its peak of 1.003 just past the boundary, 0.99 at 2618 veh/h, and at 0.8 s it amplifies, 1.06 at 3064 veh/h.

Case 5, at 0.6 s, is the chapter's central demonstration, shown in Figure 14.4. Each car on its own is damped at 0.496, while the criterion peaks at 1.306 at 1.12 rad/s, a period of 5.6 s. Car 10 bottoms out at 36.3 km/h at 16.1 s and the line measures 1.20 at car 3. The tap takes the lead's acceleration to −2.93 m/s^2, the middle car's, car 6, to −3.20 m/s^2 at 10.80 s and the last car's to −4.02 m/s^2 at 14.60 s. The last car then commands its 2 m/s^2 limit from the control period at 16.20 to the one at 21.50 s, its gap peaks at 27.23 m at 28.84 s, it reaches its own set speed of 86.4 km/h at 31.85 s, and it brakes again, to −2.87 m/s^2 at 33.05 s, as it closes the gap, a second disturbance that the catching up creates by itself. The lane would carry 3692 veh/h, and the line cannot hold together at that density. A spacing gain of 3 moves the peak without lowering it (1.305, measured 1.21), and a rate gain of 2 lowers it without removing it (1.089, measured 1.06), because at the top of the slider, $k_p = 3$, the

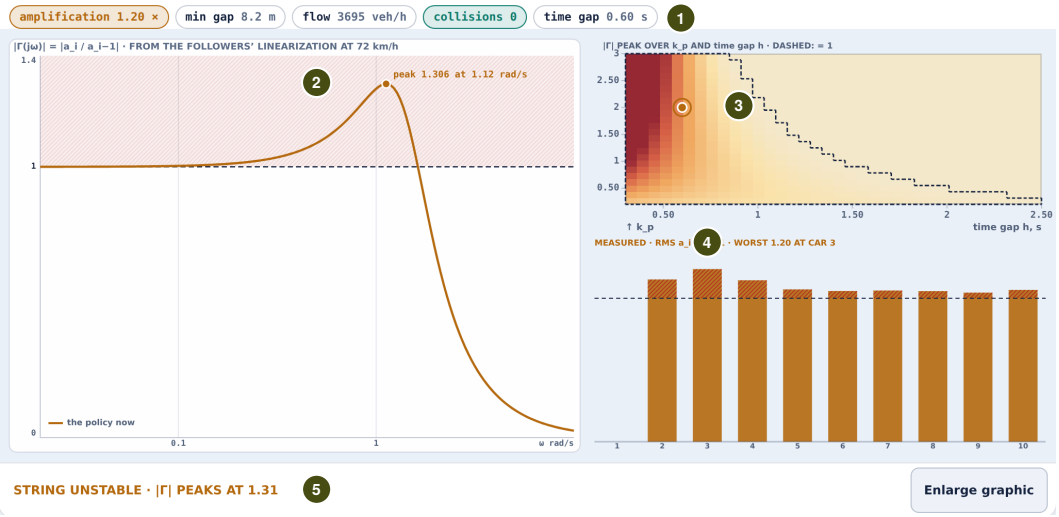


Figure 14.4 Case 5, the gap shrunk to 0.6 s, at $t = 60.00$ s in the $|\Gamma(j\omega)|$ view, run in AUTO with no control changed, cropped to the plant view. The badges (1) read amplification 1.20×, min gap 8.2 m, flow 3695 veh/h, collisions 0 and time gap 0.60 s. The left panel draws $|\Gamma(j\omega)|$ from the followers' linearization at 72 km/h, with its peak of 1.306 at 1.12 rad/s (2) marked in the hatched region above one, and the map beside it places the operating point (3), $k_p = 2$ and $h = 0.6$ s, left of the dashed contour where the peak equals one. Below the map every measured bar (4) from car 2 to car 10 is above one, the worst 1.20 at car 3, and the status line (5) reads String unstable, $|\Gamma|$ peaks at 1.31. The run's trends and scorecard are cut, and Table 14.5 gives its measured values.

boundary $\sqrt{2/k_p}$ is still 0.82 s.

14.6.3 The Radio and Its Failures

Case 6 keeps the 0.6 s gap and the gains and adds the radio at $k_a = 1$. The tap dies out, car 10 bottoming out at 63.7 km/h at 12.9 s and every car passing back at most 0.95, and the lane still carries 3692 veh/h, 92 % more than the 1.5 s ACC of case 4. At 0.4 s, just past the 0.413 s boundary, the peak is 1.033 and the line measures 1.00 at 4645 veh/h. At 0.3 s it amplifies again, 1.378 and measured 1.15, and half the feedforward gain leaves a peak of 1.048 at 0.6 s, measured 1.05.

> **TRY IT IN THE APP** **6 · Cooperative ACC at 0.6 s**
>
> Load the scenario, press Run and watch the speed bars after the brake tap at 5 s. Car 10 bottoms out at 63.7 km/h, where the radar-only line of case 5 took it down to 36.3 km/h, and every car passes back at most 0.95 of the acceleration it receives. The gap and the gains are those of case 5. Only the radio is new.

Case 7 slows the radio to 400 ms with one packet in five lost, an effective 412.5 ms. The radio term now arrives too late to help at 0.6 s. The peak is 1.204 at 1.64 rad/s, the line measures 1.07 at car 5, the stable gap moves out to 0.737 s, and at 0.8 s the line measures 0.93 at 3063 veh/h. Random loss of one packet in five almost never stops a link, because a stop needs ten silent periods in a row, a chance of 0.2^{10}, about 10^{-7}, at each packet, and none of the nine links stops before the burst at 40 s, while all nine stop during it, at the control periods at 40.85 and 40.90 s. Car 6's link counts as stopped at 40.90 s, its time gap widens to 1.160 s, the burst ends at 46 s and the link counts as live again at 46.50 s, and the gap is back within 0.01 s of 0.6 s by 72.0 s. The fallback is not free. When every gap widens at once the last car must fall back by all of them, and car 10 slows from 72 to 37.6 km/h at

47.5 s, deeper than its low of 52.2 km/h after the tap. Plain CACC, which rides out the silence on radar alone at 0.6 s with no wave to amplify, keeps car 10 above 71.1 km/h after 38 s.

> ✦ DID YOU KNOW? **A Fallback That Deepens the Dip**
>
> From the simulator: in case 7 the fallback widens every gap when the radio burst silences all nine links, and car 10's lowest speed in the run is 37.6 km/h. Plain cooperative ACC rides out the same silence on radar alone, and its lowest, 52.2 km/h, comes from the brake tap (Table 14.5).

14.6.4 A Cut-In

In case 8 a car without a radio cuts into the middle of the 15 m gap ahead of car 4 at 10 s, leaving 5.24 m on either side. With the gap regrowing at 1 m/s, car 4 brakes at no more than 2.02 m/s^2 and the cars behind it slow to 59.0 to 59.1 km/h and recover. Asked for the whole gap at once, car 4 faces a shortfall of 9.88 m and brakes at 5.96 m/s^2, close to its limit, the car that cut in brakes at 5.70 m/s^2 for its own 1.5 s gap, car 4 drops to 26.9 km/h and the cars behind it to between 30.7 and 36.5 km/h. Either way the cut-in costs flow, from 3691 veh/h before it to 3117 veh/h over the last 30 s, because car 4 now keeps 1.5 s behind a car with no radio. Nothing in the loop differs between the two runs except what the supervisor asks of it.

14.6.5 The Ring

Case 9 puts 22 drivers on the 437 m ring at 30 km/h, each displaced by at most 10 cm. Linearized there, the drivers peak at 1.198, far above their 1.024 on the open road at 72 km/h. The spread between the fastest and slowest car grows from 0.4 km/h at 30 s to 2.3 at 60 s and 11.3 at 90 s, the first car stops at 98.7 s, and from then on a stop-and-go wave runs backward around the ring with up to six cars standing at once. The ring carries 810 veh/h over the last 30 s of five minutes, against 1510 veh/h at the uniform start. With a reaction delay of 0.6 s the first stop comes at 132.0 s, and with 0.8 s at 75.8 s. Radar ACC at 1.5 s on every car keeps the same start uniform, with a spread of 0.1 km/h at 1494 veh/h. The jam also lowers the line's capacity. In a model run of 16 cars whose scripted tap is replaced by a hard stop at 5 s, these drivers discharge from the standstill at 997 veh/h, against the 1766 veh/h that their equilibrium carries at its best speed, 65.9 km/h, a drop of 44 %, while radar ACC at 1.5 s discharges at 1880 veh/h against 1920, a drop of 2 %. Fired by hand at 7, 30 or 60 s of case 2 with 16 cars, after its tap, the hard stop gives 939 to 1183 veh/h for the drivers, and 1880 again for the ACC of case 4. The ramp-metering application of Volume 2, Chapter 8 declares a capacity drop of 10 %, between these two measurements of this model.

> TRY IT IN THE APP **9 · The Ring Road: A Jam From Nothing**
>
> Press Run on the Time–space view at 20× and watch for the first red stripe. The first car stops at 98.7 s, and a stop-and-go wave then runs backward around the ring. Press Reset, set the Follower policy to ACC · time gap and run again: the spread of speeds stays at 0.1 km/h.

14.6.6 Planning Against Cooperative ACC

Case 10 runs case 6's cooperative line at 0.4 s, just past the 0.413 s boundary of cooperative ACC, under the controller of Section 14.5.4 with the plan preview and a tap at 5 s. Table 14.6 compares the structures on several cases with the same seed, noise and radio in every row and the controller's weights never retuned. Cooperative ACC retuned once by the rule that chose those weights, on the same 0.3 s line, takes $k_p = 0.5$ and $k_d = 3$ (peak 1.031, 4.57 m left in a hard stop at 3 s), and the strongest classical form found feeds those gains the commanded acceleration of the car ahead, which leads the delivered one by the lag.

On the tap of case 10 the controller measures 0.980, worst at car 2, against 1.003 for cooperative ACC, its linearization stays at one where that of cooperative ACC peaks at 1.033 at 1.38 rad/s, and its RMS command jerk is 1.4 m/s^3 against 3.3. With only the one acceleration the radio carries today

Table 14.6 Cooperative ACC Against the Predictive Controller: Measured Amplification γ, Criterion Peak $|\Gamma|_{max}$, RMS Command Jerk J, Lane Flow q and Smallest Gap d_{min}

	10, h 0.4 s, tap			6, 0.6 s	6 at 0.3 s		7, 400 ms, 20 %[a]		11[a]	hard stop, d_{min} (m)	
Structure	γ	$\|\Gamma\|_{max}$	J (m/s^3)	γ	γ	$\|\Gamma\|_{max}$	γ	q (veh/h)	γ	at 8 s	at 30 s
CACC, shipped gains	1.003	1.033	3.3	0.950	1.152	1.378	1.072	3339	0.989	2.92	2.93
CACC, commanded acceleration	0.968	1.033	8.1	0.959	0.976	1.378	1.001	3347	0.992	3.01	3.04
CACC retuned, k_p 0.5, k_d 3	0.975	1.000	4.7	0.941	1.018	1.031	0.967	3306	0.989	3.01	3.02
retuned, commanded acceleration	0.954	1.000	12.7	0.946	0.961	1.031	0.977	3306	0.988	3.22	3.20
MPC, plan preview	0.980	1.000	1.4	0.939	1.028	1.051	0.987	3371	0.985	2.95	2.97
MPC, acceleration preview	0.992	1.000	1.3	0.935	1.045	1.051	1.022	3342	0.988	2.98	2.96
MPC, preview off[b]	—	1.000	0.9	—	—	1.000	—	1920	0.974	3.33	3.48

Runs of 60, 60, 60, 80 and 110 s, default seed. Hard stops on case 10's line, runs of 40 and 60 s. [a]Cooperative rows with the fallback. Every challenge row passes except preview off (1497 veh/h). [b]With the preview off the supervisor widens every gap to $h_{fb} = 1.5$ s, the tap's pairs fall inside the quiet time of Equation (14.9) and γ has no value, and the criterion describes the law at 1.5 s. Every MPC run solved 7,200 to 19,799 plans, none rejected, at 0.88 to 0.90 ms for each period's nine solves in Node.

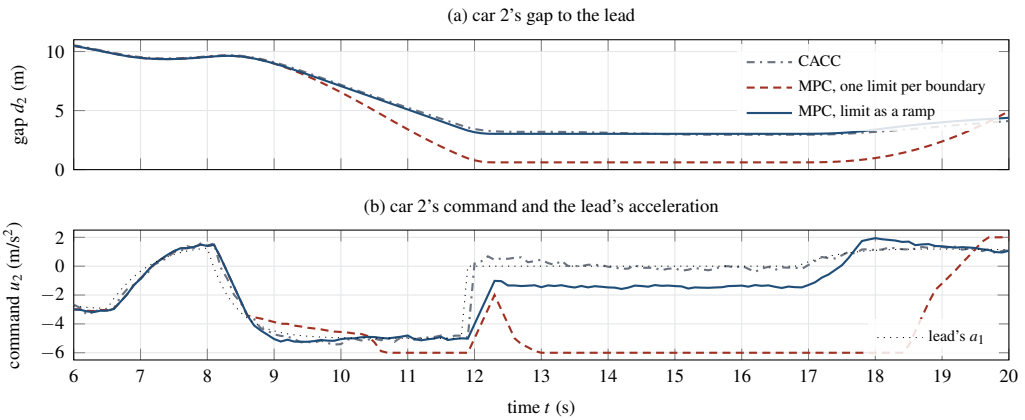


Figure 14.5 Car 2 in a hard stop fired at 8 s on the 0.4 s line of case 10. (a) Its gap and (b) its command beside the lead's acceleration (dotted), under cooperative ACC (dash-dotted, smallest gap 2.96 m), the predictive controller's first build, one jerk limit per block boundary (dashed, 0.61 m), and the controller with the limit as a ramp (solid, 3.03 m).

it measures 0.992, and 0.935 against 0.939 on case 6, so most of the gain is the controller, not the message. The plan counts where the radio is slow, 0.987 against 1.022 in case 7. Car 2, which always previews the lead's one acceleration, is the worst car under the plan in cases 7 and 10. With the preview off the widened gap leaves no wave to grow, but the lane carries 1920 veh/h instead of 4648, and with the widening removed, a bound kept out of the application, the controller measures 1.116.

Cooperative ACC retuned once by the controller's own rule beats it on amplification on case 10, 0.975 against 0.980, at 0.3 s, 1.018 against 1.028 with a peak of 1.031 against 1.051, and on case 7, 0.967 against 0.987. With the commanded acceleration it reaches 0.954 on case 10, the best of any row, at an RMS jerk of 12.7 m/s^3, nine times the controller's. The controller's advantage is therefore not its amplification but the smoothest commands of the table, 0.9 to 1.4 m/s^3 on case 10 against 4.7 for the retuned law, limits held as constraints and a plan shown to the car behind. Every row but the radar-only controller passes the challenge, at 0.985 to 0.992.

The hard stop is the case the first build lost, as shown in Figure 14.5. Fired at 8 s, it left 0.61 m under the plan preview and 0.46 m under the acceleration preview, a stop from cruise at 30 s left 0.30 m and a collision, and the radar-only controller collided nine times. The ramp form keeps 2.95, 2.98 and 3.33 m. The same stop was then fired at 17 instants from 1 to 30 s on four lines, each run continued 25 s after it. On cases 10, 6 and 7 and at 0.3 s no car under the controller collides at any

Table 14.7 The Challenge of 110 s in Five Configurations

Configuration	γ (car)	d_{min} (m)	q (veh/h)	Target
as shipped, ACC 0.6 s	1.13 (10)	1.75	2283	not met
ACC 1.5 s	0.98 (10)	5.50	1497	not met
CACC 0.6 s	1.04 (4)	1.75	2281	not met
CACC with fallback	0.99 (8)	1.75	2158	met
MPC, plan preview	0.99 (10)	1.75	2156	met

instant with any preview, the smallest gaps being 2.31, 0.46 and 2.01 m under the plan, acceleration and radar-only previews, while cooperative ACC at 0.3 s collides at four instants. The first build had collided at 9 and 13 instants under the plan preview, at 0.3 s and in case 7, at 8 to 14 on three lines under the acceleration preview and at 15 or 16 on every line with the preview off. On the challenge's line a stop before 5 s now leaves no collision, the cut-in being refused. A stop at 5 to 12 s still does under every structure (radar-only 5 to 6.5 s), as the radio-less car that cut in, still regrowing its gap, hits car 3 at up to 4.69 m/s, and none collides without the cut-in.

14.6.7 Stop and Go With a Cut-In

The challenge runs ten cars at 30 km/h and 0.6 s for 110 s with one packet in ten lost. At 5 s a car without a radio cuts in ahead of car 4 with 1.75 m on either side, and at 40 s the lead begins a stop and go to 10 km/h for 20 s. The target asks for no collision, a measured amplification of at most 1.00, no gap below 1.5 m and at least 2000 veh/h over the last 30 s. Table 14.7 compares five configurations. As shipped, radar ACC fails on amplification rather than on safety. The line measures 1.13 at car 10, cars 8, 9 and 10 come to a standstill at 52.1, 52.6 and 53.2 s, the smallest gap is the cut-in's 1.75 m, and the flow is 2283 veh/h. A 1.5 s gap is stable, at 0.98, but carries 1497 veh/h. Plain CACC is stable everywhere except behind the car that cut in, where car 4 runs on radar alone at 0.6 s and measures 1.04. CACC with the fallback, at the shipped gains, widens only car 4's gap to 1.5 s and passes, measuring 0.99 with no gap below 1.75 m at 2158 veh/h. This is the challenge's suggested setting. Its measured amplification of 0.989 leaves margins of 0.011, 0.25 m and 158 veh/h on the three targets, whether Run follows the suggestion directly or after Reset. Four other seeds of the radar noise and the radio losses give 0.988 to 0.991 for it, and 1.122 to 1.133 for the shipped radar ACC, so the verdict does not depend on the seed. The controller of Section 14.5.4 passes at 0.985, because the supervisor widens car 4's gap behind the car with no radio as it does for cooperative ACC.

14.7 Technology in Practice

What Measures It

Each follower of this chapter reads three things: the gap and its rate of change from a radar, its own speed and acceleration, and the acceleration of the car ahead over a radio. In service the gap comes from a millimeter-wave radar at the front of the car, which measures distance and radial velocity directly [23]. The radio is a vehicle-to-vehicle link whose messages carry position, speed and braking status [24]. The chapter's radar is a sampled link 50 ms late with 0.05 m and 0.05 m/s of noise and perfect target selection, and its radio is the same link model with a delay, a jitter and a loss probability. Figure 14.6 draws one follower's loop on that hardware.

What Moves It

The final element is the car's own throttle and brakes, which the platoon controller reaches only through the vehicle's low-level controller [25]. The chapter lumps that controller, the engine and the brakes into one first-order lag of 0.4 s, the same for throttle and brake, and clips the command to -6 to 2 m/s^2. A real driveline brakes faster than it accelerates, as the Assumptions note.

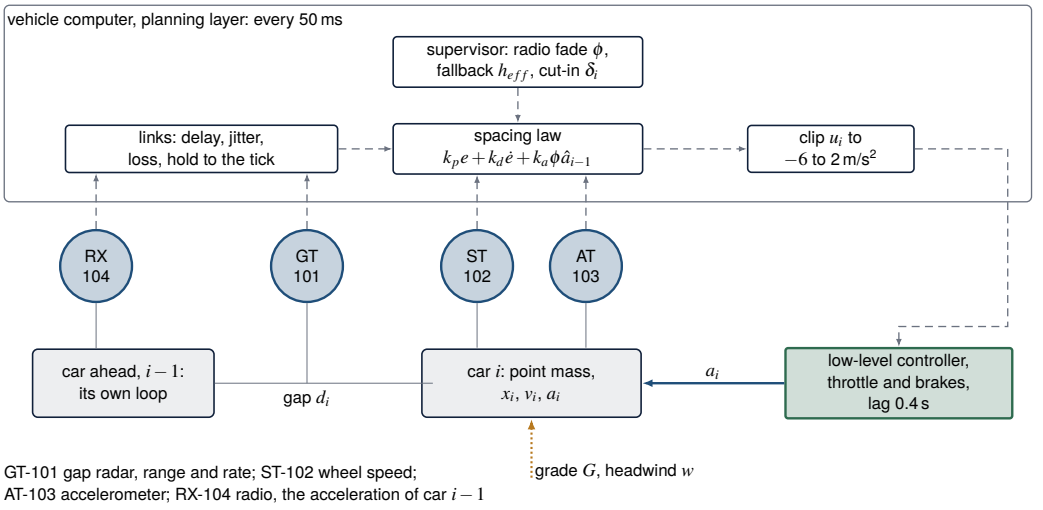


Figure 14.6 One follower's loop on vehicle hardware. The gap radar GT-101 and the radio receiver RX-104 reach the controller through the sampled links, the wheel speed ST-102 and the accelerometer AT-103 directly. The spacing law runs under the supervisor's fade, fallback and cut-in allowance, and its clipped command goes to the low-level controller of the throttle and brakes, the chapter's 0.4 s lag. Notice that the car ahead enters twice, through the gap the radar sees and through the acceleration the radio reports.

> **TECHNOLOGY** **Sensor: A Millimeter-Wave Radar at the Bumper**
>
> An automotive radar measures the distance and the radial velocity of a target directly, and with a suitable antenna system its angle. Its functions grew from distance warning to adaptive cruise control and emergency braking. Most new sensors work at 76 to 81 GHz and send fast sequences of frequency chirps, whose echoes a two-dimensional Fourier transform turns into a map of range and velocity [23]. A federal rulemaking notice gave a perception range of up to 200 m for a 77 GHz radar, with a field of view of 18 degrees [24]. The chapter's radar is 50 ms late, carries a little noise and always sees the right car.

Where the Controller Runs

In cooperative adaptive cruise control research the vehicle is organized in sensing, planning and actuation phases. Data from the radio unit and the onboard sensors are gathered on the vehicle's data bus, and the planning phase holds the high-level controller where the longitudinal control law runs. In one test fleet the production adaptive cruise control stayed installed, the cooperative law ran on a separate prototyping computer connected to the data bus, and the driver could switch between the two in real time [25]. The chapter's spacing law, the supervisor's fade, fallback and regrowth, and the predictive controller all belong to that planning layer and run every 50 ms. A proposed safety standard asked each car to broadcast ten times a second, every 100 ms [24], twice the period of the chapter's radio.

> **TECHNOLOGY** **Final Element: Throttle and Brakes Behind a Low-Level Controller**
>
> A platoon controller plans an acceleration and leaves its execution to a low-level controller, which converts the target into throttle and brake actions [25]. The delay of this actuation significantly limits the shortest inter-vehicle distance at which a cooperative platoon stays string stable [25]. In this chapter one lag of 0.4 s stands for all of it, and the bound on the time gap moves with it, because the lag enters the τs^3 term of the car-to-car transfer function. Its limits are the other half of the final element: in case 5 the last car commands its 2 m/s^2 limit from 16.20 to 21.50 s while it catches up.

What Is Modernizing It

Three developments are changing the hardware around this loop. The radio band is changing. The federal regulator kept the upper 30 MHz of the 5.9 GHz band for intelligent transportation systems and opened the lower 45 MHz to unlicensed use [26], and it required the transition from dedicated short-range communications to cellular vehicle-to-everything technology, with rules for the new devices and a timeline for retiring the old [27]. The radar is moving to multiple-input multiple-output antennas, digital beamforming and digital modulation, toward images of the whole scene rather than single reflections [23]. And the automated car is becoming an actuator of the traffic flow itself. On a ring track a single automated vehicle dampened the stop-and-go waves of about twenty human drivers, which suggests that flow control may be possible with a few such vehicles, less than 5 %, long before most cars are automated [10]. The chapter's ring of case 9 shows the same wave with every car under one policy.

> **IN PRACTICE** **What a Proposed Rule Asked of the Radio**
>
> A rulemaking notice of 2017 proposed Federal Motor Vehicle Safety Standard No. 150 for vehicle-to-vehicle radios. Each car would broadcast a basic safety message with its position, speed and braking status ten times a second under non-congested conditions, each transmission shifted at random by up to 5 ms to ease channel congestion. The message had to be received anywhere within at least 300 m with a packet error rate below 10 % [24]. The chapter's radio sends every 50 ms, and case 7 loses one packet in five, twice that error rate.

14.8 Limitations

The cars are identical point masses in one lane with one lag for throttle and brake, no road load and no tires, and a cut-in is instantaneous. The drivers are identical and noiseless and look one car ahead, so the reaction delay of 0.70 s is a value calibrated to stay short of the model's collisions at 0.8 s, not a measured human figure. The radar has perfect target selection, and the radio's losses are independent of how many cars share the channel. The criterion comes from one linearization at the mean follower speed, which describes small waves, and for people it understates how fast a large wave grows.

> **OPEN CHALLENGE** **Many Radars on One Road**
>
> Vehicles now carry up to twelve radars each, and no rules yet govern how sensors share the 76 to 77 GHz band. A simulation of a dense six-lane highway found many transmitters along the road reaching each receiver, which can lower sensitivity and accuracy and in rare cases blind the sensor. Detecting and repairing disturbed signals, active avoidance and cooperation between sensors are being studied [23]. Progress would be a radar whose gap and closing rate stay trustworthy in the densest traffic, where a platoon runs its shortest gaps. The chapter's radar has no interference.

The predictive controller's model is one car at the declared 0.4 s lag, which meets any other lag, grade or wind only through β and omits the stopped-car rule and collisions. Its preview cannot foresee the lead's brake, in error by 0.62 m/s^2 RMS one second ahead in a hard stop, and its criterion holds only while no limit binds.

One limitation concerns use rather than physics. Reset keeps every setting the student changed and replays the case's scripted events, its tap included, while reloading the scenario restores the case's own settings. A new time gap or policy lays the line out at the gap it wants only while the run stands at $t = 0$. A time gap moved while the line is running acts on the running line instead. For radar ACC the closing-up then enters the measurement, which reads 1.39 instead of 0.99 for 1.0 s chosen at 2 s, and for cooperative ACC the supervisor's gradual move pauses it. On the ring a policy chosen at $t = 0$ keeps the start speed of the drivers, 30.0 km/h, while one carried through Reset starts the ring at its own, 29.7 km/h, and the quoted figures of case 9 reproduce either way.

Every threshold is a teaching value. The challenge's 1.00 and 1.5 m, the 0.5 s link timeout and the fallback gap are not calibrations of any vehicle, and no number here describes how a production system performs.

OPEN CHALLENGE **A Radio Channel Shared by Every Car**

Every nearby car talks on the same channel, so a dense platoon is also a crowded channel. The 2017 notice called for research to identify the congestion that could impair safety applications and to develop mitigation. The strategies it weighed lower the number of messages a device sends each second, and the power or range of each one, when the device detects congestion [24]. A cooperative follower could then hear the car ahead less often just when its gap is shortest. Progress would be cooperative following that stays string stable under message rates that adapt to traffic density. In this chapter the losses do not depend on how many cars share the channel.

14.9 Conclusion

A line of individually well-damped cars can turn a brake tap into a stop, and this simulator measures by how much and why. At a 0.6 s time gap each radar follower is damped at 0.50 while the criterion peaks at 1.306, and car 10 dips to 36.3 km/h behind a lead that dipped to 57.3. Constant spacing amplifies for every gain, the best of 90 pairs peaking at 1.050. A time gap is stable above $\sqrt{2/k_p}$, 1.05 s with the radar's latency and the hold, at a cost in flow, 1920 veh/h at 1.5 s, and the predecessor's acceleration over a 50 ms radio regains 3692 veh/h at 0.6 s until 400 ms of delay and 20 % loss move the boundary to 0.74 s. Drivers 0.70 s late, barely unstable by the linearization at 1.024, collide once the delay reaches 0.8 s, and on a ring of 22 the first car stops 98.7 s after a uniform start. The challenge improves from 1.13 with three cars stopped to 0.99 at 2158 veh/h. A predictive controller in every follower measures 0.980 at a 0.4 s gap against 1.003 for cooperative ACC at 1.4 against 3.3 m/s^3 of command jerk, but the classical law retuned by its rule reaches 0.975, so it adds smoothness and constraints rather than amplification. The model is a single lane of identical, lagged cars whose drivers watch one car ahead. Further work would add mixed vehicles and drivers who watch several cars ahead, channel loss that grows with the number of cars, and a fallback that estimates the predecessor's acceleration from the radar. The platoon makes string stability a quantity that can be read off a chain of loops rather than inferred from one of them.

KEY POINTS

- A line of cars that are each well damped can still turn a brake tap into a much deeper slowdown at the back of the line.
- Constant spacing amplifies for every gain, and a time gap is string stable above $\sqrt{2/k_p}$, 1.05 s with the radar's latency and the hold, at a cost in lane flow.
- The predecessor's acceleration over a 50 ms radio regains 3692 veh/h at 0.6 s until 400 ms of delay and 20 % loss move the boundary to 0.74 s.
- Drivers 0.70 s late are barely unstable by the linearization, yet they collide once the delay reaches 0.8 s.

Where else this applies. Volume 9, Chapter 5 tells each gate's predictive controller the latest plan of the gate below, as each car here passes its plan to the car behind.

References

[1] L. Peppard, "String stability of relative-motion PID vehicle control systems," *IEEE Transactions on Automatic Control*, vol. 19, no. 5, pp. 579–581, 1974.

[2] D. Swaroop and J. K. Hedrick, "String stability of interconnected systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 349–357, 1996.

[3] G. Gunter, D. Gloudemans, R. E. Stern, S. McQuade, R. Bhadani, M. Bunting, M. L. Delle Monache, R. Lysecky, B. Seibold, J. Sprinkle, B. Piccoli, and D. B. Work, "Are commercially implemented adaptive cruise control systems string stable?" *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6992–7003, 2021.

[4] P. Seiler, A. Pant, and K. Hedrick, "Disturbance propagation in vehicle strings," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1835–1841, 2004.

[5] G. Gunter, D. Gloudemans, R. E. Stern, S. McQuade, R. Bhadani, M. Bunting, M. L. Delle Monache, R. Lysecky, B. Seibold, J. Sprinkle, B. Piccoli, and D. B. Work, "Are commercially implemented adaptive cruise control systems string stable?" arXiv preprint 1905.02108v1, 2019. [Online]. Available: https://arxiv.org/abs/1905.02108v1

[6] J. Ploeg, N. van de Wouw, and H. Nijmeijer, "$\mathscr{L}_p$ string stability of cascaded systems: Application to vehicle platooning," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 786–793, 2014.

[7] H. L. Lee, V. Padmanabhan, and S. Whang, "Information distortion in a supply chain: The bullwhip effect," *Management Science*, vol. 43, no. 4, pp. 546–558, 1997.

[8] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam," *New Journal of Physics*, vol. 10, no. 3, p. 033001, 2008.

[9] G. J. L. Naus, R. P. A. Vugts, J. Ploeg, M. J. G. van de Molengraft, and M. Steinbuch, "String-stable CACC design and experimental validation: A frequency-domain approach," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4268–4279, 2010.

[10] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.

[11] W. Levine and M. Athans, "On the optimal error regulation of a string of moving vehicles," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 355–361, 1966.

[12] R. E. Fenton and R. J. Mayhan, "Automated highway studies at the Ohio State University-an overview," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 100–113, 1991.

[13] S. E. Shladover, "PATH at 20—history and major milestones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 4, pp. 584–592, 2007.

[14] D. Swaroop, J. K. Hedrick, C. C. Chien, and P. Ioannou, "A comparision of spacing and headway control laws for automatically controlled vehicles," *Vehicle System Dynamics*, vol. 23, no. 1, pp. 597–625, 1994.

[15] R. Rajamani, H.-S. Tan, B. K. Law, and W.-B. Zhang, "Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 4, pp. 695–708, 2000.

[16] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2324, no. 1, pp. 63–70, 2012.

[17] J. Ploeg, E. Semsar-Kazerooni, G. Lijster, N. van de Wouw, and H. Nijmeijer, "Graceful degradation of cooperative adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 488–497, 2015.

[18] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

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

[20] W. B. Dunbar and D. S. Caveney, "Distributed receding horizon control of vehicle platoons: Stability and string stability," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 620–633, 2012.

[21] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 899–910, 2017.

[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] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive radar—from first efforts to future systems," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 135–148, 2021.

[24] National Highway Traffic Safety Administration, "Federal motor vehicle safety standards; V2V communications, notice of proposed rulemaking, docket no. NHTSA-2016-0126," Federal Register, vol. 82, no. 8, pp. 3854–4019, 12 January 2017, 2017. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2017-01-12/pdf/2016-31059.pdf

[25] Z. Wang, G. Wu, and M. J. Barth, "A review on cooperative adaptive cruise control (CACC) systems: Architectures, controls, and applications," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2884–2891.

[26] Federal Communications Commission, "Use of the 5.850–5.925 GHz band, final rule," Federal Register, vol. 86, pp. 23281–23299, 3 May 2021, 2021. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2021-05-03/pdf/2021-08802.pdf

[27] ——, "Use of the 5.850–5.925 GHz band, final rule," Federal Register, vol. 89, no. 240, pp. 100838–100856, 13 December 2024, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2024-12-13/pdf/2024-28980.pdf

14.10 Exercises

14.1 Concept check · 2 min

In case 5 each radar follower behind a car at constant speed has the closed-loop roots $-0.684 \pm 1.197\,j$ and $-2.632\,\mathrm{s}^{-1}$, a damping ratio of 0.496, yet the brake tap takes car 10 from 72 km/h down to 36.3 km/h. Which statement explains this?

(a) The damping ratio leaves out the radar's latency, and with the latency included each follower's own loop is unstable

(b) The tap drives every follower onto its acceleration limits, and the line amplifies only because the limits clip the commands

(c) A damping ratio of 0.496 overshoots by about 17 %, and nine such overshoots compound into the dip at car 10

(d) The car-to-car gain $|\Gamma|$ exceeds one from zero up to 1.63 rad/s, peaking near the loop's own natural frequency

14.2 Concept check · 2 min

For radar ACC with constant spacing, Section 14.4 finds $|\Gamma(j\omega)|^2 \approx 1 + 2\omega^2/k_p$ near zero frequency. What follows for the follower's gains?

(a) No pair of positive gains makes the line string stable, and the best of 90 pairs on the grid still peaks at 1.050

(b) A spacing gain above about $2\,\mathrm{s}^{-2}$ makes the low-frequency term negative and the line string stable

(c) The rate gain decides it, and any k_d larger than k_p brings the peak below one

(d) The line becomes string stable once the radar's latency is set to zero, since the expansion holds only with delay

14.3 Concept check · 2 min

The app measures the amplification of a car as the ratio of its RMS acceleration to that of the car ahead, Equation (14.9), and not as a ratio of peak accelerations. Why?

(a) Peak accelerations carry the radar's noise of 0.05 m/s, while RMS values average that noise away

(b) The criterion bounds the RMS ratio by Parseval's theorem, while a string-stable line can pass back larger peaks

(c) A peak ratio cannot be formed during a cut-in, so an RMS ratio taken over the whole run replaces it

(d) Peak ratios exceed one only on the ring road, so they cannot tell the open-road cases apart

14.4 Theory and calculation · 7 min

Radar ACC with a time gap and no delays has the transfer function of Equation (14.11), with $\tau = 0.4\,\mathrm{s}$ and $k_d = 1\,\mathrm{s}^{-1}$.

(a) Use the coefficient c_2 of Equation (14.12) to find the shortest string-stable time gap for k_p of 1, 2 and $3\,\mathrm{s}^{-2}$. *Checkpoint:* 1.414 s, 1.000 s and 0.816 s.

(b) At $h = 1.0\,\mathrm{s}$ and $k_p = 2\,\mathrm{s}^{-2}$, evaluate c_2 and c_4, and say whether the full condition of Section 14.4 holds. *Checkpoint:* $c_2 = 0$, $c_4 = 1.6$; it holds, on the boundary.

(c) At $h = 0.6\,\mathrm{s}$ with $k_p = 2$, evaluate $|\Gamma(j\omega)|$ at 1 rad/s, and find the k_p that the low-frequency bound would need at this gap. *Checkpoint:* 1.213; $k_p \geq 5.56\,\mathrm{s}^{-2}$.

14.5 Theory and calculation · 6 min

A radar follower at $h = 0.6\,\mathrm{s}$ with $k_p = 2\,\mathrm{s}^{-2}$ and $k_d = 1\,\mathrm{s}^{-1}$ follows a car at constant speed. Its characteristic polynomial is the denominator of Equation (14.11).

(a) Write the polynomial with numbers, and check that $s = -2.632$ is nearly a root. *Checkpoint:* $0.4s^3 + 1.6s^2 + 2.2s + 2$; the residual is about 3×10^{-4}.

(b) Divide out the real root and find the natural frequency and the damping ratio of the remaining pair. *Checkpoint:* $s^2 + 1.368s + 1.899$; 1.378 rad/s and 0.496.

(c) The $|\Gamma|$ peak lies at 1.12 rad/s. Find the period of a wave at that frequency, and say what case 5's sine profile at a period of 5.6 s probes. *Checkpoint:* 5.61 s.

14.6 Theory and calculation · 6 min

The lead holds 72 km/h. Use $L = 4.5\,\mathrm{m}$ and $r = 3\,\mathrm{m}$.

(a) Find the wanted gap and the flow q_h of Equation (14.8) at time gaps of 1.5, 0.6 and 0.4 s. *Checkpoint:* 33, 15 and 11 m; 1920, 3692 and 4645 veh/h.

(b) Case 7's radio has $\theta = 400\,\mathrm{ms}$ and $p = 20\,\%$ with no jitter. Find the effective delay θ_c the criterion uses. What does a radar latency of 120 ms become in the app? *Checkpoint:* 412.5 ms; 150 ms.

(c) A link counts as stopped after ten periods without an arrival. Find the chance that ten packets in a row are lost at $p = 0.2$, and the expected number of stops from random loss alone on nine links in the 40 s before case 7's burst, counting one chance at each packet. *Checkpoint:* 1.0×10^{-7}; about 7×10^{-4}.

14.7 App challenge · 8 min

Load scenario 4, *Constant time gap at 1.5 s*, at the default seed. Run to $t = 90\,\mathrm{s}$ at a playback speed of 5, pause, read the Measured amplification and Lane flow tiles, and export the CSV. Then press Reset, set Time gap to 1.0 s before Run, and run to 90 s again. Repeat with a time gap of 0.8 s. Car 10's lowest speed is the minimum of the column `v_last_mps` (the Slowest any car went tile shows the lead here). Tolerances: 0.01 on amplification, 5 veh/h on flow, 0.2 km/h on speed.

(a) Read the three values at 1.5 s. *Checkpoint:* 0.93 at car 10, 1920 veh/h, 66.8 km/h.

(b) Read them at 1.0 and 0.8 s, with the $|\Gamma|$ peak tile. *Checkpoint:* 0.99 at car 10, 2618 veh/h, 61.2 km/h, peak 1.003; 1.06 at car 6, 3064 veh/h, 54.5 km/h, peak 1.084.

(c) The peak at 1.0 s is above one, yet the line measures 0.99. Explain, using the boundary of the app's criterion. *Checkpoint:* The boundary is 1.050 s, and the excess is tiny and at 0.37 rad/s.

14.8 App challenge · 8 min

Load scenario 5, *The gap shrunk to 0.6 s*, at the default seed and run to 90 s at a playback speed of 5. Read the Measured amplification, $|\Gamma|$ peak and Slowest any car went tiles. Then load scenario 6, *Cooperative ACC at 0.6 s*, run to 90 s and read the same tiles, the Lane flow tile, and car 10's lowest speed from the column `v_last_mps` of the CSV. Finally press Reset, set Time gap to 0.4 s before Run, and run to 90 s. Tolerances as in the previous exercise.

(a) Read case 5. *Checkpoint:* 1.20 at car 3; 1.306; 36.3 km/h.

(b) Read case 6 at 0.6 s, and say what changed between the two cases. *Checkpoint:* 0.95 at car 10; 1.000; 3692 veh/h; car 10 at 63.7 km/h.

(c) Read case 6 at 0.4 s and relate the result to the cooperative boundary. *Checkpoint:* 1.00 at car 5; 1.033; 4645 veh/h.

14.9 App challenge · 9 min

Load the challenge, *Challenge, stop and go with a cut-in*, at the default seed. The run stops itself at 110 s, and a playback speed of 5 takes 22 s of wall time. Before each later run press Reset, choose the Follower policy, then press Run. Leave every other setting as loaded and do not reveal the suggestion. Read the Measured amplification, Smallest gap, Lane flow and Challenge target tiles (tolerance 0.01, 0.01 m, 5 veh/h).

(a) Run as loaded, on ACC with a time gap. Which target fails? *Checkpoint:* 1.13 at car 10, 1.75 m, 2283 veh/h; amplification fails.

(b) Run with the policy CACC. *Checkpoint:* 1.04 at car 4, 1.75 m, 2281 veh/h; not met.

(c) Run with CACC + fallback. Then press Reset, set the policy back to ACC · time gap and run once more. What does the last run show? *Checkpoint:* 0.99 at car 8, 1.75 m, 2158 veh/h, met; the last run repeats (a).

14.10 Application area · 4 min

In case 8 a car with no radio cuts into the 15 m gap ahead of car 4 of a cooperative line at 0.6 s and 72 km/h. With the gap regrowing at 1 m/s, car 4 brakes at no more than 2.02 m/s^2. Asked for the whole gap at once, it brakes at 5.96 m/s^2. The lane carries 3691 veh/h before the cut-in and 3117 veh/h over the last 30 s.

(a) Find the fraction of the flow lost, and say why the flow falls although no car collides. *Checkpoint:* 15.6 %.

(b) What should a fleet engineer specify for the supervisor of a short-gap platoon, and why must the platoon plan for cut-ins at all? *Checkpoint:* A rate-limited regrowth of the gap, and a fallback gap behind any car without a radio.

14.11 Application area · 3 min

After a hard stop in a model run of 16 cars, drivers 0.70 s late discharge from the standstill at 997 veh/h, against the 1766 veh/h that their equilibrium carries at its best speed. Radar ACC at 1.5 s discharges at 1880 veh/h against 1920 veh/h. The ramp-metering app declares a capacity drop of 10 %.

(a) Find both drops. *Checkpoint:* 44 % and 2 %.

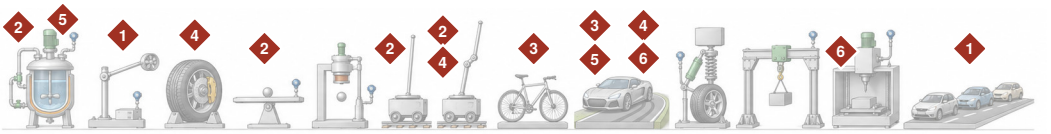
(b) What does the comparison suggest to a traffic engineer, and what in Section 14.8 keeps it from being a forecast? *Checkpoint:* A queue of these drivers discharges far below capacity, and the model's drivers are identical and look one car ahead.

14.12 Review the agent · 5 min

Asked how far a brake tap travels down case 2's line of drivers, an AI assistant answers: "The linearized criterion peaks at 1.024 for drivers 0.70 s late, so each car passes on at most 2.4 % more than it receives. Over nine cars the dip grows at most $1.024^9 = 1.24$ times. The lead dips 14.7 km/h, from 72 to 57.3, so car 10 dips at most 18.2 km/h and stays above 53.8 km/h. Reaction delays near 0.7 s are therefore safe in a ten-car line." Find the flaw, and say how you would check the claim in the app and what the check shows.

CHAPTER 15

Summary and Outlook



Open gaps 1 margins 2 regions 3 scheduling 4 estimation 5 equilibria 6 models

Notice that the drifting car carries four of the six gaps, and that the gap of recoverable regions joins the reactor, the ball and beam and both carts. Illustration, not to scale.

IN PLAIN TERMS **What Thirteen Unsteady Machines Taught**

The main idea. Each chapter of this book held a machine that would fall, spin or run away on its own, or that kept swinging after it was moved. A controller had to watch it and push it back, many times a second.

The challenge. How far each machine could be pushed before it was lost was set by the machine, its motor and its sensors, not only by the controller's settings. Past that edge the loss came suddenly, and the usual checks did not always see it coming.

Why it matters. Measuring that edge on the part that fails tells a safe design from one that only looks safe.

Thirteen chapters held thirteen plants that would not hold themselves, or that rang when they were moved. Chapter 1 asked two questions of each plant. What makes it hard, and what holds it by how much? Table 1.2 answered the first and named the structure and the margin of each chapter without numbers. This chapter adds the numbers and the structures that failed, draws a short guide for choosing a stabilizing design, 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.

15.1 What the Chapters Showed

Tables 15.1 and 15.2 compare the chapters. Each row names the difficulty, the structure that held the plant, a structure or setting that failed, and the margin with the table, figure or section of the chapter that reports it. The first table holds the plants whose gain changes with the operating point and the plants with one unstable mode. The second holds the plants with two unstable modes or a speed that moves the poles, the oscillating plants and the line of cars. Figure 15.1 sets one like pair of numbers from nine of the rows side by side.

The columns of the two tables make five points.

The margin ends at a cliff. Most of the measured margins are thin, and the loss past them is sudden. One kelvin of setpoint separated an overshoot of 12.18 K from one of 69.07 K in Chapter 2. One millisecond of sample interval separated a held double pendulum from a fallen one. At 0 % comfort the leveling loop of Chapter 11 decays at 11 N/mm and grows at 12 N/mm. The line of cars of Chapter 14 held at a reaction delay of 0.5 s and collided at 0.8 s. A stabilizing loop has no gentle region of poor performance between good and lost, so the distance to the edge is the number to report.

The usual margin did not always see the edge. The aeropendulum's PID kept a phase margin of 49 to 55° up to 105° and went unstable at 105.8°, because the mode that failed was a slow one the margin does not describe. The bicycle's schedule was stable by a continuous analysis and grew at up

Table 15.1 What Held Each Plant and by How Much, Chapters 2 to 7

Chapter	Difficulty	What held it	What failed	Margin measured
2, stirred reactor	Every temperature from 335.654 to 379.611 K is open-loop unstable, the 340 K setpoint included.	Reverse-acting PI from a step-test fit, $K_c = 4$, $\tau_I = 1.333$ min. Predictive control on the linearization.	A setpoint step 1 K larger. A lower gain of 1.6, which raised the peak and left a relaxation cycle.	An 11 K step overshoots 12.18 K, a 12 K step 69.07 K and trips (Table 2.5). Largest safe step 11.5 under PI, 13.5 under MPC (Section 2.8).
3, propeller arm	Gravity's stiffness changes sign at horizontal. At 120° the arm diverges at 4.06/s.	Gravity feedforward, a gain schedule, an integral zone and a derivative filter, all four together.	One PID tuned at 30°. Its slow mode is unstable from 105.8° with a phase margin of 49 to 55°.	Program settled in 2.26 s with all four, never without the schedule or the feedforward (Table 3.6). Phase margin down to 19.1° at 150° (Figure 3.2).
4, braked wheel	Slip held at the friction peak, where the gain vanishes. Equilibrium lost at $\lambda_c = 0.1661$.	PI on slip. A sliding-mode law with its boundary layer sized to the brake lag.	A locked wheel, 47.9 m. A rule-based cycle, locking in 201 of 348 runs.	Stop 33.0 m against 31.8 m ideal (Table 4.6). Phase margin 40.6 at 27 m/s, 15.2° at 3 m/s (Table 4.4). Locks in 60 of 300 runs against 80 of 336 (Table 4.9).
5, ball and beam	Two integrators and a servo lag, no damping.	A filtered derivative. A PID on a profile with feedforward.	Proportional alone, each swing 1.37 to 1.42 times the last. An unfiltered derivative, ball flung 66 to 94 mm.	Phase margin 48.5° (Table 5.6), 14.4° lost to the camera's delay (Section 5.6). Relay race 0 arrivals against 15 (Table 5.8).
6, levitated ball	One real pole at 31.3 rad/s, a 10 ms coil, a 3 A ceiling.	A derivative time longer than the coil's, with an equilibrium feedforward.	Integral action, at any setting. A gain above the magnetic stiffness of 88.9 A/m with too little derivative.	Critical gain 107.0 A/m, found between 105 and 110. Critical derivative time 10.30 ms (Section 6.6). A 1 mm drop at 28 mm uses 93 % of the ceiling (Section 6.7).
7, cart and pole	An unstable root of 5.58/s, one e-fold every 179 ms, a wrong-way zero, a rail.	A cascade, angle inside cart position. An LQR on the full state.	The angle loop alone, at the bumper in 4.11 s. The cascade on a 150 °/s push.	Gain window closed at about 105 ms of sensing lag (Figure 7.2). Largest push kept off the bumpers 130 for the cascade, 229 °/s for the LQR (Table 7.6).

to 0.157/s once the 10 ms hold was counted. On the line of cars the linear peak rose only from 1.017 to 1.084 over the range of driver delay in which the measured ratio reached 1.29 and the cars collided (Table 14.5). Table 1.1 showed the same disagreement on one pendulum. The margin worth quoting is the one measured on the mode and the path that fail.

Derivative information holds the plant, and its source limits it. The ball under the magnet and the ball on the beam were held by derivative action that had to act faster than a lag. The levitated ball needs a derivative time above the coil's 10 ms, and no integral setting replaces it. The same derivative differentiates noise. Unfiltered, it flung the ball on the beam 66 to 94 mm off its setpoint. On the double pendulum, rates differenced from encoder counts made the force chatter by 7.17 N, and a Kalman filter on the same counts left 0.43 N. The velocity a stabilizing loop uses is a design choice as much as its gain.

Table 15.2 What Held Each Plant and by How Much, Chapters 8 to 14

Chapter	Difficulty	What held it	What failed	Margin measured
8, double pendulum	Two unstable modes, the fast one an e-fold every 90 ms, and one motor.	An LQR on a Kalman filter.	A loop on each link. Its gains may sum to 12.26 N/rad, the lower link needs 58.6 N/rad. Differenced rates, 7.17 N of chatter against 0.43.	Longest stable sample interval 12.3 ms at a 30 ms drive lag (Figure 8.2). After a switch, 12 ms holds and 13 ms falls (Table 8.5).
9, bicycle	Self-stable only inside a speed window. Countersteering is a wrong-way zero.	A lean PD scheduled on speed. An LQR scheduled on speed.	Gains tuned at 2 m/s, unstable above 3.45 m/s once the 10 ms hold counts. Any of 123 designs at 0.3 m/s.	The PD schedule leaves a band from 2.81 to 4.29 m/s growing at up to 0.157/s (Figure 9.2). LQR roots below −2.76/s from 0.5 to 10 m/s (Section 9.5).
10, drifting car	Roots of 1.95 and 0.21/s at an equilibrium with the rear tire saturated.	A two-input LQR on steer and drive, scheduled on a friction estimate.	A grip controller, spun at 3.61 s. The LQR scheduled for the dry pad, spun on a wet patch at 10.06 s.	Untouched, 5° off in 2.98 s (Section 10.6). A 10° kick back within 2° in 0.92 s (Table 10.4).
11, quarter car	One force on two masses fixes the body at 11.25 Hz, near the 11.9 Hz wheel hop.	A skyhook and groundhook blend on an active strut. Road preview.	A semi-active strut set for comfort, off the road 60 times in 5 s. Any strut on a pothole edge, 1579 m/s^2 asked of 98.1 m/s^2.	Hump rise 36.7 against 47.7 mm, 5.9 with preview (Table 11.4). Leveling gain decays at 11, grows at 12 N/mm (Section 11.5).
12, gantry crane	A swing with a damping ratio of 0.0579 and a frequency set by the cable. One drive.	Two or three impulses shaping the reference of a PID position loop.	A shaper set for 4 m on an 8 m cable. A saturated drive, 7.7° with shaping on or off.	Residual swing 7.87 unshaped, 0.30° with three impulses (Table 12.4). Mistuned, 44.6 % of unshaped, 43.7 % predicted (Figure 12.2).
13, end mill	A regenerative delay turns an 800 Hz tool mode into chatter.	An adaptive feed for the load, a chatter watch that moves the spindle speed.	The feed loop alone on a worn tool, load within 1.6 % and 72 mm of marks.	Limit 2.10 mm, 1.5 mm quiet, 3 mm chatters at 823 Hz. Pocket 12.63 against 16.36 s with 26 mm marked (Table 13.4).
14, line of cars	Each car stable, the line amplifying. A radar latency and a driveline lag.	A time gap above $\sqrt{2/k_p}$, 1.05 s. The predecessor's acceleration over a radio.	Constant spacing, best of 90 gain pairs 1.050. Radar only at 0.6 s, peak 1.306, car 10 down to 36.3 km/h.	Peak 1.000 at 1920 vehicles an hour by radar, at 3692 by radio (Table 14.5). Radio boundary 0.737 s at 400 ms and 20 % loss (Figure 14.2).

Structure decides before tuning on underactuated plants. The cart's angle loop alone balanced the pole and drove the cart into a bumper. On the double pendulum no pair of loops can hold the plant at any tuning, because the sum the cart allows is smaller than the gain the lower link needs. The drift's grip controller spins the car at every request of 10° or more. A full-state regulator held all three, and on the cart it kept a push of 229 °/s off the bumpers where the cascade managed 130. When one actuator serves two outputs, the first question is which structure can hold the plant at all.

Some limits belong to the plant, and only a change to the plant or its command moves them. No strut follows a pothole edge that asks 1579 m/s^2 of a wheel with 98.1 m/s^2 available from its spring and weight. No tuning stops a flicked ball on a beam limited to 3°. No feed setting moves the stability lobe of Chapter 13, while the spindle speed does. The crane's swing was removed by

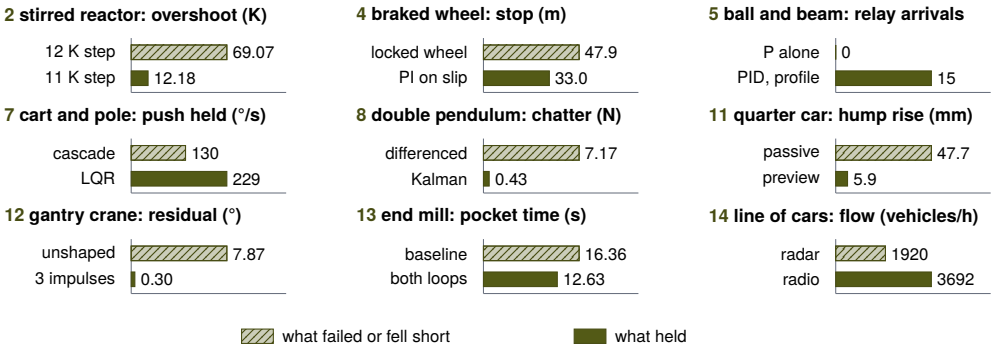


Figure 15.1 One like pair of numbers from nine rows of Tables 15.1 and 15.2, digit for digit, each panel on its own scale. The hatched bar is the setting or structure that failed or fell short, the solid bar the one that held. Higher is better on the ball and beam, the cart and pole and the line of cars, lower on the other six. The aeropendulum, the levitated ball, the bicycle and the drifting car have no like pair in their rows, because each reports an edge or a single margin. Notice that every better number came with a change of structure, command or setting, not with a retuned gain alone.

shaping the command, not by feedback, and a saturated drive gave the same 7.7° with shaping on or off. These are the limits of Section 1.4.7 in their physical form. The engineer's lever is the command, the operating point or the equipment.

15.2 Choosing a Method

The chapters suggest an order of questions for a plant that must be held or kept from ringing. Each names the chapter where its answer decided the result.

1. *Is the operating point stable, and is it the only one?* Find the steady states and the eigenvalues at the chosen one before choosing a controller (Section 1.4.2). The reactor's setpoint sits on an unstable branch, and a stable-looking hot state was an unstable focus.

2. *Can the actuator turn the unstable mode back?* Compare the largest disturbance with $u_{\max}/p$ and check $p\,\tau_d < 1$ (Section 1.4.7). The levitated ball at 28 mm spent 93 % of its current ceiling on a 1 mm drop.

3. *Is a lag inside the loop longer than the derivative time?* Then no gain helps (Equation (1.9)). Shorten the lag, lengthen the derivative time, or filter less, as Chapter 6 and Chapter 5 show.

4. *Does the gain move or change sign over the range?* Use feedforward of the holding input and a schedule, or feedback linearization (Chapter 3). At a peak where the gain vanishes, regulate below it with a margin (Chapter 4).

5. *Is the plant underactuated or nonminimum phase?* Try a cascade with the inner loop faster than $2p$ and the outer slower than $z/2$. If no loop pair satisfies both, as on the double pendulum, design on the full state with an observer.

6. *Does a speed or a friction move the poles?* Schedule on it and check the sampled loop at every point of the schedule, not only the continuous one (Chapter 9, Chapter 10).

7. *Is the problem a resonance the command excites?* Shape the command when the frequency is known to within the shaper's tolerance, which the crane's 3.3 % error already reached. Use preview when the disturbance can be seen coming (Chapter 11).

8. *Is the limit a delay inside the process, such as the regenerative cut?* Move the operating point out of the unstable region (Chapter 13).

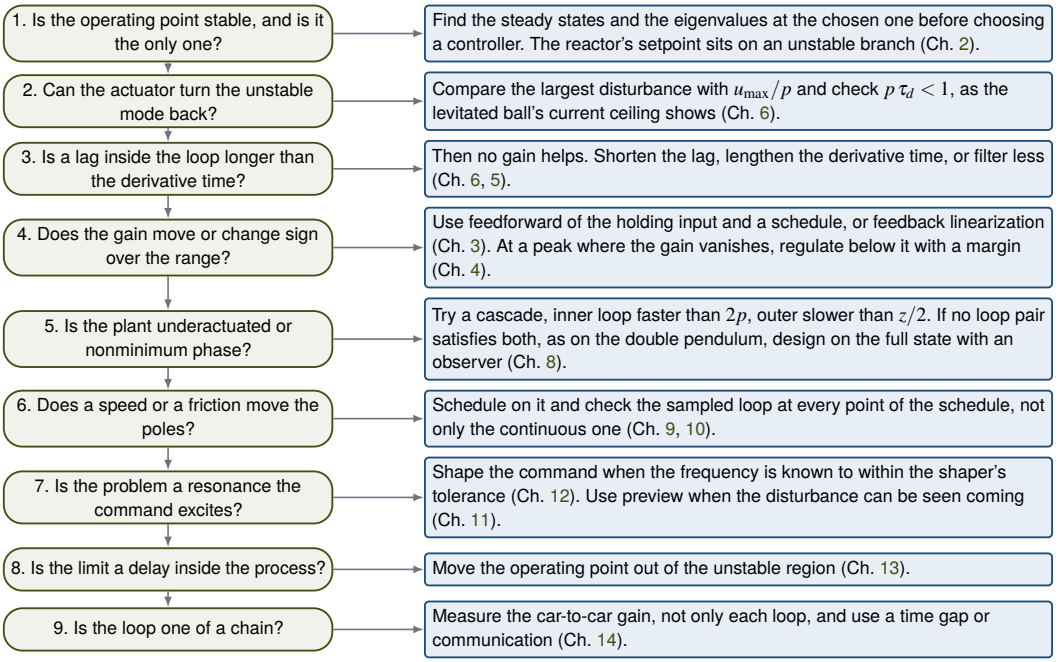


Figure 15.2 The nine questions of this section in order, with the list's answer and the chapters where it decided the result. Notice that the operating point and the actuator's reach are checked before any structure is chosen: the reactor's setpoint sits on an unstable branch, and the levitated ball spent most of its current ceiling on one small drop.

9. *Is the loop one of a chain?* Measure the car-to-car gain, not only each loop, and use a time gap or communication (Chapter 14).

A predictive controller earned its place where a limit bound during the transient. On the reactor it raised the largest safe setpoint step from 11.5 to 13.5 K (Section 2.8). On the cart it stopped a kick at 0.768 m where the regulator crossed the line to 0.809 m (Table 7.7), and on the crane it quieted the standard move 2.16 s sooner than three impulses (Table 12.5). Away from the limits it added little or lost. On the levitated ball the same law clipped without look-ahead held the headline step at 0.325 mm s against 0.319 mm s (Table 6.6). On the line of cars the retuned classical law reached 0.975 against 0.980 (Table 14.6). The double pendulum's MPC fell in five cases its regulator held (Table 8.6). Volume 9 treats the choice in full. Figure 15.2 draws the nine questions in order.

15.3 Where Practice Is Heading

Stabilization beyond the handling limit. A controller has drifted a modified car along a path of varying curvature, speed and sideslip, derived without reference to a nearby equilibrium. It used the yaw acceleration to stabilize the sideslip, and a test run without wheel-speed control showed that the dynamics of the saturated rear tire must be accounted for [1]. Chapter 10 held its car about one equilibrium at a time with a schedule. A review of suspension control finds semi-active and active suspensions implemented by industry and road preview becoming viable for production vehicles [2], the structure that cut the hump rise of Chapter 11 from 36.7 to 5.9 mm. In the editor's view the vehicle chapters of this volume describe functions that are moving from test tracks into series production, and the work moves with them to proving the margin over the whole envelope of road and tire.

String stability leaves the simulator. On a circuit with no bottleneck, a uniform flow of cars broke into a jam that traveled backward like a wave once the density passed a critical value [3]. Later experiments on a circular track with more than 20 vehicles damped such waves by controlling the velocity of a single vehicle, and the authors suggest that a few mobile actuators, less than 5 % of the traffic, could control the flow [4]. A review of platoon control shows that the definitions and analysis methods of string stability differ across the literature and sets out the equivalences among them [5]. The ring of 22 drivers in Chapter 14, whose first car stopped 98.7 s after a uniform start, is that experiment in a browser.

Chatter is still the limit of the machine. A review of chatter research since the stability laws of the 1950s calls chatter the main obstacle to raising the productivity and quality of machined parts, and it covers process damping and cutters with variable pitch designed to suppress it [6]. The simulator of Chapter 13 has neither, which is why its lobes are pessimistic at low speed. Command shaping for flexible systems, the method of Chapter 12, has a published history of more than fifty years [7].

Sensors that come closer to the state. Two chapters were limited by how the state reached the controller. The levitated ball is read by an infrared beam that it partly blocks, a sensor separate from the coil, and the camera on the ball and beam cost that loop 14.4° of phase margin. A review of self-sensing magnetic bearings, which estimate the gap from the coil's own inductance, finds them no longer chiefly a research concept but in commercial use in turbomolecular pumps and elevator guideways [8]. A survey of event cameras reports a latency of the order of microseconds and less motion blur than a frame camera, and finds that new methods are needed to process their output [9]. In the editor's view, a sensor that removes a lag lets a stabilizing loop carry more derivative action before noise or delay stops it.

Electric actuators at the brake and the handlebar. In aircraft braking, electromechanically actuated brakes have found wide acceptance as an alternative to hydraulic ones, and advanced antiskid controllers use them [10]. The United States light-vehicle brake standard requires that where an electric vehicle's regenerative braking is part of the service brake, the antilock system control it [11]. A steering motor added to a consumer electric bicycle stabilized the riderless bicycle down to just below 4 km/h, and with the assistance on, 26 riders on a treadmill at 6 km/h recovered from handlebar pushes significantly more often [12]. The slip loop of Chapter 4 and the scheduled steer of Chapter 9 are the laws such actuators carry.

Feedback from the load, not only from the drive. An inertial unit mounted on the load closed an anti-sway feedback loop on a five-ton industrial overhead crane. It damped the first swing mode while the setpoint stood at zero, and once the hoist moved it kept that mode from being excited, at the cost of slower tracking of the trolley's velocity setpoint [13]. The crane of Chapter 12 removes its swing by shaping the command alone, which works only while the cable length is known to within the shaper's tolerance. Figure 15.3 sets these developments and those above in three threads.

Optimization and learning wrap a stabilizing core. Fast solvers for embedded optimal control are available as open-source software [19]. A review of learning-based predictive control finds most work learning the prediction model from recorded data, and it describes predictive control used to add constraint satisfaction to controllers that were learned [20]. Control barrier functions verify and enforce safety properties in optimization-based controllers [21]. In the editor's view the plants of this volume will keep a linear stabilizing law at their core, a PD, a cascade or an LQR, with optimization above it to respect limits. The chapters point the same way, since the predictive controllers paid where a limit bound and not elsewhere.

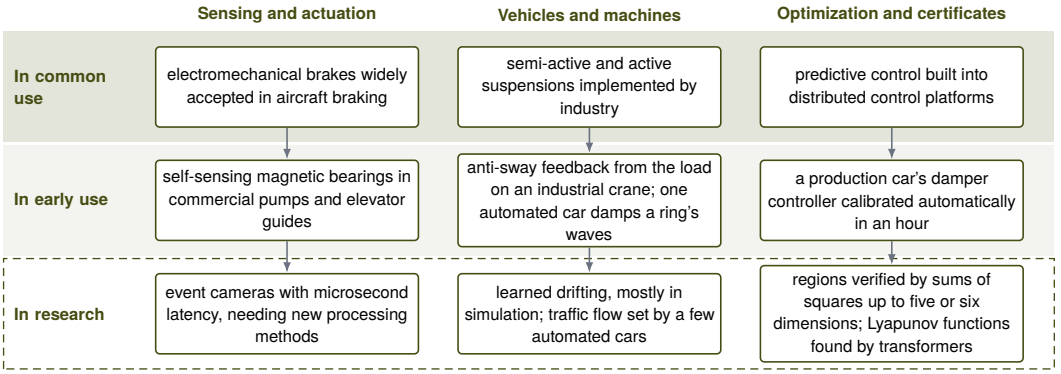


Figure 15.3 Developments in three threads. The top lane holds what the cited source calls widely accepted, implemented by industry or a major change in industrial practice, the middle lane what it reports in commercial use in a few applications or tested on production or industrial equipment, and the bottom lane what it reports from simulation or names as a need, in the editor's reading. Sources by thread: sensing and actuation [8–10], vehicles and machines [2, 4, 13, 14] and optimization and certificates [15–18]. No date is drawn because the sources give none for these lanes. Notice that each thread ends in research that must still prove a margin on hardware.

Certificates and designs proposed by machines. Sequence-to-sequence transformers trained on synthetic samples have found Lyapunov functions for polynomial systems more often than algorithmic solvers and people, and have found new ones for nonpolynomial systems [18]. Agents built on language models tune controllers iteratively against stated requirements on stability, performance and robustness [22]. In the editor's view a certificate or a stabilizing gain proposed by software is a claim that a computation can check, and this volume shows which computation. Check the eigenvalues of the sampled loop, the margin of the mode that fails and the response from the edge of the region it claims.

> ⧖ HISTORY **2010 and 2022: From Local Laws to Regions With Proofs**
>
> The history boxes of Chapter 1 end with a survey of 2000 that left open how fast a schedule of local designs may move. In 2010 a method built a tree of trajectories, each held by a linear quadratic regulator, and computed the region each one recovers by convex optimization. Its authors noted that the weakness of the older picture of funnels had been the difficulty of computing, or estimating by trial and error, the region of each local controller [17]. A review of 2022 found the region of attraction used in learning-based control as a means to guarantee safety, the set of states from which the controller drives the system back [23]. In the editor's reading the thirteen chapters stand at that point. Most of their stabilizing laws are linear about one point, and the regions they report were found by runs of the simulator, not by a proof.

15.4 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 [24]. In the editor's view, agents can now carry out much of the work that used to fill the hours of a control project. They derive a model from balances or fit one to plant data. They write the simulator and its tests, code the controller, tune it against stated criteria, script commissioning checks, write the documentation and watch running loops for degradation. The apps and chapters of this series were produced that way under the editor's direction, and the account that follows this section reports what the agents got wrong in this volume and how it was caught.

In the editor's view, this changes where an engineer's value lies. When a model, a controller and a report can each be produced in minutes, producing them is no longer the scarce skill. The work moves to five things.

Quality control. An agent's output arrives as a claim: this model fits, this tuning settles in four minutes, this code implements the equation. The engineer checks claims against the plant or the simulator. A reported number is measured again on a fresh run at the stated seed. The code is read for what it clamps, what it initializes and what it reads, because a display that shows the controller's model in place of the plant looks correct until the two differ. A tuning is tried on a disturbance the agent did not test. The most useful checks are independent of the author, made by a second agent or a person who has not seen the first answer, and they are checks that can fail.

Supervision. The engineer sets the scope of each task and decides what an agent may touch. Reading a historian and proposing a tuning is one level of access. Writing a setpoint, downloading a configuration or changing an interlock is another, and it passes through the plant's management of change with a named approver. Supervision also fixes the order of work: a baseline before a change, one change at a time and a comparison that shows what moved.

Creativity. An agent answers the question it is given. Choosing the question remains the engineer's task: which loop is worth improving, whether a disturbance calls for feedforward, a cascade or a change to the process, which step test the operation can tolerate, and which comparison would show that a predictive controller earns its place over a retuned PID loop. The experiment that separates two explanations is designed by someone who can imagine both.

Process insight. A model contains what its author put in. The equipment does more. A valve sticks, the dead time of an analyzer grows when its sample flow drops, a thermowell lags, a heat exchanger fouls and an operator intervenes at the same hour each night. The engineer who knows the unit recognizes a result that is too good, asks what the model left out and knows which measurement to distrust first.

Accountability. Responsibility does not transfer to a tool. Functional safety standards assign the activities of the safety lifecycle to persons and organizations of stated competence, and they require an assessment that includes someone independent of the design team [25]. 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.

15.4.1 How This Volume Was Made

Each of the thirteen apps was built and tested by agents. An agent then wrote the paper on it, keeping verification notes that trace every number to a run, and a referee report was written on each paper. At least five of the thirteen reports are self-reviews by the agent that wrote the paper. Defects went into a log for each app with a severity, a cause, a suggested fix and a status. For this book, one agent per chapter reran every number an exercise uses. The editor set the scope of each round and what each agent could change. The record holds the following errors and the checks that caught them.

The defect logs. The thirteen logs hold 144 entries with a severity. Of these, 32 are wrong behavior of an app (a wrong output, a wrong lesson or a controller reading what it should not), 85 are documentation, 22 are cosmetic or display, and 5 concern completion rules or the interface. The cases below were chosen for this volume's topic. Each is a claim about stability, an equilibrium or a margin.

A stable state that was not. The reactor's help text said that at a jacket temperature of 300 K the reactor could ignite to a second, stable state at 369.7 K. While writing the paper, an agent computed the Jacobian there and found eigenvalues 1.357 ± 1.540j, an unstable focus. Released from 369.7 K, or from anywhere up to 380 K, the run returned to 324.475 K every time. The same log records a footnote that named the nearest root of the energy balance as the temperature the reactor would settle at, without examining its stability, and a tip that advised lowering the gain against a runaway. Measured, the lower gain raised the peak from 451.44 to 469.88 K.

A stability window computed badly. The first gain window of the cart and pole was computed by power iteration. It produced a floor that wandered from 18.3 to 18.9 N/rad and three ceilings the plant cannot have. A first grid of nonlinear runs, read too early, labeled slowly settling loops as rocking. In review the boundaries were recomputed from the spectral radius of the sampled loop, checked against exact eigenvalues, and the window was set against 40 nonlinear runs, which agreed at 39. The same review found that numbers lifted from the app's guide were wrong, among them a gain of 15 N/rad called "below the floor" of 11.3 N/rad.

A design claim that the plant did not support. Comments in the suspension app said that a stiff proportional leveling gain made the body unstable. In the self-review the agent tested it. At the default comfort setting no gain up to the slider's 50 N/mm destabilized the body. At 0 % comfort it did, between 11 and 12 N/mm, through a mechanism the review then derived. The first draft had also given both passive modes a damping ratio near 0.32 from uncoupled formulas, where the eigenvalues give 0.25 and 0.33.

An equilibrium on the wrong branch. The drift controller designs its LQR about an equilibrium that a solver finds for the requested speed and sideslip. While writing the paper, the agent swept every target a learner can set. On the dry pad 89 of 1,485 returned a yaw rate of the wrong sign, a car turning against its own drift, and the regulator would have been designed about that point. None of the 9,090 targets the shipped cases reach was affected. The fix accepts only a root with the right sign and adds a test over 5,940 targets.

Controllers that read the plant. Three logs record a controller that took values a real controller could not have. The levitation PID computed its holding current from the run's own ball. A 10 % heavier ball scored an IAE of 0.108 mm s when it came through the disturbance slider and 0.008 mm s when it came through the parameter, which the controller read. The braking controller's low-speed exit read the true road. On identical samples, a change from dry asphalt to ice moved its held torque from 1336.9 to 200.0 N m. The fix moved every run that reaches the exit, 10 of the 13 scenarios, and removed wheel locks, among them 0.15 and 0.23 s under the PI loop. The reactor's controllers started from the run's balances, and 10 of 11 edits made before Run moved the controller's start while no run showed it. On the double pendulum, a test meant to catch such reads in the predictive controller missed seven of eight versions doctored to read the truth, and it was rewritten until it caught every doctored version tried. These were found by audits that asked what each controller reads, and by tests built to fail.

A headline chosen after the run. The levitation paper's comparison step was chosen from twelve candidates after both controllers had run on them. The referee asked whether the headline was fair. The paper now says how the step was chosen and carries the counterfactuals. The predictive law clipped without look-ahead held the same step, so state feedback, not the optimization, won it, and the chapter states the result that way.

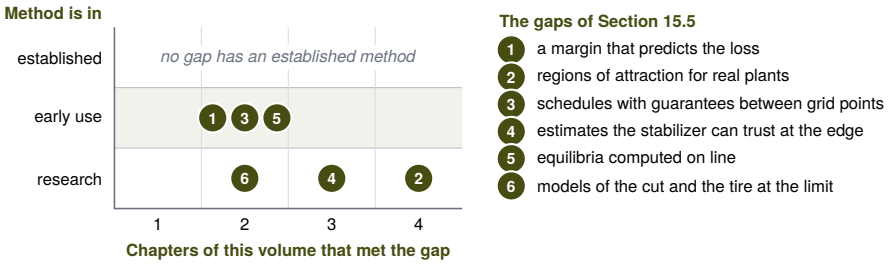


Figure 15.4 The six gaps of this section. Across is the number of chapters that each item names as evidence: two for gap 1 (Chapters 3 and 14), four for gap 2 (7, 8, 2 and 5), two for gap 3 (9 and 10), three for gap 4 (4, 10 and 8), two for gap 5 (10 and 2) and two for gap 6 (13 and 10). Up is the maturity of the method available, the editor's reading of the sources cited in each item and in the Open challenge box. Notice that the gap met by the most chapters, the region a controller can recover, is still in research.

The book round. The thirteen chapter records list 215 rows of numeric verification, each a stated value beside its reproduction. No reported value drifted beyond a rounding (a minimum gain printed 1.70, since corrected to 1.69) and three counts of the apps' own checks, which have grown since the papers. The bicycle agent found a briefing step that a student could follow faithfully and learn the opposite lesson from. The "hands-off" run kept the loop's last torque and fell in 0.55 s. It then appeared to fall sooner than the wrong-signed loop's 0.61 s, where a true hands-off run takes 1.66 s. The chapter's exercise tells the reader which button to press, and the briefing was later corrected. The book round's defect list held 10 such app or paper items for this volume. A later repair pass fixed eight, confirmed one as printed and kept one, a count of checks, as a historical value.

The pattern is the one Section 15.4 describes. Every error above was caught by a check that could fail, such as eigenvalues computed at a claimed equilibrium, a sweep over every target, a nonlinear run set against a linear window, or a hidden value moved to see whether the controller moved. For a stabilizing loop the check matters more, because a held plant hides what its controller knows. Four working rules follow. Compute the eigenvalues of every state the text calls stable. Sweep the whole range a user can reach. Read what a controller reads. Let a person decide what changes.

15.5 Gaps in Nonlinear Stabilization

Six gaps belong to this volume's topic, and the gaps common to agent-era practice follow them in Section 15.5.1. 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 15.4 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. *A margin that predicts the loss.* The aeropendulum lost a slow mode behind a healthy phase margin, and the line of cars collided at a linear peak of 1.031. A textbook notes that the gain and phase margins must both be given to show that a loop stays clear of the critical point, and it defines the stability margin as the shortest distance to it [26]. A review of platoon control finds that the definitions and analysis methods of string stability differ across the literature [5]. The gap is hard because each of these numbers describes one linear loop at one point, while the chapters lost a slow mode, a sampled loop or a chain of saturating drivers. A single robustness number that ranks designs by the disturbance, delay or parameter change that actually loses a nonlinear, sampled, saturating plant is not in routine use.

2. *Regions of attraction for real plants.* The cart and pole's largest recoverable push, the double pendulum's catchable band and the reactor's largest safe step were all found by search on a

simulator, and on the ball and beam no tuning keeps a flicked ball on a beam limited to 3°. The Open challenge box after this list takes the gap further.

3. *Scheduling with guarantees between the grid points.* The bicycle's schedule left a band of growth between design speeds, and the drift regulator scheduled for dry tires spun on a wet patch. A survey of gain scheduling found the relaxation of the slow variation restriction still very much an open question, judged the robustness analysis of gain-scheduled controllers very immature, and named soundly based interpolation between operating points and help in choosing the scheduling variable as pressing needs [27]. The gap is hard because each design point is linear and stable while the loop as it moves is neither. Guarantees for the scheduled loop as the parameter moves, checked on the sampled loop between the grid points, are thin for plants like these.

4. *Estimating what the stabilizer needs.* The braking law needs the friction peak, the drift schedule a friction estimate, and the double pendulum rates that the encoder does not give. A study of antilock braking reports that uncertainty in the road and vehicle states is often neglected or handled apart from the controller, which leads to sub-optimal braking, and that the friction peak is harder to find on snow and ice [28]. For aircraft braking, friction estimation is called an extremely challenging and still open task [10]. The gap is hard because the quantity is visible only near the edge the loop must avoid. An estimator that is fast enough to stabilize with and honest about its error at the edge of the operating region is an open design problem.

5. *Equilibria computed on line.* The reactor's ignited state at the published jacket temperature is an unstable focus, not a second place to operate, and at low speed and small sideslip the drift's equations admit a mirror-image root that turns the wrong way, which its solver rejects by the sign of the yaw rate. A drifting controller on a passenger car took its equilibria from maps computed offline, with an initial guess that steers the solver to the drifting equilibrium it wants [29]. The gap is hard because a solver that converges says nothing about which branch it found or whether it is stable. Controllers that compute their own operating point need a routine check of branch, sign and stability before they act on it.

6. *Models of the machine at the cut and the tire at the limit.* The milling model has no process damping, and the drift tire's friction does not fall as it slides. A review of chatter finds that stability predictions still suffer from uncertain measurements, time-varying dynamics and a worn tool, and asks for physics-based prediction made off line to be joined with on-line adaptive learning and tuning [6]. A drift controller tested on a sports car names tire model fidelity and friction estimation as what still limits it [14]. The gap is hard because the limits that matter most in this volume sit where models are least certain. Identifying them safely from the machine is still research.

OPEN CHALLENGE Certifying the Region a Controller Can Recover

Gap 2 runs through four chapters. Three found a recoverable push, band or step by search, and in the fourth no tuning catches a flick. Verifying regions by sums of squares has succeeded on systems of up to five dimensions, and checking by sampling costs less but loses the guarantee [17]. With a bounded input, a saturated feedback's basin of attraction lies inside the set of states from which some admissible input can reach the origin [30]. Guarantees for learned controllers rest on assumptions hard to verify before a robot runs [23]. In the editor's view the region is hard to certify because it belongs to the whole saturating loop. Progress would be a certified region that a run from its edge confirms.

15.5.1 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 [23]. 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 [25]. 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 [31]. 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 [32]. 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 [33]. 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

- Each machine could be pushed only so far, and past that edge it was lost at once: the distance to the edge is the number to report.
- The usual margin did not always see the edge. The margin worth quoting is the one measured on the mode and the path that fail.
- On underactuated plants the structure decided before the tuning, and the source of the derivative information limited what the loop could hold.
- Every error in the record was caught by a check that could fail, such as eigenvalues at a claimed equilibrium or a sweep over every target a user can reach.

References

[1] J. Y. Goh, T. Goel, and J. C. Gerdes, "Toward automated vehicle control beyond the stability limits: Drifting along a general path," *Journal of Dynamic Systems, Measurement, and Control*, vol. 142, no. 2, p. 021004, 2020.

[2] J. Theunissen, A. Tota, P. Gruber, M. Dhaens, and A. Sorniotti, "Preview-based techniques for vehicle suspension control: A state-of-the-art review," *Annual Reviews in Control*, vol. 51, pp. 206–235, 2021.

[3] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks: Experimental evidence for the physical mechanism of the formation of a jam," *New Journal of Physics*, vol. 10, no. 3, p. 033001, 2008.

[4] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.

[5] S. Feng, Y. Zhang, S. E. Li, Z. Cao, H. X. Liu, and L. Li, "String stability for vehicular platoon control: Definitions and analysis methods," *Annual Reviews in Control*, vol. 47, pp. 81–97, 2019.

[6] Y. Altintas, G. Stepan, E. Budak, T. Schmitz, and Z. M. Kilic, "Chatter stability of machining operations," *Journal of Manufacturing Science and Engineering*, vol. 142, no. 11, p. 110801, 2020.

[7] 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.

[8] E. H. Maslen, "Self-sensing for active magnetic bearings: Overview and status," in *Proceedings of ISMB10, International Symposium on Magnetic Bearings*, 2006. [Online]. Available: https://www.magneticbearings.org/app/uploads/publications/ismb10/ismb10_submission_key03.pdf

[9] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.

[10] F. Crocetti, G. Costante, M. L. Fravolini, and P. Valigi, "Tire-road friction estimation and uncertainty assessment to improve electric aircraft braking system," in *29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 330–335.

[11] National Highway Traffic Safety Administration, "Standard no. 135; light vehicle brake systems, 49 CFR 571.135," Code of Federal Regulations, Title 49, Chapter V, edition of October 1, 2024, U.S. Government Publishing Office, 2024. [Online]. Available: https://www.govinfo.gov/content/pkg/CFR-2024-title49-vol6/pdf/CFR-2024-title49-vol6-sec571-135.pdf

[12] M. T. Haitjema, L. Alizadehsaravi, and J. K. Moore, "Automatic bicycle balance assistance reduces probability of falling at low speeds when subjected to handlebar perturbations," *IATSS Research*, vol. 50, no. 1, pp. 701–711, 2026.

[13] V. Helma and M. Goubej, "Active anti-sway crane control using partial state feedback from inertial sensor," in *2021 23rd International Conference on Process Control (PC)*. IEEE, 2021, pp. 137–142.

[14] F. Werner, T. Heintzenberg, M. Lienkamp, and J. Betz, "Drifting in the future: Stabilizing path following drifting on high-latency vehicle systems," in *2026 IEEE International Conference on Robotics and Automation (ICRA)*, 2026, pp. 12 763–12 770.

[15] 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.

[16] G. Savaia, Y. Sohn, S. Formentin, G. Panzani, M. Corno, and S. M. Savaresi, "Experimental automatic calibration of a semi-active suspension controller via Bayesian optimization," *Control Engineering Practice*, vol. 112, p. 104826, 2021.

[17] R. Tedrake, I. R. Manchester, M. Tobenkin, and J. W. Roberts, "LQR-trees: Feedback motion planning via sums-of-squares verification," *The International Journal of Robotics Research*, vol. 29, no. 8, pp. 1038–1052, 2010.

[18] A. Alfarano, F. Charton, and A. Hayat, "Global Lyapunov functions: A long-standing open problem in mathematics, with symbolic transformers," arXiv preprint arXiv:2410.08304, 2024.

[19] R. Verschueren, G. Frison, D. Kouzoupis, J. Frey, N. van Duijkeren, A. Zanelli, B. Novoselnik, T. Albin, R. Quirynen, and M. Diehl, "acados: A modular open-source framework for fast embedded optimal control," *Mathematical Programming Computation*, vol. 14, no. 1, pp. 147–183, 2022.

[20] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 269–296, 2020.

[21] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," in *Proc. 18th European Control Conference (ECC)*, 2019, pp. 3420–3431.

[22] X. Guo, D. Keivan, U. Syed, L. Qin, H. Zhang, G. Dullerud, P. Seiler, and B. Hu, "ControlAgent: Automating control system design via novel integration of LLM agents and domain expertise," arXiv preprint arXiv:2410.19811, 2024.

[23] 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.

[24] 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.

[25] 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.

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

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

[28] B. Sullivan, J. Jiang, G. Mavros, and W.-H. Chen, "An exploration-exploitation approach to anti-lock brake systems," arXiv preprint 2306.14730 [eess.SY], 2023. [Online]. Available: https://arxiv.org/abs/2306.14730

[29] S. Meijer, A. Bertipaglia, and B. Shyrokau, "A nonlinear model predictive control for automated drifting with a standard passenger vehicle," in *2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, 2024, pp. 284–289.

[30] Y. Aoustin and A. Formal'skii, "Ball on a beam: stabilization under saturated input control with large basin of attraction," *Multibody System Dynamics*, vol. 21, no. 1, pp. 71–89, 2009.

[31] 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.

[32] 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.

[33] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

15.6 Exercises

15.1 Discussion · 8 min

An agent hands you a PID tuning for a new propeller arm with the report "phase margin 50 to 55° at every angle from 15 to 105°, so the loop is robust over the range." Using Chapter 3 and Table 1.1, say why that report is not enough, and list what you would ask the agent to compute or run before the tuning is used.

15.2 Discussion · 7 min

On the double pendulum of Chapter 8 a loop on each link beside a cart loop cannot hold the plant at any tuning. A colleague proposes to let an agent search the six PID gains overnight on the simulator. Explain why the search cannot succeed, what result it would return, and how you would redirect the work.

15.3 Discussion · 8 min

Section 15.1 says that some limits belong to the plant. Take two of the following and, for each, name the limit, say why no feedback controller of the chapter removes it, and say what does: the pothole edge of Chapter 11, the flicked ball of Chapter 5, the stability lobe of Chapter 13, the crane swing of Chapter 12 with a saturated drive. Then name one plant from your own field with a limit of the same kind.

15.4 Discussion · 8 min

An agent proposes replacing the LQR of a balancing machine with a predictive controller, citing a case where the predictive controller won. Using the results gathered in Section 15.2, design a comparison that would convince you, and say what result would make you keep the LQR.

15.5 Discussion · 6 min

Section 15.4.1 reports that the levitation PID computed its holding current from the run's own ball, and that no shipped run showed it. Explain why a controller that reads the plant's true parameters is harder to detect on an unstable plant than on a stable one, and describe a probe that would detect it in an agent-written controller.

15.6 Discussion · 7 min

The drift solver of Chapter 10 once returned equilibria on the wrong branch for 89 of 1,485 targets, and the reactor's help text once called an unstable focus a stable operating point. Write a short checklist that a person or an agent should apply to any computed operating point before a stabilizing controller is designed about it, and say which item would have caught each of the two errors.

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 60° and 120°. Gains $+0.5$ and -0.5 rad per rad/s^2. Eigenvalues $\pm 1.41\,\mathrm{j}$ at 60° and ± 1.41 at 120°.

1.3 (a) $s^2 + 4s + 6$, $\omega_n = 2.45\,\mathrm{rad/s}$, $\zeta = 0.816$. (b) 120°. (c) $k_p = 14$.

1.5 $L = [20,\ 104]$. Eigenvalues -2, -3, -10 and -10.

1.7 (a) 0.4 s, with a pair at $\pm 4\,\mathrm{j}$. (b) 0.5 s for proportional feedback, 0.347 s for the relay.

1.9 1 s. At 0.6 s the coefficient is $1 - 0.36 = 0.64 > 0$, so slow waves grow from car to car.

Chapter 2. Exothermic Reactor at an Ignition Boundary

2.1 (a). The jacket enters only the energy balance, so the first response to a jacket move is positive everywhere, and the 0.05 min control period is far shorter than the residence time.

2.3 (c). After the 3 min hand-over the IAE falls from 54.7 to 49.6 K min, while after the 4 min one it rises from 6.28 to 6.47 K min.

2.5 (a) $K_c > 0.267\,\mathrm{K/K}$. (b) $s^3 + (2.0921K_c - 0.5595)s^2 + (4.664K_c - 0.5567)s + 2.321K_c$, stable for $K_c > 0.57\,\mathrm{K/K}$. (c) It is stable at 340 K, but the minimum gain grows along the middle branch, so an upset that moves the reactor carries it out of the region the gain can hold.

2.7 (a) 328.79 K at 6 min, 328.70 K at 20 min, largest 328.85 K. (b) 2.11 K/K, 63.2 % reached about 0.94 min after the step. (c) The chord gain grows with the step size. A 4 K step passes the ignition fold.

2.9 (a) 15.07 K, 57.94 K min: the reactor ignites. (b) 0.573 K, 2.26 K min. (c) 0.41 K, 1.42 K min.

2.11 (a) 559.2 K and 601.0 K. (b) A feed 14 % richer than design removes the cold state. The alarm has no latch and takes no action on the plant.

Chapter 3. Aeropendulum: The Setpoint Chooses the Plant

3.1 (d). The margin is read near 10 rad/s, and the mode that loses its damping lives near 1.6 rad/s, where the stiffness counts.

3.3 (b). Full thrust, 0.3818 N, is below the 0.3989 N that holds horizontal, so no angle above the 73.16° stall angle can be held.

3.5 (a) 0.3989 N and 72.97 %; 68.46 %, which also holds 120°. (b) 0.3818 N, 4.3 % short, stall at 73.16°.

3.7 (a) 4.25° and 17.94 %. (b) 15.40 s. (c) No hit, ζ 0.14 and 27°: the edge moves to about 140.6°, past the 120° setpoint.

3.9 (a) Met at 2.26 s, 9.2°, 0.43 % and 1.19 %. (b) Chatter 2.32 % breaks the 2 % limit. (c) 3.79 s and 18.2° break the 3.0 s and 12° limits.

3.11 (a) 80.8° and 87.7°; count 187 lifts the arm past horizontal. (b) Count 746 holds 87.0° and count 745 84.6°, a step of 2.4° against 6.9°.

Chapter 4. Anti-Lock Braking at the Friction Peak

4.1 (c). Only the wheel speed is instrumented, and Equation (4.10) adds the noise to it.

4.3 (a). Dry to ice drives slip to 0.655. Wet to dry gives no excursion above the setpoint.

4.5 (a) $1287.8 + 49.1 = 1336.9\,\mathrm{N\,m}$, the wheel term 3.7 %. (b) $\mu' = -0.444$, $p = +4.05/\mathrm{s}$, a time constant near 0.25 s. (c) pv holds no speed, so $p = 4.05 \times 27/5 = 21.9/\mathrm{s}$.

4.7 (a) 32.7 m with no lock, and 32.9 m with 0.27 s locked. (b) 34.8 m with no lock, and 34.1 m with 0.87 s locked. (c) 0.092 and 0.155, both below $\lambda_c = 0.1661$.

4.9 (a) 0.086, nearly three times the band, so the run fails without a lock. Slip reaches only 0.023 before the ice, and the car meets it at 15.43 m/s. (b) 0.0007, a pass, with the car at 8.45 m/s at the ice. (c) The retuned loop unwinds its integral 33 times faster.

4.11 (a) 0.0981 m/s^2, 20 % of the ice value and 0.85 % of the dry value. (b) 54.4 m against 65.2 m.

Chapter 5. Ball and Beam Balance

5.1 (c). A delay τ_e moves the roots of the spring-like loop to $\omega^2\tau_e/2 \pm j\omega$, a growth of $e^{\pi\omega\tau_e}$ a swing.

5.3 (a). At 3° the beam can decelerate the ball at most 0.295 m/s^2, so stopping from 30 cm/s takes 15.3 cm.

5.5 (a) $\omega = 1.830\,\mathrm{rad/s}$, period 3.43 s. (b) 1.45 and 6.7 mm. At 0.15 °/cm, 1.30 and 13.3 mm, about 1.19 for an 11.4 cm swing. (c) $\zeta = 0.64$, $\tau_D = 1.09\,\mathrm{s}$ for $\zeta = 1$, and $\tau_D > \tau_s$.

5.7 (a) 16.41 s, −37.3 cm after −26.3 cm, ×1.42, 3.48 s. (b) ×1.17, 4.88 s. (c) 37.56 s, ×1.25. The period goes as $1/\sqrt{K_c}$, and the fixed loss to the rolling resistance is a smaller share of a wider swing.

5.9 (a) 0 arrivals, BALL OFF at 1.19 s. (b) 15 of 15, target met. (c) 12 of 15, missing legs 10, 12 and 14, the legs to the left mark after the bend.

5.11 (a) 0.49°, 0.25° and 0.16°. (b) Only the shipped dither carries the beam past the breakaway. Dither wears the actuator and shakes the load.

Chapter 6. Magnetic Suspension of an Unstable Ball

6.1 (b). The controller sees only the sensor output $\hat{h}$ and sets the current command i_{cmd}.

6.3 (d). Condition (6.18) needs $\tau_D > \tau_c$ whatever the gain, and 6 ms fails it.

6.5 (a) $s^2 + (BK_c - p^2) = 0$, a constant oscillation at 58.6 rad/s. The PI loop has a zero s^2 coefficient and is never stable. (b) 107.0 A/m, and the default budget is 34.3/s against 2/s. (c) 10.30 ms.

6.7 (a) Loss of capture at 0.17 s, contact at 0.10 s. (b) Caught, from a gap of 21.8 to 23.1 mm, with a peak of 2.4 to 2.9 A. (c) Lost, from about 27 mm.

6.9 (a) Lost at 3.94 s as shipped and at 3.68 s with 1.000 A/mm. (b) Held, worst error 1.01 mm, just outside the band. (c) 0.92 mm with no saturation and a 1.94 A peak, which meets the target. The suggested tuning gives 0.18 mm and 1.87 A.

6.11 (a) The lookup, about 47 µs against 380 µs. (b) With a 100 A/s amplifier the table loses the ball after a 1 mm drop that the PID holds, and a step to 27 mm defeats it while a ten-move MPC holds it.

Chapter 7. Inverted Pendulum Balance on a Finite Rail

7.1 (a). With friction the constant term of Equation (7.9) is $-b_c m\ell g$, so a real positive root exists whatever the gains.

7.3 (c). By Equation (7.19) the zero offsets are optimal, so the first move is $-K\tilde{x}_0$.

7.5 (a) $a = 0.487$, one count is 0.0879°. (b) 0.075 rad/s, 0.37 N through the damping and 0.046 N through the gain. (c) 70 ms or 0.39 of the e-folding time, against about 105 ms or 0.59 where the window closes.

7.7 (a) Fell at 4.54 s, 5 rail hits. (b) 0.128 and 0.520 m/s, a growth rate of 0.47 s^{-1} against the estimate of 0.52 s^{-1}. (c) 0.046 m.

7.9 (a) Fell at 3.04 s. (b) 0.509 m. (c) 0.5 s falls at 7.54 s, 0.6 s holds at 0.793 m. The rig needs a longer horizon than the 0.46 s linear crossing.

7.11 (a) 132 °/s. (b) 0.25 m/s, and 95 °/s with $e = 0$.

Chapter 8. Double Inverted Pendulum Model-Based Balance

8.1 (d). Equation (8.8) caps $K_{p1} + K_{p2}$ near 12.26 N/rad while it asks for more than 58.6 N/rad on the lower link.

8.3 (b). Without feedback the stored force hands over only on the exact rig, and with feedback the same run hands over 2.03 s into the trajectory.

8.5 (a) 90.3 ms, then 1.057, 1.248 and 9.15. (b) 3.86 Hz, far below 25 Hz. (c) 120.8 ms and 36.1 ms. The interval is limited by the lag, not by the resolution of the sampler.

8.7 (a) Fell at 1.61 s, 0.615 s after the release. The upper link is past 2° 0.205 s after the release. (b) Fell at 1.53 s, 0.530 s after the release. (c) Held, with 0.015 m of cart travel and the upper link within 0.64°.

8.9 (a) 5.93 s and 6.55 N, late and noisy. (b) 6.27 s and 0.40 N, quiet but late. (c) 4.81 s and 0.41 N, met. (d) 6.00 s and 0.54 N, still late.

8.11 Payloads up to +27 % are handled. Heavier ones are handed over but not held, and need a balance design for the present links, which requires the payload to be known.

Chapter 9. Riderless Bicycle Balance

9.1 (d). Above $v_c = 6.024\,\mathrm{m/s}$ the capsize root is positive, at $0.143\,\mathrm{s^{-1}}$ at 8 m/s, and every other root decays.

9.3 (b). At 0.3 m/s the grip $gK_{0\varphi\delta} + v^2K_{2\varphi\delta}$ is −18.6 N m/rad against −794 N m/rad of gravity on the lean.

9.5 (a) 0.197 s and 7.8 s. (b) 6.98 s; 20.9 s against 20.2 s measured. (c) 2.96 s, a factor of at most 1.59.

9.7 (a) It falls at 0.61 s, the bars steering away from the lean. (b) 1.66 s hands off. (c) At most 2.8 N m, and within 0.5° from about 0.93 s.

9.9 (a) 2 gates, fall at 17.04 s, 4.9 % saturated. (b) 8 gates, no fall, 0.1 %, target met. (c) The fixed LQR passes (8 gates, 0.1 %). The scheduled PD stays up but passes only 5 gates.

9.11 (a) The passenger is folded into the rear frame, so every matrix the physics integrates changes, while the controller keeps the declared empty bicycle's matrices. (b) Set the passenger to 20 kg, read the window and the closed loop on the Map view, and rerun the cases across speed; if the margin is short, schedule on a measured load or design for the whole load range.

Chapter 10. Autonomous Drifting

10.1 (c). On the circle $F_{yr} = \sqrt{(\mu F_{zr})^2 - F_{xr}^2}$, so every newton of drive is paid for in lateral grip.

10.3 (a). The yaw balance asks the front for $(b/a)\mu_R F_{zr} = \mu_R F_{zf}$, its whole circle, so the longitudinal balance cannot close.

10.5 (a) 0.355 s and 3.36 s; 1.18 s. (b) $\ln 10/1.16 = 1.98/\mathrm{s}$, the slide's 1.95/s. (c) 0.51 s; $p_1\tau_d = 0.39$; about 3.9 rad/s.

10.7 (a) −25.0°, 0.93 rad/s, 3.72 kN; roots +1.95 and +0.21/s. (b) SPUN at 4.29 s; rows at 1.84 s and 3.00 s, about 1.98/s. (c) About 3.16 s.

10.9 (a) SPUN at 32.54 s, 6 cones, score 0. (b) OFF PAD at 46.98 s, 7 cones, both transitions. (c) Score 58, no cone, target met. (d) SPUN at 37.41 s. The LQR holds the slide, the path loop keeps the circle, and the $\hat{\mu}$ schedule carries the drift across the wet patch.

10.11 (a) On grip the force is inside the circle, so force over load is the friction in use, not the friction available. A car that never slides gives this estimator nothing to learn from. (b) It schedules the equilibrium and the gain on the wrong friction until the filtered estimate catches up.

Chapter 11. Active Suspension Comfort and Road Holding

11.1 (b). At ω_1 the wheel drops out of the first line of Equation (11.9), which leaves $-\omega_1^2 M_s Z_s = k_t Z_r$ with no strut term.

11.3 (d). The filtered proportional force is nearly in phase with the body velocity and acts as a damping of about $-K_c/(\omega_b^2\tau_h)$.

11.5 (a) 1579 m/s^2 and 98.7 m/s^2. (b) 98.1 m/s^2, 160.6 m/s^2 with the actuator. (c) 2.99 m/s and 3.83 m/s.

11.7 (a) 0.885 m/s^2 against 1.197, 8.1 % against 11.0, 1 J. (b) 0.345 m/s^2, 4.3 %, 138 J, 61 of 80 mm. (c) With preview the actuator does the work of compressing the spring over each hump.

11.9 (a) −35.9 % with 6 liftoffs, not met. (b) −17.4 %, no liftoff, not met. (c) −28.5 %, no liftoff, met, 191 J.

11.11 (a) Switch the preview off or widen its window when the preview force grows. (b) Biased misreadings, which do not average out.

Chapter 12. Input Shaping on an Overhead Crane

12.1 (a). A shaper convolves the setpoint, and a constant setpoint comes out unchanged, so only the position loop removes the swing.

12.3 (c). Cancellation needs a linear path from command to swing, and a drive on its limit breaks it.

12.5 (a) $\omega_c = 0.678\,\mathrm{rad/s}$, $\zeta_c = 1.00$, a factor of 2.3. (b) $\zeta_c = 0.43$ as loaded, about 22 % overshoot, and $\zeta_c = 1.09$ suggested. (c) 13.2 kN/m, asking 79 kN against a 4 kN rating.

12.7 (a) 14.14°, 2.81°. (b) ZV 6.29°, 0.26°, 2.01 s; ZVD 3.23°, 0.07°, 4.01 s. (c) The coupled mode of trolley and load is 3.3 % faster than $\sqrt{g/\ell}$, just outside the two-impulse tolerance band.

12.9 (a) Position (0.14 m) and sway (4.32°) fail. (b) 0.05 m passes, sway 3.85° and 0.5 % on the limit fail; peak 18.55°, pinned 0.16 s. (c) Peak 4.26°, no time on the limit, largest amplitude 0.16°: all three targets met.

12.11 (a) $\arctan(120/(150\times 9.81)) = 4.662°$; neither, since the loop cannot reach θ and the wind is not a command. (b) $W/K_c = 120/620 = 0.194\,\mathrm{m}$, removed by integral action.

Chapter 13. CNC Milling Stability Lobes and Adaptive Feed

13.1 (a). Equation (13.7) holds the tool, the cutting coefficient, the engagement and the tooth period, and no feed.

13.3 (c). The ring-down is not synchronous with the teeth, and the indicator has no growth test.

13.5 (a) 43.2 $\mathrm{cm^3/min}$ and 0.639 kW. (b) 6.76 kW, 1.26 kW more than the spindle has. (c) 5252 rpm.

13.7 (a) 0.64 kW, no marks. (b) 1.32 kW, 88 mm, 60.1 µm, limit 2.10 mm. (c) 1.27 kW, no marks, 8.8 µm, limit 17.9 mm. The load follows the depth, the chatter follows the limit.

13.9 (a) 3.65 kW, the override swinging between its limits. (b) 3.19 kW, override at least 87.6 %. (c) 2.08 kW. Only removing the delay recovers the peak.

13.11 (a) 512 Hz and 10 240 rpm, near the 10 314 rpm the chatter control found in case 5. (b) Tap-test the tool in its holder to measure its mode. The model has one mode in one fixed direction and no process damping.

Chapter 14. String Stability in a Vehicle Platoon

14.1 (d). String stability is a property of the gain from car to car, which the poles of one follower do not show.

14.3 (b). The RMS ratio is the $\mathscr{L}_2$ gain the criterion bounds, and the ratio of peaks is not bounded by $|\Gamma|_{max}$.

14.5 (a) $0.4s^3 + 1.6s^2 + 2.2s + 2$, residual about 3×10^{-4}. (b) $s^2 + 1.368s + 1.899$, $\omega_n = 1.378\,\mathrm{rad/s}$, $\zeta = 0.496$. (c) 5.61 s: the sine drives the line at its most amplified frequency.

14.7 (a) 0.93 at car 10, 1920 veh/h, car 10 down to 66.8 km/h. (b) 1.0 s: 0.99, 2618 veh/h, 61.2 km/h; 0.8 s: 1.06 at car 6, 3064 veh/h, 54.5 km/h. (c) 1.0 s lies just inside the unstable side of the 1.050 s boundary, and the peak of 1.003 sits at a low frequency where the tap carries little energy.

14.9 (a) Amplification 1.13 at car 10 fails, while 1.75 m and 2283 veh/h pass. (b) 1.04 at car 4, not met. (c) 0.99 at car 8, 1.75 m, 2158 veh/h: met. The run back on ACC repeats 1.13, 1.75 m and 2283 veh/h, so Reset carries no state from the earlier policy.

14.11 (a) $(1766-997)/1766 = 44\,\%$ and $(1920-1880)/1920 = 2\,\%$. (b) Holding a merge short of breakdown protects its flow, because a stopped queue of drivers discharges far below what it carried moving. The drivers are identical, noiseless, one-lane and look one car ahead, so the numbers bracket the declared 10 % without predicting it.

Chapter 15. Summary and Outlook

15.1 A good answer notes that in Chapter 3 the same report was true and the loop was unstable from 105.8°, because a slow mode that the phase margin does not describe lost its damping. It asks for the closed-loop

eigenvalues of the sampled loop at every angle, with the least damping ratio, not only the crossover margin. It asks for the delay margin and the lowest gain factor as well, since Table 1.1 shows the measures ranking gains differently. It asks for nonlinear steps toward and past the edge of the range, and for the margin quoted on the mode that fails.

15.3 A good answer states each limit as a number against a capability. The pothole asks 1579 m/s^2 of a wheel that falls at 98.1 m/s^2 from spring and weight, so the tire leaves the road. The ball needs 15.3 cm to stop at the 3° limit and has 10 cm. The lobe depends on depth and spindle speed, not feed, so the speed rule moves the cut out of the unstable region. A saturated drive makes the shaped and unshaped moves alike, 7.7° either way. What helps is a change of equipment, operating point or command, and the example from the reader's field should have the same structure.

15.5 A good answer notes that the controller holds the plant either way, so the trends look normal, and the read only shows when a parameter is changed in a way the measurements alone would not reveal. The probe changes a hidden parameter before Run (or the same change through a disturbance input), keeps every measurement identical where possible, and asks whether the controller's first output or its integral moves. The levitation numbers, IAE 0.108 through the slider and 0.008 mm s through the parameter, show the signature.

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.

active magnetic bearing A bearing that holds a spinning shaft in the air with electromagnets steered by feedback from position sensors. *Chapter 6.*

active suspension A car suspension with an actuator that can push or pull between the body and the wheel, rather than only resist motion. *Chapter 11.*

actuator saturation The condition in which a controller asks for more than the motor, valve or brake can give, so the extra request has no effect. *Chapter 1.*

adaptive control Control that adjusts its own settings while it runs, as a feed controller does when the cutting conditions change. *Chapter 13.*

adaptive cruise control A speed control that measures the gap to the car ahead and slows down to keep a safe distance. *Chapter 14.*

aeropendulum A light arm on a pivot that a propeller at its end lifts to a commanded angle. *Chapter 3.*

alpha-beta filter A simple estimator that corrects a predicted position and speed by fixed fractions of each new measurement error. *Chapter 5.*

anti-lock braking A brake control that releases and reapplies the pressure so that a braked wheel keeps turning instead of locking and skidding. *Chapter 4.*

back-calculation An anti-windup method that resets the integral of a controller to agree with what the actuator actually delivered. *Chapter 1.*

ball and beam A tilting beam along which a ball rolls, held at a target position by changing the angle of the beam. *Chapter 5.*

Bryson's rule A first choice of the weights of a regulator: one over the square of the largest acceptable excursion of each state and input. *Chapter 1.*

capsize mode A slow motion in which a riderless bicycle at high speed leans over and falls by itself. *Chapter 9.*

cascade control An arrangement in which one controller sets the setpoint of another, faster one, as an angle loop inside a position loop. *Chapter 1.*

chatter A self-excited vibration of a cutting tool that leaves a wavy surface and can break the tool. *Chapter 13.*

CNC Computer numerical control: a machine tool that moves its axes by a stored program. *Chapter 13.*

complementary filter An estimator that blends a fast but drifting sensor, such as a gyro, with a slow but unbiased one, such as an accelerometer. *Chapter 9.*

continuous stirred-tank reactor A well-mixed vessel with steady inflow and outflow, in which the reaction runs at the conditions of the outlet. *Chapter 2.*

controllability The property that the inputs can drive every state of a model to any chosen value. *Chapter 1.*

cooperative adaptive cruise control Adaptive cruise control that also receives the acceleration of the car ahead by radio and acts on it. *Chapter 14.*

countersteering Steering briefly the wrong way to start a lean, so that a bicycle or a car can then turn the way it should. *Chapter 9.*

damping ratio A number that says how quickly an oscillation dies out, from zero for none to one for a response with no overshoot. *Chapter 3.*

delay The time between an action and the first effect a sensor can see, which limits how fast a loop may act. *Chapter 1.*

derivative filter A low-pass filter on the derivative term that keeps measurement noise from jolting the actuator. *Chapter 5.*

describing function The gain of a nonlinear element, such as a relay, on a sinusoid of a given amplitude, used to predict a limit cycle. *Chapter 1.*

distributed control system (DCS) A supervisory level of control over a group of local controllers that together run the loops of a plant. *Chapter 2.*

dither A small added oscillation that keeps an actuator moving so that friction does not hold it still. *Chapter 5.*

double integrator A plant whose input sets an acceleration, so that its position follows the input integrated twice. *Chapter 5.*

double inverted pendulum Two hinged links balanced upright, one above the other, on a moving cart. *Chapter 8.*

drifting Driving a car with its rear tires sliding at a large angle, held at an unstable equilibrium by steer and throttle. *Chapter 10.*

dynamic inversion The name used in flight control for feedback linearization. *Chapter 3.*

eigenvalue A rate of the natural motion of a linear model. One with a positive real part belongs to a motion that grows. *Chapter 1.*

electronic control unit (ECU) A small computer with inputs and outputs that runs one or more control functions of a car. *Chapter 4.*

electronic stability control A car system that brakes individual wheels to help the driver when the car begins to spin out or plow out. *Chapter 10.*

encoder A sensor on a shaft or rail that counts pulses to report a position or an angle in fixed steps. *Chapter 1.*

energy swing-up Pumping energy into a hanging pendulum until it reaches the energy of the upright position, where another controller catches it. *Chapter 7.*

extremum-seeking control Control that varies its own reference to find the peak of a curve rather than hold a fixed value. *Chapter 4.*

feedback linearization A control law that cancels the nonlinearity of a plant through a model, so that a linear law can act on what remains. *Chapter 1.*

feedforward Moving an actuator from a known command or disturbance before an error appears, as gravity feedforward holds the arm. *Chapter 1.*

fold bifurcation A point where two steady states meet and vanish as a setting changes, so that beyond it no steady state exists. *Chapter 1.*

friction circle The limit on the total grip of a tire, shared between driving or braking and cornering. *Chapter 10.*

gain scheduling Changing the settings of a controller with a measured condition, such as an angle or a speed, so that one design holds everywhere. *Chapter 1.*

gantry crane A crane whose trolley runs along a beam on legs and carries a load that hangs from a cable. *Chapter 12.*

groundhook damping A suspension law that damps the wheel as if it were tied to the road, to keep the load on the tire steady. *Chapter 11.*

Hopf bifurcation The point where a pair of eigenvalues crosses into the right half plane and a steady oscillation is born. *Chapter 1.*

input shaping Splitting a command into timed parts whose oscillations cancel, so that a lightly damped system stops without swinging. *Chapter 1.*

integrator windup The growth of the integral of a controller while its actuator sits at a limit, which causes a large overshoot afterward. *Chapter 3.*

invariant point A frequency at which no suspension force can change a response, such as the acceleration of the body near wheel hop. *Chapter 11.*

inverted pendulum A pole balanced upright on a moving cart, which falls unless a controller moves the cart under it. *Chapter 7.*

Kalman filter An observer whose correction gain is chosen from the statistics of the process noise and the measurement noise. *Chapter 1.*

limit cycle An isolated, repeating oscillation whose size is set by the system rather than by where it started. *Chapter 1.*

linear quadratic regulator (LQR) A state-feedback gain that minimizes a weighted sum of the squared state errors and the squared input over time. *Chapter 1.*

linearization A straight-line model of a nonlinear plant that holds near one operating point. *Chapter 1.*

Lyapunov function An energy-like function that always decreases along the motion of a system, which proves that the motion settles. *Chapter 8.*

magnetic levitation Holding an object in the air by magnetic force alone, with no contact. *Chapter 6.*

model predictive control (MPC) Control that computes the best inputs over a future horizon at each step and applies only the first of them. *Chapter 2.*

moving horizon estimation (MHE) An estimator that fits a model to a window of recent measurements at each step, within bounds on the states. *Chapter 2.*

observability The property that the measurements, taken over time, reveal every state of a model. *Chapter 1.*

observer A model run beside the plant and corrected by the measurement error, to estimate the states that are not measured. *Chapter 1.*

open-loop instability The behavior of a plant that drifts away from its operating point unless a controller holds it there. *Chapter 1.*

operating point The steady state at which a plant runs and around which its model is linearized. *Chapter 1.*

phantom traffic jam A jam that forms on a road with no bottleneck, from small changes of speed that grow from car to car. *Chapter 14.*

phase margin How much extra lag a loop can take at its crossover frequency before it becomes unstable. *Chapter 1.*

PID control A controller that acts on the error, its integral and its rate of change, called PD when the integral is left out. *Chapter 1.*

pole placement Choosing the gains of a state feedback so that the closed loop has the eigenvalues the designer asks for. *Chapter 1.*

Posicast control An input shaper of 1957 that splits

a step into two parts spaced so that their ringing cancels. *Chapter 12.*

preview control Control that measures a disturbance, such as the road ahead of a wheel, before it reaches the plant. *Chapter 11.*

programmable logic controller (PLC) A rugged computer that reads its inputs, runs a program in a fixed cycle and sets its outputs. *Chapter 12.*

quantization Rounding a measurement or a command to fixed steps, such as the counts of an encoder. *Chapter 3.*

quarter car A model of one corner of a car: a body mass and a wheel mass joined by a spring and a damper, on a tire. *Chapter 11.*

regenerative chatter Chatter fed by the wavy surface that the tool left on its previous pass. *Chapter 13.*

region of attraction The set of starting states from which a controlled system returns to its equilibrium. *Chapter 1.*

relay feedback Feedback that switches the input between two values by the sign of an error. *Chapter 1.*

resonance A frequency at which a lightly damped system responds strongly to a small repeating input. *Chapter 1.*

Riccati equation The matrix equation whose solution gives the gain of a linear quadratic regulator. *Chapter 1.*

right-half-plane pole (RHP pole) A pole with a positive real part, which makes the open-loop response grow. *Chapter 1.*

right-half-plane zero (RHP zero) A zero with a positive real part, which makes a response start in the wrong direction. *Chapter 1.*

Routh stability criterion A test on the coefficients of a characteristic polynomial that tells whether every root has a negative real part. *Chapter 1.*

sample time The interval between two executions of a digital controller. *Chapter 3.*

self-stability The ability of a moving bicycle with no rider to stay upright within a range of speeds. *Chapter 9.*

semi-active damper A damper whose resistance can be changed quickly but which can only absorb energy, never supply it. *Chapter 11.*

separation principle The result that a state feedback and an observer can be designed apart, and the closed loop keeps the poles of both. *Chapter 1.*

sideslip The angle between the direction a car points and the direction it travels. *Chapter 10.*

skyhook damping A suspension law that damps the body as if it were tied to a fixed point in the sky. *Chapter 11.*

sliding-mode control A switching control that drives a chosen combination of the states to zero and holds it there. *Chapter 4.*

slip The fraction of the speed of a tire at its contact patch that is sliding rather than rolling. *Chapter 4.*

spacing policy The rule that sets the gap a following car wants, such as a fixed distance or a time gap. *Chapter 14.*

stability lobes A chart of the largest chatter-free depth of cut against spindle speed, with pockets of stable speeds. *Chapter 13.*

state-space model A model written as first-order equations for a set of states driven by inputs. *Chapter 1.*

steady-state multiplicity The existence of more than one steady state for the same constant input. *Chapter 1.*

string stability The property of a line of following cars, or of loops in a chain, that a disturbance does not grow as it passes along. *Chapter 1.*

thermal runaway A rise of temperature that speeds a heat-releasing reaction, which raises the temperature further. *Chapter 2.*

thermowell A closed tube that protects a temperature sensor in a process line or vessel and slows its reading. *Chapter 2.*

time gap The spacing between cars measured in seconds of travel, so that the wanted gap grows with speed. *Chapter 1.*

underactuated system A system with fewer inputs than motions to control, such as a cart that must also hold a pole. *Chapter 7.*

variable frequency drive An electronic drive that sets the speed of an alternating-current motor by the frequency of the power it feeds. *Chapter 1.*

weave mode A swaying and steering oscillation of a bicycle that grows at low speed. *Chapter 9.*

wheel hop The bouncing resonance of a wheel between the tire and the suspension spring. *Chapter 11.*

ZV and ZVD shapers Input shapers of two impulses half a period apart, or of three, that cancel the leftover swing of one mode. *Chapter 1.*

Index of Control Topics

actuator authority, 233
actuator saturation, 144, 204, 232, 256, 280, 283, 304
adaptive control, 318
  identified gain, 324
AI agents
  engineering practice, 372
  open problems, 376
  reproduction, 375
  truth audit, 374
  verification record, 373
  while studying, 16
alpha-beta filter, 116
anti-sway control, 371

back-calculation, 87, 273, 324
Bryson's rule, 116, 141, 229, 251
bullwhip effect, 340
bumpless transfer, 15, 34, 64, 116, 172, 230, 250, 274, 300, 325, 350

cascade control, 8, 25, 165, 229, 251
  inner-loop speed, 108
  inner-loop stability, 120
characteristic polynomial, 195
chatter, 368, 371
coefficient of restitution, 162, 193
commissioning, 152, 305
complementary filter, 228
conditional integration, 63
controllability, 6
controller action, 31, 87, 228, 274, 350
crane safety, 309

damping ratio
  against phase margin, 61
data export, 16
dead time, 330
delay differential equation, 321
derivative action, 367
derivative filter, 114, 125
describing function, 9, 61, 90
dither, 109, 180
double integrator, 107, 113
drifting, 370

eigenvalues, 5, 226, 248, 295
energy swing-up, 160, 161, 190
equilibrium
  solver, 374
extremum regulation, 77
extremum-seeking control, 80

feedback linearization, 56, 64, 71, 109
feedforward, 289
feedforward control
  model error, 141
fold bifurcation, 4, 29
friction
  destabilizing, 165
  stick-slip, 163, 181, 194
friction estimation, 101, 251

gain floor, 165, 196
gain scheduling, 3, 5, 32, 64, 72, 90, 217, 220, 251
graceful degradation, 350
gravity feedforward, 63
groundhook damping, 266, 273

Hopf bifurcation, 9, 31
human operator, 236

infeasible setpoint, 68, 249
input shaping, 8, 288, 368
  robust, 310
  zero-vibration (ZV) shaper, 296
  ZVD shaper, 297
integrator windup, 63, 88
invariant point, 263, 271

Kalman filter, 6, 15, 197, 276, 301, 351
  extended, 142
  input disturbance, 34

limit cycle, 9, 31
  contact loss, 320
linear quadratic regulator, 6, 64, 116, 368
  discrete, 168, 198, 251
  scheduled, 229, 251
  two-input, 241
linearization, 5, 60, 138, 248, 346
loop interaction, 250

LuGre friction model, 80
Lyapunov function, 191, 372

mass matrix, 191
measurement noise, 94, 112, 136
Merritt criterion, 321
model predictive control, 34, 47, 169, 199, 274, 300
  against state feedback, 370
  distributed, 350
  explicit, 141
  in the apps, 15
  learning-based, 371
  move blocking, 200
  soft constraints, 170, 200, 276
  terminal cost, 171
moving horizon estimation, 15, 35

networked control, 344, 360, 361
Newton iteration, 248
noise amplification, 145, 203

observability, 6
observer, 6, 367
offset-free control, 142, 201, 276, 301, 351
open-loop instability, 2, 22, 60, 158, 188, 242
operating point
  process gain, 139
  stability that changes with, 53

performance limitations, 7, 115, 126, 152, 348, 376
phase margin, 61, 86
  loss with dead time, 114, 330
PID control
  anti-windup, 14, 33, 63, 140
  derivative action, 114, 143, 295
  derivative filter, 140
  I-PD, 147
  proportional kick, 147
  velocity form, 14, 33, 62, 140, 250, 273, 324, 349
PID tuning
  internal model control, 32
pole placement, 6, 182
Posicast control, 291
preview control, 263, 265
process safety, 46

quantization, 59, 70, 112, 163, 193
quantized actuator, 59

rate limit, 135, 273, 350
reference governor, 207
region of attraction, 7, 183, 191, 211, 372
relay autotuning, 36
relay feedback, 10
repetitive process, 316
residual vibration, 298
resonance, 8, 270, 293, 320
Riccati equation, 251
right-half-plane pole, 7, 86, 131, 165, 195
right-half-plane zero, 7, 165, 195, 227
Routh stability criterion, 61, 142

sample time, 12, 68, 203
sampled-data control, 166, 196, 230
scenarios, 16
  completion rules, 16
scorecard, 16
self-stability, 217
sensor delay, 112, 371
simulation engine, 14
simulator framework, 12
skyhook damping, 265, 273, 283
sliding-mode control, 89
  boundary layer, 89
  reaching condition, 89
spacing policy, 343
stability boundary, 86, 321, 333
stability lobes, 315, 332
stability margin, 11, 366
stabilization, 1, 2
  design procedure, 369
  open problems, 375
state observer, 109
state-space model, 5
static map, 58
steady-state gain
  nonlinear, 3
  sign change, 31, 77, 226
steady-state map
  fold, 86
steady-state multiplicity, 3, 22, 374
step test
  on a nonlinear plant, 32, 64
string stability, 10, 338–340, 371
synchronous averaging, 325
system identification, 236

time delay, 166, 229, 258, 316, 344
time-varying feedback, 198
trajectory feedforward, 109, 198

trajectory optimization, 199

underactuated system, 158, 188, 288, 368
unstable steady state, 139, 241

valve stiction, 108
van Heerden diagram, 29

wheel slip control, 77

Index of Applications

accelerometer, 220, 281, 331
active magnetic bearing, 133, 152
active suspension, 263
adaptive cruise control, 339, 359, 360
aeropendulum, 53
    rig and kit, 56
aircraft antiskid, 79
anti-lock braking, 77, 99, 371
    low-speed exit, 88
    rule-based pressure cycle, 88
Arrhenius kinetics, 23

balancing robot, 159
ball and beam, 107
bicycle, 217
    benchmark, 219
bounce test, 280
breakaway friction, 111
brushless motor, 54
bump stop, 266
bumper, 162, 193, 210

capacity drop, 356
capsize mode, 226
cart and pole, 158, 188
cast-iron insert, 330
castering mode, 226
center of percussion, 163
chatter detection, 325
chatter marks, 326
container terminal, 291
continuous stirred-tank reactor, 22
control valve
    split range, 45
cooling jacket, 25
cooperative adaptive cruise control, 341
countersteering, 219
    in a drift, 248
crane drive, 293, 308
crane operator, 290
current amplifier, 135, 151

damper wear, 280
deck stop, 193
double inverted pendulum, 188
driveline lag, 342, 359

dual-antenna navigation unit, 250, 256

electric drive, 210, 247
electromagnetic suspension, 133
electronic speed controller, 58, 70
electronic stability control, 80, 243, 244, 257
encoder, 11, 163, 193, 210
end mill, 316
exothermic reactor, 22

feed override, 324
figure-eight, 251
friction circle, 246

hobby servo, 110, 124
hoist, 294, 308
hydraulic brake modulator, 83, 99

impact test, 323
intelligent driver model, 343
inverted pendulum, 158
iterative LQR, 199

load leveling, 273

machine tool, 315
magnetic encoder, 59
magnetic levitation, 131, 133
magnetorheological damper, 266, 281
milling, 315
multirotor vehicle, 54

overhead crane, 288

passive twin, 269
path following, 251
payload, 291
pendulum, 3
phantom traffic jam, 341, 356
position sensor
    camera, 110
    resistance wire, 110, 124
pothole, 268
propeller actuator, 58
propeller thrust, 58

quarter car, 263

quarter-car model, 80

rate gyro, 181, 220, 234
regenerative chatter, 315, 318
residual sway, 305
ride comfort
    ISO 2631-1, 269, 282
rider model, 229
road profile, 268
road roughness
    ISO 8608, 265
rolling radius, 110
rolling resistance, 111

semi-active damper, 265, 266
sensor noise, 150
single-track vehicle model, 244
skidpad, 241
spindle drive, 321, 332
    power limit, 330
spindle speed selection, 325
steer motor, 220, 235

thermal runaway, 24, 25
thermowell, 29, 44
tire
    brush model, 246
    Burckhardt model, 82
    combined slip, 246
    contact loss, 268
    Magic Formula, 80
tool wear, 321
torque motor, 112
trail, 222
trolley, 291

vehicle drifting, 241
vehicle platoon, 338

washboard road, 268
weave mode, 226
wheel hop, 264
wheel-speed sensor, 83, 98
Whipple bicycle, 217
wind load, 305

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.