TUNING AND CONTROL LOOP PERFORMANCE

TUNING AND CONTROL LOOP PERFORMANCE

FOURTH EDITION

GREGORY K. MCMILLAN

MOMENTUM PRESS

MOMENTUM PRESS, LLC, NEW YORK

Tuning and Control Loop Performance, Fourth Edition

First published by Momentum Press®, LLC
222 East 46th Street, New York, NY 10017
www.momentumpress.net

ISBN-13: 978-1-60650-170-2 (print)
ISBN-13: 978-1-60650-171-9 (e-book)

Momentum Press Manufacturing and Engineering Collection

DOI: 10.5643/9781606501719

Cover and interior design by Exeter Premedia Services Private Ltd. Chennai, India

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Abstract

The proportional-integral-derivative (PID) controller is the heart of every control system in the process industry. Given the proper setup and tuning, the PID has proven to have the capability and flexibility needed to meet nearly all of industry's basic control requirements. However, the information to support the best use of these features has fallen behind the progress of improved functionality. Additionally, there is considerable disagreement on the tuning rules that largely stems from a misunderstanding of how tuning rules have evolved and the lack of recognition of the effect of automation system dynamics and the incredible spectrum of process responses, disturbances, and performance objectives. This book provides the knowledge to eliminate the misunderstandings, realize the difference between theoretical and industrial application of PID control, address practical difficulties, improve field automation system design, use the latest PID features, and ultimately get the best tuning settings that enables the PID to achieve its full potential.

The book provides up to date comprehensive concepts and extensive details on the practical application of not only the PID but all of the automation system components that affect loop performance from a user's viewpoint that is not available elsewhere in the literature. The book starts with a summary of the essential fundamentals. The reader gains a foundation of the PID functionality, the major types of process responses, the dynamics that originate in the automation system, and the major tuning rules. Since the complex interrelationships and common misconceptions can be overwhelming, a unified methodology provides the focus and procedures to tackle the most challenging applications. While the unified methodology advocates the use of a particular set of tuning rules, the general approach can be still be used with other tuning rules based on the knowledge gained elsewhere in the book. The book moves on to provide considerable depth on performance objectives and the effect of process, controller, measurement, and valve dynamics so that the practitioner can improve not only tuning but the setup of the PID and the design of the field equipment and automation system. The book provides insight and details on how disturbances, nonlinearities, and interactions affect tuning and performance. The book finishes up with how the role of the PID can be expanded through cascade control, advanced regulatory control, process control improvement, auto tuners, adaptive controllers, and batch optimization techniques. Test results at the conclusion of many of the chapters illustrate and quantify the effect of dynamics, application problems, and tuning solutions. The appendixes provide supporting information that the reader can explore on an as needed basis to better understand the nature and source of the concepts and details presented.

KEY WORDS

adaptive control, advanced regulatory control, analyzer response, auto tuner, automation system, batch optimization, bioreactor control, cascade control, compressor control, control loop performance, control valve response, external reset feedback, feedforward control, inverse response, lambda tuning, level control, measurement response, pH control, PID control, PID execution rate, PID filter, PID form, PID structure, PID tuning, pressure control, process control, process disturbances, process dynamics, process interaction, process metrics, process nonlinearity, process performance, process response, proportional-integral-derivative controller, reactor control, runaway reaction, temperature control, valve deadband, valve position control, valve resolution, variable frequency drive response, wireless control, wireless response

Contents

List of Figures

List of Tables

Acknowledgments

There were many Monsanto and Solutia Directors and Senior Fellows who played a key role technically and personally in my journey as an automation professional. Foremost among those who shaped my career was the late Dr. Jim Fair, Honorary Professor Emeritus at the University of Texas, the creator and first director of the Engineering Technology (ET) Department. Being part of the ET Department allowed access to the best technical minds in the world, along with the freedom to explore and the encouragement to publish. My second ET director, Dr. Stan Proctor, shared his extensive knowledge in modeling and control and provided direction and support. The appreciation of my independence and humor by my last ET director, Dr. Adrian Hutton, another Texan, enabled me to publish many funny and creative articles and books. The knowledge I inherited from the late Vernon Trevathan and gained from conversations with Senior Fellows at Solutia and Monsanto, including Dr. Henry Chien, Dr. Bob Otto, Dr. Terry Tolliver, and Stan Weiner, has been of inestimable value.

When I retired from Solutia, I was fortunate to have the opportunity to become a consultant to Emerson Process Management at their headquarters in Austin in 2004. Here I have had access to some of the best minds on the supplier side, including James Beall, Terry Blevins, Mark Coughran, Lou Heavner, Mike Ott, and Willy Wojsznis. Jim Cahill has helped me to effectively use social media. The support by Mark Nixon, manager of DeltaV R&D, and Grant Wilson, vice president of DeltaV technology of my endeavors has enabled me to continue my journey. Most recently my role as a simulation consultant at MYNAH technologies has opened the door to seeing to fruition the development of dynamic models that can be deployed to gain better process understanding and recognition of control opportunities. Mart Berutti, Nobin William, and Dave Sextro have created an inspiring and friendly environment to promote creativity and achievement with the next generation of engineers to get the most out of simulation.

Special thanks go to Pierce Wu at MYNAH Technologies who developed in MiMiC the test cases to illustrate the key differences in process and valve responses. The trend chart of the open loop response of dead time dominant, self-regulating, integrating, and runaway processes in Chapter 1 shows the dramatic differences caused by the degree and type of feedback internal to the process. The trend charts in Chapter 7 shows the dynamic implications of backlash and stick-slip in the open loop and closed loop response of valves for various processes.

Editing and comments by key participants in the ISA Mentor Program including Brian Hrankowsky, Danaca Jordan, Leah Ruder, Hector Torres, and Keneisha Williams increased the usability and the technical questions by Brian and Hector improved the pertinence of the technical content. Hector also provided the 3D plots of ultimate period and ultimate gain in Chapter 4 and the curve fit for the test results for the reduction in peak error by cascade control for Chapter 11. My long time associate Owen Campney reviewed and commented on the equations entered.

Finally, the patience and understanding of my wife, Carol, have made this book and the sharing of what I have learned a part of our life.

Preface

Most plants depend entirely upon proportional-integral-derivative (PID) control. Even plants with model predictive control still depend on the PID to deal with valve nonlinearities and provide fast control of flows and pressures. Yet most of the capability of the PID is not effectively used due to disagreements as to PID tuning rules and a lack of guidance on how to meet different objectives and deal with process and automation system dynamics. This book seeks to provide an intensive and extensive view of what is needed to get the most out of the PID based on 44 years of experience in applying PID control in the process industry. New tools for analyzing and tuning loops offer automatic identification of loop dynamics, PID settings, and statistics on variability. This book seeks to take advantage of the new ability of tools today to know the dynamics and choose tuning rules and factors to free up the individual to focus more on selecting PID options and setting parameters to meet the challenges and objectives of the application.

The first edition in 1984 focused on the tuning methods prevalent at the time with the objective of minimizing the peak and integrated error from unmeasured load disturbances at the process input. Tuning techniques to minimize these errors were aggressive. In practice the PID gain was cut in half to provide more robustness (e.g., gain margin increased from two to four). While minimizing these errors is still important for preventing activation of safety instrumentation systems, relief devices, preventing environment violations, and preventing off-spec product in columns, crystallizers, evaporators, and reactors, there are many other considerations. The book has been almost completely rewritten in the fourth edition to meet other objectives such as minimizing overshoot in the setpoint response, minimizing the propagation of oscillations due to interactions and resonance, maximization of variability absorption for surge tank level, and maximizing the coordination of loop responses to reduce transients in the material and energy balances. Also provided is the recognition of when measurement, valve, and variable frequency drive (VFD) dynamics and resolution limits present a problem and the use of PID tuning and options to minimize the consequences until the source of the problem is fixed. Tuning is not meant to be a cover up but a symptom of a system problem. Dynamics and tuning and can be used to find how to improve the equipment, piping, measurement, PID, valve, or VFD design and installation.

The fourth edition is exceptionally long because the topic is so huge and practical expertise that plays such a key role in the success largely remains in the brains of senior specialists, consultants and technologists. The principles, simple algebraic equations, recommendations, examples, test results, and key points presented seek to make the information more accessible. While some of the foundation is derived from the frequency domain, the details and understanding is based on relationships are detailed in the time domain since this is what is seen on trend recordings and can deal with discontinuities from analyzers and valve backlash and stick-slip. The analysis in the time domain is more consistent with the thought process of practitioners opening up a dialog between process control, operations and process design. This is not to discount the value of frequency domain analysis and power spectrums available in new tools that can build upon what is learned in university courses on control theory.

If I had to pick chapters to read first, I would recommend Chapters 1–3 to provide the basis and overview of the total solution. Subsequent chapters provide more of the details for a spectrum of application considerations, problems and solutions. The appendices provide significant technical support.

Chapter 1 starts out with the basics and provides an overview of tuning rules that are representative of the more than 100 rules documented in the book *proportional-integral* (*PI)* and *PID Controller Tuning Rules* by O'Dwyer. The fourth edition seeks to show how many of these rules converge to provide the same PID gain and reset time with slightly different factors when the objective is the minimization of errors from load disturbances. The author apologizes if a favorite tuning method is not shown. The important point is that the loop should be tuned and the egos behind particular rules should take a back seat to the solution. A good practitioner will get good results by adjusting factors and using PID features. The fourth edition seeks to document the practical expertise that is more important than the rule. Test results are used in most of the chapters to explore, discover and verify concepts.

Chapter 2 offers a comprehensive overview of the total system solution. A unified methodology is presented to tie together the whole solution. Chapter 3 provides simple equations to estimate the effect of tuning settings on peak and integrated errors for load response and the ability to get to a new set point quickly. The relationships between tuning and performance seen in the equations are more important than the actual use of the equations to predict particular values.

The effect of process and mechanical design is analyzed in Chapter 4 and the effect of automation system components are extensively discussed including tables of typical dynamics in Chapters 5–7. The effect of disturbances, nonlinearities and interactions are detailed in Chapters 8–10. The book concludes with a discussion in Chapters 11–15 of how to get the most out of cascade control, advanced regulatory control, process control improvement, auto tuners and adaptive control, and batch optimization.

CHAPTER 1

Fundamentals

1.1 INTRODUCTION

The proportional-integral-derivative (PID) as well as the proportional-integral (PI) controller is the workhorse of the process industry. More than 99.9 percent of basic control systems rely on this controller. Even when model predictive control (MPC) is widely employed, a PI controller is normally used for flow control to deal with the nonlinearities of the installed flow characteristic and to enable flow ratio control.

1.1.1 PERSPECTIVE

For unmeasured disturbances, the PID has been proven to provide essentially optimal control (Bohl and McAvoy 1976). When fast and aggressive control is needed to prevent activation of a safety instrumented system (SIS) or relief device, a PID controller is necessary. A correctly tuned PID controller with a gain greater than 10 and a derivative time greater than one minute are sure signs that the immediate action of the proportional and derivative modes of the PID are needed.

In contrast, the MPC response to an unmeasured disturbance is more like the integral mode in terms of the prolonged gradual correction. The MPC output is a scheduled series of moves (changes in the manipulated variable) based on the difference between the setpoint and the MPC predicted profile that is computed from past moves with no knowledge of unmeasured disturbances. A fraction of the error between the predicted profile and the actual profile caused by the unmeasured disturbance is used to bias the predicted profile, to provide feedback control. The more gradual action and the inherent multivariable capability of MPC is particularly advantageous to interacting systems and gas or plug flow volumes. In these applications, a large primary process time constant or a small integrating process gain that would smooth out control actions and provide a separation of dynamics to minimize the consequences of interaction does not exist. MPC also provides exact dynamic compensation of measured disturbances.

This book will show that a high PID gain and rate time setting is indicative of a process time constant or integrating process gain that is extremely slow compared to the dead time. Often these dynamics are in a single loop or cascade loop, responsible for quality control. Overdrive of the output (output driven past the balance point) is needed to correct for a disturbance or to achieve a new setpoint in these near-integrating (large process time constant), true integrating, and runaway processes (highly exothermic reactors). Integral action provides the overdrive too late. Integral action has no sense of direction or anticipation. Excessive integral action can be unsafe in integrating and runaway processes. The most frequent problem observed for these processes is a reset time too small.

A PID module execution time that is less than one second is indicative of a loop with a process dead time and time constant less than one second. Since the fastest execution time of an industrial MPC is one second and the MPC latency is generally much greater than PID latency due to the complexity of the calculations, the MPC will be the largest source of dead time for these loops. The ultimate limit to the integrated error for an unmeasured disturbance on the process input will be shown in Chapter 3 as being proportional to the total loop dead time squared.

The PID controller has not been effectively used, primarily because PID performance depends almost entirely upon tuning and the correct application of PID options assuming the control strategy and configuration is correct. MPC is increasingly displacing PID because the MPC tuning is minimal and implementation is more automated. MPC is also better understood by chemical and petroleum engineers since the focus is on the steady-state response. In contrast, the expertise required for tuning and use of PID features for diverse applications is considerable and largely undocumented. The PID time frame is much shorter with the delay and initial rate of change of the response being more important than the final value of the process variable (PV) for the more important loops (e.g., concentration and temperature). Proper tuning and utilization of the PID depends upon the knowledge of the dynamics of the process excursion whereas the MPC performance is much more dependent upon the accuracy of the steady-state gains.

The tuning rules are largely a subject of continual argument. Over 100 tuning rules have been published, and all the authors are convinced their rules are the best. This strange situation is the result of an incredibly diverse range of dynamics and objectives in industrial process applications and the lack of understanding of the relationship between tuning and plant performance objectives. Gamesmanship is also at play. Everyone wants to win and show their methods are the best. Because there is no consensus on tuning rules and how performance depends upon tuning, a person can prove almost any point by how the PID is tuned for the case presented.

The PID has incredible flexibility. However, the powerful features responsible for this extensive capability are not extensively utilized because of the lack of understanding and guidance. These features can prevent oscillations from deadband (e.g., backlash), split range discontinuities (e.g., transition from steam to cooling water valve), resolution or threshold sensitivity limits (e.g., stick-slip), slow secondary loops, at-line and offline analyzers, nonlinearities, recycle streams, and interactions.

Chapter 1 starts with an introduction to the basic functionality of PID control and the advantages and disadvantages of each mode. The chapter moves on to detail the major tuning rules and to discuss the relative merits. The chapter concludes with a general solution to reduce down to two sets of tuning rules to be used in conjunction with software that can identify the process dynamics. The proper tuning gained from Chapter 1 sets the stage for Chapter 2 to concisely detail a unified methodology to get the most out of the PID for industrial process applications.

1.1.2 OVERVIEW

Most of the tuning rules were designed with one or two particular types of dynamic responses in mind when in fact there are four major types of the dynamic process response. The significance of the size of the actual or equivalent primary process time constant is a commonly misunderstood key characteristic of process responses. The lack of recognition of the types of responses, nonlinearities, and process objectives has led to major differences in tuning rules. This chapter discusses the relative capabilities of methods for various processes in preparation for reaching a general solution. Without completely giving away the answer we itemize the functional recommendations.

1.1.3 RECOMMENDATIONS

1. Carefully select the control action (reverse or direct) based on the process action (reverse or direct). Note that for a fail open valve if the valve action (increase-close) is not set in the PID, analog output (AO) block, current to pneumatic Transducer (I/P), or positioner reversing the direction of the output signal, the control action must be changed to the opposite direction to account for the valve action.
2. Determine whether output scale is in engineering units or percent and set the output limits to match the operating limits.
3. If separate anti-reset windup (ARW) limits exist (e.g., PRoVOX and DeltaV PID), set these ARW limits equal to the output limits using proper units. An exception to this would be when recovery from an output limit needs to be fast, set the respective ARW limit inside the output limit.
4. Maximize the use of the proportional mode for processes that have a slow gradual response.
5. Maximize the use of derivative mode to compensate for secondary time constants especially in integrating and runaway processes.
6. Maximize the use of the integral mode for processes that have a fast or abrupt response.
7. Determine the tuning units and convert settings necessary as per Appendix K.
8. Determine the PID Form used and convert settings necessary as per Appendix K.
9. If the International Society of Automation (ISA) Standard Form is used, ensure the rate time does not exceed the reset time and is less than the total of the dead time and secondary lag. For nearly all processes, the ISA Standard Form rate time should be less than one-fourth the reset time.
10. For processes with a slow gradual response, set the rate time equal to or greater than the secondary lag to greatly improve control.
11. Minimize PV filter time and transmitter damping settings (e.g., just large enough to keep output fluctuations within valve dead band). Large measurement time constants can delay the PID controller's response and lead to undesired effects in the field even giving the illusion of better control.
12. Use the inverse of the maximum product of total dead time and process excursion rate per percent change in PID output for setting the PID gain.
13. For the best rejection of unmeasured fast disturbances (e.g., reactor pressure or temperature control), *maximize* the transfer of variability from the PV to the PID output by tuning the PID for aggressive reaction. An exception to this would be to avoid overshoot of final resting value of PID output for plug flow volumes (volumes with no back mixing and consequently no process time constant).
14. For the greatest absorption of variability where tight control is not required (e.g., surge tank level control), *minimize* the transfer of variability from the PV to the PID output by

tuning the PID to provide a gradual response or adding a secondary flow loop setpoint rate limit to slow down the change in the manipulated flow. The slowest allowable rate of change in the PID output is the available change in PID output (e.g., difference between current output and output limit) divided by the time to an alarm or constraint violation.

1.2 PID CONTROLLER

A PID controller has proportional, integral, and derivative modes. The term PID is often still used even when a mode is omitted. Frequently, the derivative mode is not used but the controller is still called a PID. Whether a mode is turned off or whether a mode acts on error (difference between setpoint and PV) or just on the PV depends upon the structure chosen as detailed in Section 2.2.2.

In nearly all industrial controllers, the PID algorithm uses a PV and setpoint that is a percent of an input scale and provides an output that is a percent of an output scale. The lack of understanding that the signals are percent has caused a lack of recognition of the effect of final control element [e.g., valves and variable speed drive (VSD)] characteristics and size and measurement calibration span on PID tuning.

PID controllers in some university labs, simulations, and studies may use signals in engineering units and rarely have the extensive options or features in industrial controllers. The richness and complexity of proprietary industrial PID algorithms translate to a need to use virtual or actual plant tests to fully see the true PID capability. Most of the test results in this book use a virtual plant where the actual industrial PID algorithm is downloaded.

Figure 1.1 shows the contribution of each mode to the total output response of a PID controller with a structure of PID on error for a step change in setpoint. A block valve in series with the control valve is assumed to be closed so that the PV is constant, enabling the observation

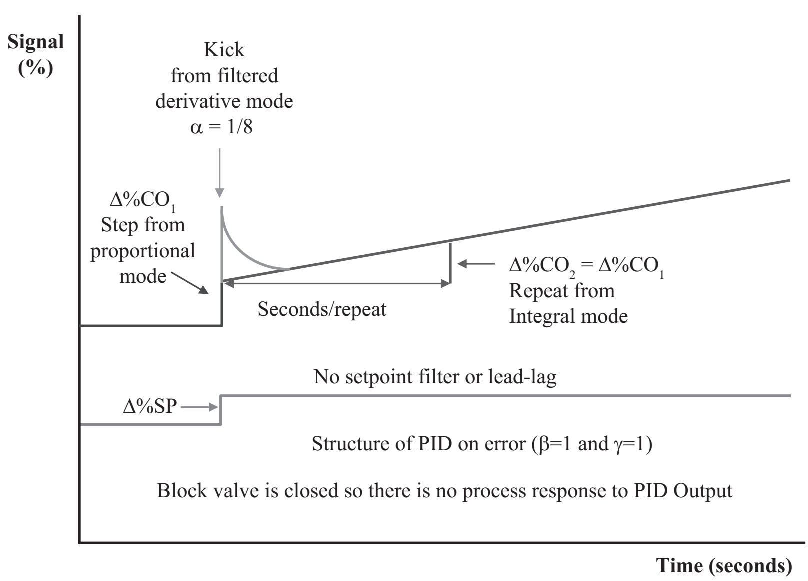


Figure 1.1. Contribution of each mode for a step change in the setpoint.

of the effect of each mode on the PID output without the complication of the process response. The filter on derivative action is one-eighth the rate time setting ($\alpha = 1/8$).

In Figure 1.1, the step change in the setpoint causes an immediate step change in the PID output from the proportional mode. A coincident kick from the derivative mode occurs but this contribution to the output decays out to zero. The kick is proportional to the slope of the setpoint change which is smoothed by the derivative mode filter set by α. The ramp in the PID output is from the integral mode. When the change in PID output from the integral mode is equal to the change in PID output, the integral mode has repeated the action of the proportional mode. All integral mode settings in industrial controllers have units of repeats per minute or second or the inverse, which are minutes or seconds per repeat. When the integral setting units are minutes or seconds per repeat, the *per repeat* is often dropped from the terminology.

If a block valve is not closed or a prime mover (e.g., pump, fan, or compressor) is not shutdown, the PV will start to respond after one dead time in response to the change in PID output. The contribution from the proportional mode will decrease as the PV approaches setpoint. The contribution from derivative mode will change sign as the error decreases. The contribution from the integral mode will continue to increase in the same direction and will not change sign until the error changes sign.

The general equations that dictate the contribution of each mode for an ISA Standard Form and a PID on error structure show that the controller gain is a multiplying factor for each mode. Many academic studies and some older controllers use a parallel structure described in Appendix K where the PID gain only affects the proportional mode.

$$\underline{\text{P}}\text{roportional (gain) } P = K_c * E \tag{1.1a}$$

$$\underline{\text{I}}\text{ntegral (reset) } I = K_c * \frac{1}{T_i} * \int E * dt \tag{1.1b}$$

$$\underline{\text{D}}\text{erivative (rate) } D = K_c * T_d * \frac{dE}{dt} \tag{1.1c}$$

where

E = error (PV minus setpoint) (%)
P = proportional mode contribution to PID output (%)
I = integral mode contribution to PID output (%)
D = derivative mode contribution to PID output (%)
K_c = controller gain (dimensionless)
T_i = integral time (reset time) (sec)
T_d = derivative time (rate time) (sec)

It is critical that the user understands the differences in the types of settings and units used for each mode. The settings also depend upon the Form of the PID algorithm discussed in Section 2.2.1. Appendix K shows how to convert settings between the various Forms and more information on the effect of tuning setting types and units. Lack of this knowledge can cause settings to be orders of magnitude too small or too large leading potentially dangerous operation.

Controllers may use proportional band (PB) in percent instead of a dimensionless controller gain (K_c). Controllers may use a reset setting (R_s) in repeats per minute or repeats per second instead of an integral time (T_i) also known as reset time in minutes or seconds per repeat. The time units for both the reset and rate settings can be minutes or seconds depending upon the supplier. Settings must be converted to the proper type and units. The following equations provide some guidance.

$$PB = \frac{100\,\%}{K_c} \tag{1.2a}$$

The conversion of integral time or reset time (seconds) to reset setting (repeats/minute):

$$R_s = \frac{60}{T_i} \tag{1.2b}$$

where

K_c = controller gain (dimensionless)
T_i = integral time (reset time) (sec)
PB = proportional band (%)
R_s = reset setting (repeats/min)

In this book, the proportional mode setting will be a dimensionless gain, the reset setting units will be seconds (seconds per repeat), and the rate setting units will be seconds.

The official definition of a lag is any phase lag that can be due to a dead time or a time constant. In most publications today, lag time is used interchangeably with time constant. In this book, lag will be used as a concise term in tables to denote time constant. Delay will be used as concise term for dead time.

Due to space and time limitations, we cover the highlights here in terms of PID modes and actions assuming a certain degree of familiarity with the PID. Please see Appendix B for a discussion of the basics of PID controllers including notable bullet points.

1.2.1 PROPORTIONAL MODE

The proportional mode provides an immediate reaction. For a step change in the PV or setpoint, the proportional mode gives a step change in output for a PID on error structure. Step changes in setpoint are quite common. Step changes in a measurement only occurs in fast processes from step changes in PID setpoints and outputs or from open and closing of on-off valves due to sequences or SIS. For slow processes where there is a large primary process time constant to slow down a process excursion (e.g., back mixed liquid volumes), step changes in measurements are isolated to the rare case of disturbances that occur on the process output rather than the process input.

Step changes in measurements are extremely disruptive. A step change in the PV does not give time for PID controller to respond and catch up with the deviation from setpoint. Most of the control literature shows step disturbances often on the process output. This is the worst case scenario.

For applications involving fast processes such as gas volumes and plug flow volumes commonly seen in the oil, gas, and petrochemical unit operations, step changes in the PID output are minimized. Consequently, the PID structure and tuning is chosen to provide a gradual change in PID output and a gradual approach to setpoint.

For applications involving slow processes such as mixed liquid volumes commonly seen in the specialty chemical business, step changes in PID output enable upper level loops for composition and temperature control to reach setpoint faster. An overshoot of PID output beyond the final resting value is seen as beneficial in terms of reducing batch cycle times and startup times. In contrast, overshoot of the PID output is rarely appreciated in fast processes.

Abrupt changes in the PID output afforded by the proportional mode also helps minimize the peak and integrated error for disturbances on the input to slow processes. Chapter 3 and Appendix C provide the equations that show the beneficial effect of greater proportional action (larger PID gain settings). For highly exothermic reactors, the aggressive immediate reaction helps prevent a runaway of the temperature response avoiding the activation of relief devices and SISs. The response of runaway processes starts out slow but accelerates if uncorrected.

Since the contribution from the proportional mode will decrease at a rate proportional to the approach of the PV to setpoint, the proportional mode is reacting to the rate of change that is the slope of the PV trajectory. Thus, the proportional mode provides preemptive action in changing the direction of the PID output change before the PV crosses setpoint. If the proportional action is much greater than the integral action, the PV will falter (momentarily hesitate) in the approach to setpoint.

To summarize, when the rate of change of the PV is slow, processes can benefit from rapid changes in the PID output and the anticipatory action of the proportional mode. For other processes, gradual changes in the PID output are sought by shifting feedback correction from the proportional mode to the integral mode for PV and setpoint changes.

Note that in industrial PID controllers, an increase in gain setting (decrease in PB) will proportionally increase the contribution from the integral and derivative modes as detailed in General Equations 1.1a to 1.1c for the modes.

1.2.2 INTEGRAL MODE

The integral mode provides a gradual action in a direction consistent with the sign of the error. The change in output does not change sign until the PV crosses setpoint. This action is more in tune with customer expectations promoted by the design of operator displays to date that presents digital values on graphics and faceplates.

Consider a reactor that is heating up with a temperature loop that is split ranged to add either steam or coolant to the jacket. The operator sees the present temperature is digitally displayed as 48°C and the setpoint as 52°C on the faceplate shown in Figure 1.2a or a graphic display of the reactor system. Which valve will the operator expect to be open? Based on the temperature value being below setpoint, the operator wants the steam valve to be open. If the number of decimal places is increased, some operators will want the steam valve to be open even if the temperature is just a few hundredths of a degree below setpoint (e.g., PV = 51.96°C). This view is shared by almost everyone including process engineers and control engineers who you would think would know better. I got a call from two different control rooms that there was something wrong with the PID algorithm because the coolant valve opened before the PV reached setpoint.

Adding more integral action will achieve what everyone wants. Integral action will seek to keep opening the steam valve until the temperature crosses setpoint. Unfortunately, waiting to open the coolant valve until the error changes sign is too late because any change in the PID

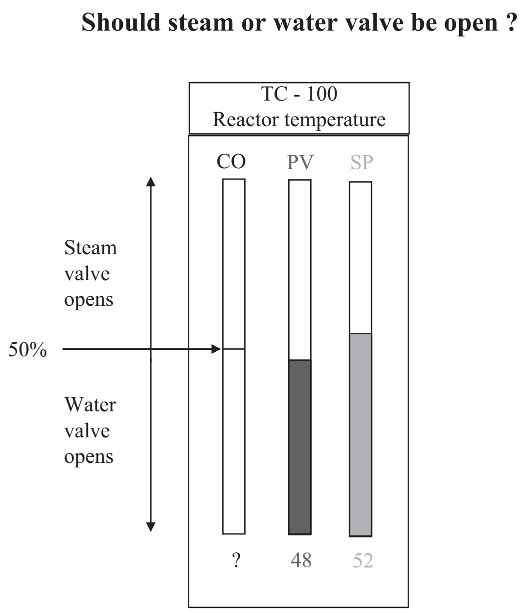


Figure 1.2a. Effect of operator interface and human expectations on tuning.

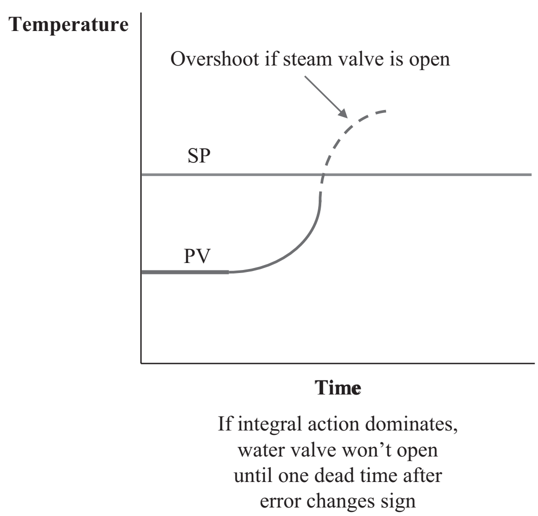


Figure 1.2b. Effect of trajectory and dead time understanding on tuning.

output does not affect the PV until after the total loop dead time. Furthermore, the contribution of the integral mode to open the steam not offset by the other modes must be undone before the split range point is reached so the coolant valve can open. Adding insult to injury, the transition is delayed and far from smooth due to the discontinuity from valve stiction that is greatest near the seat as one valve opens as the other valve closes and from changes in phases and operating conditions in the jacket. The result is a considerable overshoot of setpoint that increases as integral action is increased. For reactors, the overshoot can trigger undesirable side reactions or runaway conditions.

A trend display as shown in Figure 1.2b with an intelligent time scale and PV scale would show the operator the trajectory. A predicted future PV value would help educate the operator as to the effect of dead time and trajectory slope and value of patience. The computation of the future PV value is simple and robust and can be used to provide an intelligent adaptation of reset settings (Sections 5.3 and 12.4), a full throttle setpoint response (Section 2.2.8) and an improvement of batch operations (Section 15.2).

Integral action is needed to eliminate offset (persistent deviation between setpoint and PV for a single disturbance or setpoint change). Offset as a fraction of a disturbance or setpoint change is inversely proportional to the controller gain setting. For processes with a slow rate of change, the allowable controller gain can be quite large and offset correspondingly very small. However, the display of PV values to several decimal places can result in attempts to tune a controller to eliminate offsets within the noise band of the measurement or resolution limit of the valves. Attempts by integral action to eliminate the offset from stiction (stick-slip) and in some cases from backlash (deadband) result in a limit cycle (perpetual equal amplitude oscillation when there are no disturbances).

A case in point is the valve positioner. These positioners have traditionally been proportional-only controllers with a gain of 100 or more resulting in offsets of less than a percent. The modern positioner has added integral and derivative modes and the ability to monitor actual valve position to several decimal places. While this tuning capability offers flexibility to deal with difficult actuator and valve applications, the addition of integral action creates a limit cycle unless integral action is turned off when the valve position is within the resolution limit of the valve.

The intent here is not to make integral action a villain but instead to warn against excessive integral action for certain types of applications. For processes with a slow rate of change, the most common problem is a reset setting in seconds that is too small. For concentration, pH, level, or temperature control of liquid volumes or gas pressure control of large gas volumes, the reset time and gain is too small often by more than two orders of magnitude (e.g., reset time and gain should be increased by a factor of more than 100). Coming from the specialty chemical business, I am particularly sensitive to excessive integral action because it is the main tuning mistake for crystallizers, evaporators, columns, stirred neutralizers, and stirred reactors. The problem tends to be worse for the batch operation of these types of process equipment.

For processes with a fast rate of change, integral action is the preferred mode to reduce abrupt PID output changes from abrupt PV changes. For dead time dominant loops where the full change in PV is seen almost immediately after dead time, the integral action should be greatly increased and proportional action decreased approaching an integral only controller. This tuning provides a smoothness missing in the process and reduces the effect of noise and interactions. Dead time dominant loops often have a reset time that is too large. For concentration and temperature control of large plug flow liquid or gas volumes, the reset time and gain are too large often by nearly an order of magnitude (e.g., reset time and gain should be decreased by a factor of 10).

Subsequent sections discuss how a slow rate of change of process corresponds to a large process time constant or a small integrating process gain and that we can convert back and forth between these terms. In other words, a large process time constant can be expressed in terms of low integrating process gain and vice versa. Most of the disagreement in tuning stems from a misunderstanding of the effect of an actual or effective process time constant that is much larger or smaller than the loop dead time. Processes such as extrusion, pulp, paper sheet, and plastic sheet have plug flow dead time dominant responses. Oil, gas, and hydrocarbon processes often have a dead time about equal to the sensor time constant. Most specialty chemical processes have well-mixed liquid volumes where the dead time is much less than the process time constant. Even here a misunderstanding develops from looking at disturbances that arrive on a process output bypassing the process time constant causing an abrupt process response seen in the previously mentioned processes. This disagreement and different objectives have resulted in over a hundred different tuning rules. The authors of tuning rules are typically adamant about the value of their rules. There is not enough room in this book to present all the rules. The book presents Lambda, Internal Model Control (IMC), quarter amplitude, a Short Cut Method (SCM), and Ziegler–Nichols ultimate oscillation and reaction curve tuning rules with modifications to help deal with the diversity of applications.

Integral action can be effectively removed by setting the reset time to a very large setting (e.g., 10,000 sec). For controllers with a reset setting in repeats per minute or second, a reset setting of zero may not turn off integral action. The rule for what happens for a zero reset or gain setting varies with supplier and vintage leading to unexpected behavior. A different Structure (see Section 2.2.2) should be specified to turn off the proportional and integral modes eliminating potentially dangerous situations. The Form and Structure is not generally specified in studies on tuning making it difficult to compare results. Chapter 2 will explain how the use of proportional and derivative mode acting only on PV and integral on error (PD on PV and I on error) and two degrees of freedom (2DOF) structure will eliminate the overshoot. The 2DOF can provide a faster approach to setpoint but requires the adjustment of two factors (beta and gamma). Test results in Chapter 2 will show the effect of the factors in a 2DOF structure.

1.2.3 DERIVATIVE MODE

The derivative mode provides an abrupt action in a direction consistent with the sign of the rate of change of the PV. The derivative mode is most productive when the rate of change of the PV is slow. The derivative mode provides anticipatory action and helps prevent overshoot. The derivative mode can also help cancel out a secondary time constant (second largest time constant) often associated with a thermal lag in heat transfer surfaces. In older controllers with the "Series" Form described in Section 2.2.1, an increase in the rate time automatically resulted in less proportional and integral action. This prevented the derivative time from becoming larger than the reset time that could result in instability. The use of the "ISA Standard" Form in newer controllers eliminates this inherent protection. Keeping the rate time less than the reset time becomes the responsibility of the person tuning the controller.

Too much derivative action can result in a relatively fast oscillation as the PV approaches setpoint. The derivative mode will amplify noise and discontinuous responses from at-line analyzers with large cycle times and wireless transmitters with large reporting time intervals (default update rates). If the primary and secondary time constants are not much larger than the dead time, the benefit is marginal. Plus, rate time is a third adjustment that must be coordinated with the other adjustments in the "ISA Standard" Form. For these reasons, the use of derivative action is not widespread. Fortunately, derivative action can be turned off by simply setting the rate time equal to zero. Note that the term PID is still used in the literature even when the rate setting is zero resulting in a PI controller.

Here we summarize the major reasons why practitioners do or do not use derivative action.

How can derivative action be beneficial?

1. Canceling the effect of a secondary time constant that is more than half the dead time in a second order plus dead time approximation enabling tighter control.
2. Helping to counteract integral action to suppress overshoot.
3. Achieving a faster response for small setpoint changes by zipping the PID output through deadband from valve backlash and resolution limits from stiction.
4. Counteracting runaway conditions by reacting to the acceleration of the PV.

How can derivative action be detrimental?

1. Passing on abrupt PV changes as abrupt PID output changes in loops with large transportation delays, wireless transmitters with slow default update rates, and at-line analyzers with large cycle times.
2. Amplifying measurement noise.
3. Decreasing the signal to noise ratio for temperature control in older Distributed Control System (DCS) with 12 bit analog to digital (A/D) and wide range thermocouple input cards.
4. Increasing the wrong direction of the PID output change for an inverse response.

1.2.4 ARW AND OUTPUT LIMITS

All controllers must prevent the output from being too low or too high. High gain settings and any use of the integral mode can result in outputs going below zero percent or above 100 percent. Industrial controllers limit the contribution of the modes to be above a low limit and below a high limit. Typically, the contribution from the integral mode is first reduced so the total contribution from all modes does not go outside the limits. The algorithm is often called ARW and is a proprietary feature. The preferential reduction of the integral mode enables the controller output to come off the output limit sooner when the PV starts to approach the setpoint. How close the PV gets to setpoint before the output comes off the limit determines whether the PV will falter or overshoot the setpoint. Ideally, the contribution from the integral mode would be limited so that the output would come off the limit when the PV is within a specified margin of the setpoint. The optimal degree of limiting can be computed from a rise time (time to reach margin) estimated from the rate of change of the PV (see Section 5.3). The margin that would prevent overshoot can be computed if the open loop integrating process gain is known (see Section 1.3.3).

The PID output can come off the output limit too soon for a high gain PID with a fixed setpoint when switched from manual (MAN) to automatic (AUTO). In this case, as soon as the PV starts to approach setpoint in a startup or batch operation, the output will come off of its limit. For example, after switching an evaporator level controller from MAN to AUTO to fill up the evaporator, on startup the feed control valve would immediately start to close when the level started to increase from zero percent. The solution is to have the setpoint track PV and change the setpoint to the desired operating point after switching the mode to AUTO.

Some controllers have ARW limits besides output limits. One major manufacturer increases the reset action by a factor of 16 when the PID output is within the ARW limits and output limits. This capability was principally developed to help control valves with pneumatic positioners close by quickly going to a negative output. The offset and drift in the calibration of pneumatic positioners and lack of position readback required that the valve signal be one to two milliamps direct current (madc) below the four madc that was the zero percent signal to make sure the control valve was closed. The 16 × reset action helped the PID output quickly move through this region of uncertainty. The low ARW feature also helped surge valves with high break away torques, open in time to prevent surge. Some tight shutoff valves required a PID output of 15 percent before they would break free of the seal or seat. In this case the low ARW limit was set at 15 percent and the low output limit was set at −10 percent for pneumatic positioners and zero percent for digital positioners.

A less common application occurred when on heat-up of a vessel the temperature controller output needed to decrease steam flow as soon as possible to prevent hot spots when the temperature approached setpoint without increasing batch cycle time. The high ARW and output limits were set at 90 percent and 100 percent, respectively.

If ARW limits exist, they should be set equal to the output limits for most applications. The limits in older DCS were typically in percent and in newer DCS they are often in engineering units of the PID output scale. The default values for limits may be in percent that is totally inappropriate for the PID with output scales in engineering units.

1.2.5 CONTROL ACTION AND VALVE ACTION

If the control action and valve action is not right, nothing else matters. The PID will ramp off to an output limit. If the valve action is set correctly, the correct control action is the opposite of the process action to provide the negative feedback correction. If the process action is *direct*, where an increase in the manipulated variable (e.g., flow or lower loop setpoint manipulated by the PID output) causes an increase in the controlled variable for the PID (PV that is the PID input to be controlled at setpoint), the control action is *reverse*, so an increase in PV causes a decrease in PID output. If the process action is *reverse*, the control action must be *direct*.

The complicating factor is valve action for a fail open valve that is *increase to close* where an increase in the signal to the valve moves the valve towards the closed position. In order for the control action to stay being the opposite of the process action, the PID output signal must be reversed. The signal can be reversed in the field in the current to pneumatic transducer (I/P) or in positioner. The signal can also be reversed in the control system. In a modern DCS, the signal can be reversed in the PID block or AO block by selecting the *increase to close* option or in the setup of a signal characterizer block or split range block. The problem for the practitioner is to know if the valve action is set only once and where the valve action is set. If the valve action is not set or is set multiple times, the control action must be changed accordingly. While smart devices have increased the visibility of configuration of these devices in the field, a valve action set in the DCS configuration in the PID module can be more readily reviewed and tested by virtual plant simulations to verify if the correct valve action and control action have been set.

Tests by stroking the valve in the field from the control system should be used to verify if the valve action is correct. A process engineer should be present for the test and ascertain the process action and control action is correct before the loop is commissioned.

1.2.6 OPERATING MODES

The modern PID has many different modes. These modes are often switched based on operating states. Most loops are in the MAN mode when the unit operation is shutdown or is not at operating conditions that provide a representative measurement for feedback control by the PID. The following modes using Fieldbus standard terminology are available in a modern PID. In general, the transition between modes is bumpless.

In the MAN mode, the PID is no longer doing feedback control. The PID output is fixed unless changed by the operator. Nearly all loops must enable the MAN mode since in general you should never take away the capability for operations or maintenance to fix a PID output. To provide coordinated actions for batch operations and startup or preemptive actions for SIS, equipment protection and prevention of environmental violations, a controller may be momentarily forced out of MAN. If an algorithm needs to write to a PID output, the ROUT remote output (ROUT) mode should be used.

In the ROUT mode, the PID output is set as part of a sequence by a batch control system, sequential function chart, or a calculation. This mode provide an intelligent means to deal with abnormal operation, batch, startup, product transitions, and measurements not representative of the process. Often controllers are operated in MAN because operators feel they need to intervene to deal with situations. The ROUT mode can implement the best of the operator responses and process expertise on a more recognizable and repeatable basis allowing for continuous improvement. The effective use of this mode can dramatically reduce the time loops are in MAN and the startup and batch cycle time and improve on-stream time and yield. Sections 2.2.8 and 5.2 will provide the details to effectively use this mode.

The output tracking function can also be used to set the PID output to an externally computed signal. In this case, a track discrete is turned on and off. The use of the ROUT mode is preferred in terms of visibility and the ability of the operator to change the mode to MAN if necessary for abnormal operation or maintenance.

In the AUTO mode, the PID is getting a local setpoint from the operator. While a calculation can write to this local setpoint, the loss of the ability to change the setpoint is confusing to the operator. There is a back calculate input (BKCAL_IN) analog and status bit input signal to coordinate the PID output with the mode and output of blocks that the PID output signal goes through, to prevent the AUTO mode if there is an incorrect mode or failure. The BKCAL_IN also enables a bumpless transfer to AUTO.

In the cascade (CAS) mode, the PID is receiving a remote setpoint from another PID to form a CAS loop. As with the AUTO mode, there is a BKCAL_IN to suspend feedback control when downstream blocks are not functional and to provide a bumpless transfer of the PID output to a best value when these blocks show a "good" status.

In this book, we will use the term *upper loop* for a master, primary, or outer loop and *lower loop* for a slave, secondary, or inner loop in a cascade control system. There can be multiple levels of cascade control. We have a quadruple cascade control system for a reactor temperature PID output that is the setpoint of jacket temperature PID whose output is the setpoint of a coolant flow PID whose output is the setpoint for a digital valve controller (DVC). The option often exists for the upper loop output to track the setpoint or PV of the lower loop until the upper loop goes into AUTO and the lower loop goes into CAS to provide a smooth transition to cascade control.

Many PIDs offer the option of feedforward being active when in MAN. The operator should still have the option of operating manually. If the feedforward signal is the setpoint for a flow loop, the operator can take the flow loop out of CAS mode and put the flow loop in AUTO or MAN to automatically or manually set the flow or valve, respectively.

In the remote cascade (RCAS) mode, the PID is receiving a remote setpoint from a MPC, batch control system, or a supervisory computer. As with the AUTO mode, there is a BKCAL_IN to suspend feedback control when downstream blocks are not functional and to provide a bumpless transfer of the PID output to a representative downstream signal when these blocks show a “good” status.

1.3 LOOP DYNAMICS

The open loop response is the dynamic change in the PV (%ΔPV) to a step change in the controller output (%ΔCO). *Open loop* used to designate the response does not include feedback control and thus shows the reaction of the process and not the controller. An open loop response can be created by a change in PID output when the PID is in the MAN or ROUT modes. The open loop response is the basis for determining the tuning and capability of the loop. The open loop response is described by the values of three terms; open loop gain, time constant and dead time. The dead time term is defined the same way for all types of processes as the time for the PV to get out of the noise band after a change in PID output. The definition of the open loop gain and time constant depends upon the type of feedback within the process.

1.3.1 TYPES OF PROCESS RESPONSES

There are three fundamental types of processes depending upon the sign of feedback within the process. The sign of the feedback is apparent from whether the response approaches a final value, continues to ramp, or accelerates in the latter portion of the open loop response. The ordinary differential equations for the process documented in Appendix F show the first principle origin of the process feedback sign and the terms that define the corresponding open loop response.

The most common type of process has negative feedback. These processes are self-regulating in that they will line out at a steady-state value as shown in Figure 1.3a for a given step change in PID controller output assuming there are no process disturbances. Liquid flow, liquid pressure, continuous concentration, and open loop stable continuous temperature processes have a self-regulating response.

The primary time constant is the time to reach 63 percent of the final change in PV value (%ΔPV) after the PV starts to change for a given change in the controller output (%ΔCO). The primary time constant is the largest time constant in the loop. Ideally, the largest time constant is in the process but, for liquid flow, pressure, and some inline processes, the largest time constant originates in the sensor, transmitter, or PID filter. Some software can identify the second largest time constant that is called a secondary time constant. This identification can help in the tuning of controllers that use derivative action because setting the rate time equal to the secondary time constant cancels out its detrimental effect in slowing down closed loop control (control when PID is in AUTO, CAS, or RCAS modes). If the process response had no secondary time constant, the slope of the response would be at its maximum value immediately after the dead time. The secondary time constant creates a gradual increase (e.g., bend) in the rate of change

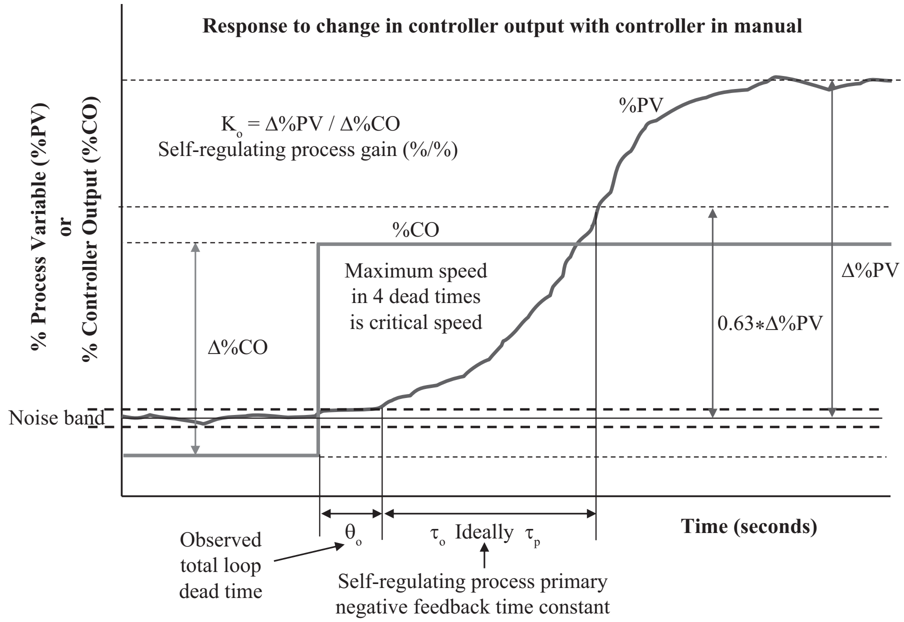


Figure 1.3a. Dynamics for a negative feedback process (self-regulating process).

of the process after the dead time. All processes have a secondary time constant. If this time constant is not identified, the secondary time constant is converted to an equivalent dead time as per the equations in Section 4.1 to provide a first order plus dead time (FOPDT) model. If the secondary time constant is identified, we have a second order plus dead time model.

If the primary time constant is less than one-fourth the total loop dead time, we have a sub category of a self-regulating process called a dead time dominant process. The tuning of dead time processes requires significantly more integral and less proportional action. Derivative action is not used. If the process time constant is greater than four times the total loop dead time, we have a subcategory of a self-regulating process called near-integrating. The tuning for near-integrating processes is the same as that for true integrating processes which have zero internal process feedback. This is based on the realization that a PID does most of its correction in the first four dead times during which the process has not appreciably decelerated. Concentration, pH, and temperature control of continuous unit operations with back mixed liquid volumes have a near-integrating response.

When the time constant is between one-fourth and four times the dead time, we have a moderate self-regulating process. Most of the control literature on the design and implementation PID, MPC, or special control algorithms are addressing moderate self-regulating processes.

The open loop steady-state gain is dimensionless and is the final change in percent PV for the step change percent change in PID output. Note that the changes in PV must be in percent of input and output scales because the PID algorithm works in percent and not engineering units.

The next most common type of process has zero internal process feedback. These processes are integrating and will ramp continually off to an equipment or process limit as shown in Figure 1.3b for a given step change in the PID output. Level, gas pressure, batch concentration, batch pH and batch temperature processes have an integrating response.

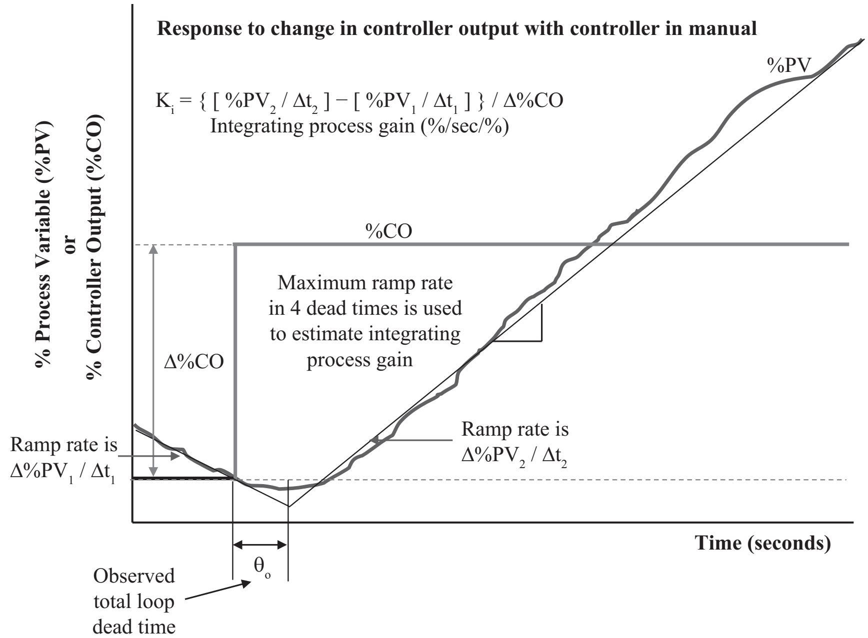


Figure 1.3b. Dynamics for a zero feedback process (integrating process).

The key characteristic of a zero feedback process is an open loop integrating process gain that is the change in ramp rate of the PV in percent per second ($\%\Delta PV_2/sec - \%\Delta PV_1/sec$) divided by the step change in percent controller output (%ΔCO) assuming there are no process disturbances. If the PID was AUTO, CAS, or RCAS modes immediately before the test, the initial ramp rate is typically zero ($\%\Delta PV_1/sec = 0$). The ramp rate of most integrating processes is extremely slow (e.g., 0.000001 %/sec).

We will see in Section 1.3.3 that the open loop integrating process gain can be converted to a primary time constant and open loop steady-state gain of a near-integrating process and vice versa. The conversion is useful for gaining a better understanding and set of tuning settings. In fact there are few really true integrating processes. If there were no equipment, process, or measurement limits, nearly all integrating processes would reach a steady state. For example, if the liquid level or gas pressure could increase without bounds, the additional head or pressure would increase the liquid flow and gas flow out of the vessel to balance the increase in flow into the vessel.

An appreciable time constant in a true integrating process is called a secondary time constant. The secondary time constant causes the process response to slowly accelerate to reach the maximum rate of change determined by the open loop integrating process gain. The identification and mitigation by the use of a rate time equal to the secondary time constant is more important in integrating processes than in moderate self-regulating processes. This definition of secondary time constant is consistent for all three major types of processes enabling the term primary time constant to be used to provide the conversion between a true integrating and near-integrating response.

The product of the PID gain and reset time must always be greater than a minimum established by the open loop integrating process gain defined by Equation 1.5e to prevent slow rolling oscillations. Most integrating processes have a gain setting much lower than permitted by the process dynamics. Since people associate oscillations with too high of a gain and decrease the gain making the problem worse when the solution was to either increase the gain or the reset time. Most integrating processes have too much integral action. The first thing to try in most cases if the period of the oscillation is more than 20 times the dead time is to simply increase the reset time by a factor of 100 to 1,000.

A process with positive feedback fortunately does not occur very often but is associated with some extremely important unit operations. These processes have an open loop runaway response where the PV will continue to accelerate as shown in Figure 1.3c for a given step change in the PID output. These processes are called open loop unstable.

The most common example is an exothermic reactor where the heat of reaction increases with temperature and can exceed the cooling rate if the temperature gets high enough. I have also experienced a runaway speed condition when a large compressor went into surge but this axial compressor may not be typical in that it had extremely low rotor inertia. The cell mass concentration control of a biological reactor will exhibit a positive feedback response in the exponential growth phase where an increase in cell concentration increases the cell growth rate. A runaway condition is not typically a concern because the acceleration in the cell mass concentration is so slow and is followed by a stationary phase where the acceleration stops due to the death rate of older cells equaling the growth rate. The pH response of a true strong acid and strong base neutralizer will dramatically accelerate as the pH approaches neutrality (the rate of change of pH increases by a factor of 10 for each pH unit in the approach to neutrality). Fortunately most systems are moderated by slight concentrations of weak acids and weak bases and most notably carbonic acid associated by exposure to what might seem as insignificant concentrations of carbon dioxide.

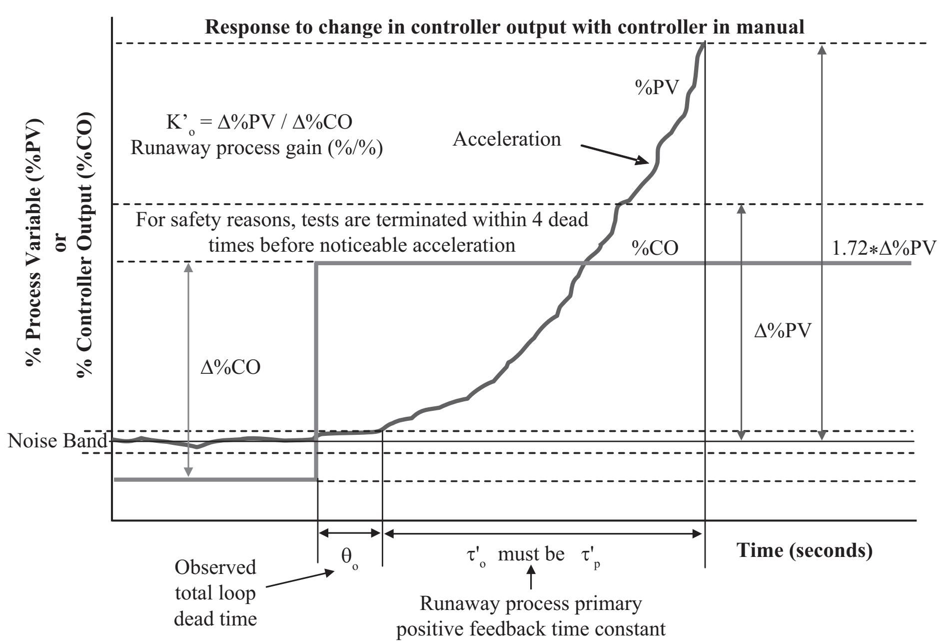


Figure 1.3c. Dynamics for a positive feedback process (runaway process).

The primary process time constant for a process with positive feedback is the time for the PV to reach 172 percent of the value established by the open loop runaway process gain. This gain can be computed as shown in Appendix F but roughly corresponds to the point where a noticeable continual acceleration becomes significant. The identification of the positive feedback time constant and open loop gain by an open loop test is rarely done in practice because of the danger of a runaway. Often these loops have to stay in AUTO mode at all times so that the PID controller provides through its control algorithm the negative feedback missing in the process that is needed for safe operation.

Runaway processes are identified as integrating processes. Even when a runaway process exists, tests are designed not to show the acceleration. The tuning must be aggressive in that the process will be a runaway if the PID gain is too small not providing enough proportional action. In Sections 4.5.3 and 6.8, equations are presented to compute the maximum and minimum gain for stability as a function of dead time, the primary time constant, and the secondary time constant. Integral action is minimized because the integral mode will not seek to add coolant until the PV has crossed setpoint on a temperature rise. Derivative action is more important than in other processes in order to help prevent excessive acceleration and to compensate for a secondary time constant.

Figure 1.4 plots a dead time dominant, near-integrating, true integrating, and runaway process response on the same time scale. The dead time in all cases was 10 seconds and the secondary time constant was zero seconds. A concurrent step change in the PID output of 10 percent occurred at 30 seconds into the test. The responses on the same trend chart reveal the fundamental ability to identify a ramp rate and open loop integrating process gain (0.01% PV per second per percent change in controller output) at the beginning of the near-integrating and runaway processes that enable these processes to be modeled and tuned as true integrating processes.

1.3.2 DEAD TIMES AND TIME CONSTANTS

Dead time is always detrimental in that it delays the recognition of a change in the PV or the correction of the manipulated flow.

The total loop dead time is the sum of all pure delays and all time constants smaller than the primary time constant converted to equivalent dead time per equation in Section 4.1.1 in a FOPDT model. In a second order plus dead time model, the secondary time constant is taken separately and not converted to equivalent dead time. As seen in Figure 2.4 in Section 2.3, the delays and time constants can be in the final control element (e.g., valve or variable speed drive), process, measurement, and PID controller. To maximize loop capability, the secondary time constant and dead time must be minimized and the primary time constant must be in the process and maximized. If the primary process time constant is in the process, it will slow down the excursion from unmeasured disturbances at the process input (load disturbances). The maximum PID gain is roughly proportional to the ratio of the primary time constant to dead time in a FOPDT model.

If there was only a primary process time constant (no dead time, no secondary time constant, and no noise), perfect control would be possible and there would be no upper limit to the controller gain for stability. Without dead time, I would be out of job. I have job security because zero dead time systems only exist in some models and academic studies.

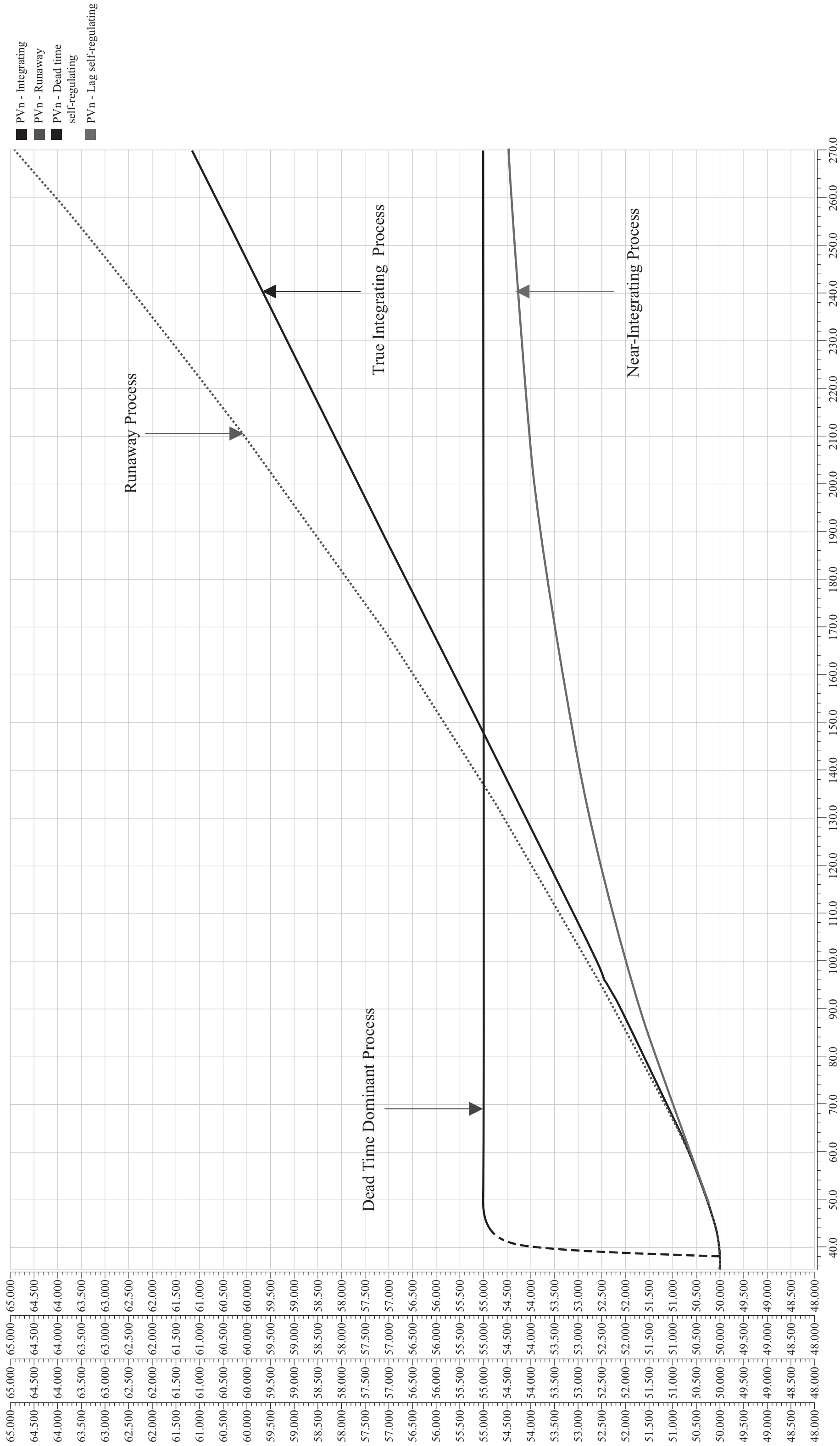


Figure 1.4. Simultaneous step responses of major types of processes.

1.3.3 OPEN LOOP SELF-REGULATING AND INTEGRATING PROCESS GAINS

The open loop gains define what the PID sees as a percent change in PV (Δ%PV) for a percent change in controller output (Δ%CO). The open loop gain is the product of the valve or VSD gain, a flow ratio gain, process gain, and measurement gain. In the literature, this open loop gain is simply referred to as the process gain with considerable inconsistency as to units. This book uses the term open loop gain to provide better recognition of the effect of different parts in the control loop and to ensure an open loop gain that is dimensionless for self-regulating processes and 1/sec units for integrating processes. The ability to segment the contributing factors enables the calculation of the effect of changes in valve or VSD capacity and installed flow characteristics, types of processes and equipment, and transmitter calibration span.

The flow ratio gain details an often overlooked contributing factor for the process and equipment gain that is inversely proportional to feed flow for continuous composition, pH, and temperature control. The PV for these processes is plotted versus a ratio of manipulated flow to feed flow. For small disturbances, the gain from the process and equipment is the slope of this curve at the setpoint divided by the feed flow. This ratio of manipulated flow to feed flow is the key to understanding the benefit of flow ratio control (flow feedforward).

The open loop steady-state gain for a self-regulating process is the final change in the percent PV input ($\Delta\%PV$) after the process has reached a new steady state for a step change in the percent controller output ($\Delta\%CO$).

$$K_o = \frac{\Delta\%PV}{\Delta\%CO} \tag{1.3a}$$

The open loop steady-state gain (K_o) is the product of the gains for the valve or VSD (K_v), flow ratio (K_r), process (K_p), and measurement gains (K_m). Usually the product of the flow ratio gain and process gain defined here is simply expressed as a process gain.

$$K_o = K_v * K_r * K_p * K_m \tag{1.3b}$$

For small changes in flow, the valve or VSD gain is the slope of the installed flow characteristic near the current operating point that is the change in flow in engineering units (ΔF_v) divided by the change in controller output ($\Delta\%CO$).

$$K_v = \frac{\Delta F_v}{\Delta\%CO} \tag{1.3c}$$

For composition, temperature, and pH loops, the flow ratio gain (K_r) is inversely proportional to the main process feed flow (F_p).

$$K_r = \frac{\Delta R}{\Delta F_v} = \frac{\Delta F_v / F_p}{\Delta F_v} = \frac{1}{F_p} \tag{1.3d}$$

The process gain is the change in PV in engineering units (ΔPV) divided by the change in ratio (ΔR).

$$K_p = \frac{\Delta PV}{\Delta R} \tag{1.3e}$$

For flow and pressure loops, the product of the flow ratio and process gain simply becomes the change in PV per change in flow (both in engineering units).

$$K_r * K_p = \frac{\Delta PV}{\Delta F_v} \tag{1.3f}$$

The measurement gain (K_m) is 100 percent divided by the calibration span of the transmitter.

$$K_m = \frac{100\%}{\left(PV_{\max} - PV_{\min}\right)} \tag{1.3g}$$

For the upper (primary) loop in a cascade control system, the PID output becomes the setpoint of a lower (secondary) loop rather than the setpoint of a valve or VSD. In Equation 1.3b, the setpoint scale gain for the manipulation of a setpoint is substituted for the gain for the manipulation of a valve or variable speed drive. The nonlinearity of the valve or VSD installed flow characteristic has been eliminated from the view of the upper loop. The gain for manipulation of a setpoint is computed as detailed in Equation 1.3i as the span of the setpoint scale of the lower loop divided by 100 percent. The setpoint scale normally matches the measurement scale of the lower loop. If the DCS output scale is in percent, this match happens automatically. If the PID output scale is in engineering units, the user is responsible for setting the upper PID output scale to match the lower PID scale. The most common lower (secondary) loop is flow in which case there is still the flow ratio gain (K_r) for composition, pH, and temperature loops.

For cascade control where an upper composition, pH, or temperature loop is manipulating a lower flow loop, the valve or VSD gain is replaced with a setpoint gain in the open loop steady-state gain giving:

$$K_o = K_{sp} * K_r * K_p * K_m \tag{1.3h}$$

The setpoint gain is the span of the lower loop setpoint scale divided by 100 percent:

$$K_{sp} = \frac{\left(SP_{\max} - SP_{\min}\right)}{100\%} \tag{1.3i}$$

An important higher cascade control loop is an upper vessel temperature loop manipulating the setpoint of a coil or jacket. In this case, the flow ratio gain has been eliminated. Also, the process gain simplifies to about 1.0 giving Equation 1.3j for a heating or cooling of a batch vessel when there is no reaction or feed. Equation 1.3j is Equation F.6h in Appendix F with a zero feed flow and zero heat of reaction. The simplification and linearization of the open loop gain is a significant advantage in addition to the compensation of cooling water or steam pressure or temperature changes by the coil or jacket loop before they appreciably affect vessel temperature.

For cascade control where an upper vessel temperature loop is manipulating the lower coil or jacket temperature loop, the open loop gain simplifies to the product of the setpoint and measurement scale gains for zero feed flow and zero heat of reaction yielding:

$$K_o = K_{sp} * K_m \tag{1.3j}$$

Note that an important check is that the open loop steady-state gain is dimensionless.

For integrating processes, the open loop self-regulating (steady-state) gain (K_o) is replaced with an open loop integrating process gain (K_i) with inverse time units (1/sec) that results from the process gain term units not cancelling out the flow time units as seen in Equation 1.4a through 1.4c for level. For level we can eliminate the ratio gain term defined by Equation 1.3d in the open loop gain (K_o) for a self-regulating process giving an equation for the integrating process gain (K_i) with the distinction of no longer being dimensionless but having units of 1/sec.

$$K_i = K_v * K_p * K_m \tag{1.4a}$$

$$K_v = \frac{\Delta F_v}{\Delta \% CO} \tag{1.4b}$$

$$K_p = \frac{1}{A * \rho} \tag{1.4c}$$

$$K_m = \frac{\Delta \% PV}{\Delta L} \tag{1.4d}$$

You can convert an open loop self-regulating process (steady-state) gain (K_o) for a process with a primary time constant (τ_o) much larger than the dead time (θ_o) to an open loop integrating process gain (K_i) by the use of Equation 1.5a developed for the near-integrator approximation to enable you to use tuning rules for integrating processes to improve the disturbance rejection. The equation can also save a huge amount of time in the identification of the process dynamics since the largest ramp rate $Max(\Delta\%PV/\Delta t)$ within four dead times in the right direction is used per Equation 1.5b to compute the integrating process gain instead of waiting for the process to essentially reach a steady state (e.g., 98 percent response time) to identify the primary time constant and open loop steady-state gain. For a time constant that is 20 times the dead time (not uncommon for well-mixed liquid reactors), the identification time is 20 times faster.

$$K_i = \frac{K_o}{\tau_o} \tag{1.5a}$$

$$K_i = \frac{Max(\Delta\%PV / \Delta t)}{\Delta\%CO} \tag{1.5b}$$

For integrating processes, the product of the controller gain (K_c) and integral time (reset time) setting (T_i) must be greater than twice the inverse of the open loop integrating process gain (K_i) to prevent the start of slow rolling oscillations.

$$K_c * T_i > \frac{2}{K_i} \tag{1.5c}$$

Since most PIDs on integrating processes have a controller gain much less than the maximum allowed, the equation is reformulated to show the minimum integral time.

$$T_i > \frac{2}{K_c * K_i} \tag{1.5d}$$

The oscillations will decay slowly unless the following inequality is enforced:

$$T_i > \frac{1}{4 * K_c * K_i} \tag{1.5e}$$

Recent test results show that the above rules apply to near-integrating and runaway processes as well. As you go from near-integrating to true integrating to runaway processes, the consequences of violating these rules get more severe in that the oscillations are larger and slower to decay. These rules also give a minimum PID gain for a window of allowable gains.

Nomenclature for PID Gain and Reset Time Calculations:

A = cross sectional area of vessel (m^2)
K_c = controller gain (dimensionless)
K_m = measurement gain (%/PV e.u.)
K_i = open loop integrating process gain (%/sec per % delta PID output) (1/sec)
K_o = open loop self-regulating process (steady-state) gain (%/%) (dimensionless)
K_p = process gain (PV e.u.)
K_r = flow ratio gain often imbedded in process gain (1/flow e.u.)
K_v = valve or VSD gain (flow e.u./%)
$\Delta\%CO$ = change in controller output converted to percent of controller output scale (%)
ΔF_v = change in valve or VSD flow (flow e.u.)
F_p = process flow at current production rate (flow e.u.)
ΔL = change in level (m)
$\Delta\%PV$ = change in PV converted to percent of controller input scale (%)
$PV_{\max}$ = maximum PV scale value (PV e.u.)
$PV_{\min}$ = minimum PV scale value (PV e.u.)
ΔR = change in ratio of manipulated flow to process flow (dimensionless)
$SP_{\max}$ = maximum setpoint scale value (SP e.u.)
$SP_{\min}$ = minimum setpoint scale value (SP e.u.)
T_i = integral time (reset time) (sec)
τ_o = primary time constant (largest time constant in loop) (sec)
ρ = liquid density in vessel (kg/m^3)

1.3.4 DEADBAND, RESOLUTION, AND THRESHOLD SENSITIVITY

In Chapters 6 and 7, which are based on the effect of measurements and valves, respectively, we will quantify the effects of deadband, resolution, and threshold sensitivity limits. These dynamic terms increase the dead time in control systems by the amount of time it takes for a PID output to change enough to make the valve, VSD, or measurement to respond. For a ramping controller output, the additional dead time is the resolution or threshold sensitivity limit or half of the dead band divided by the rate of change of PID output. Tests that make step changes in the PID output or setpoint do not show the additional dead time when the step change in signal exceeds the resolution or threshold sensitivity limit, or half of the dead band. The dead time increases as the tuning is slowed down (PID gain is decreased, reset time is increased, and rate time is decreased). The effect is mostly in terms of increasing the peak and integrated errors quantified in Chapter 3 from unmeasured disturbances at the process input (load upsets). The effect of disturbances on the process output or the rise time for setpoint changes is generally negligible unless the disturbances or setpoint changes are small and mostly integral action is used.

There is a beneficial effect of very small threshold sensitivity setting in the PID in terms of ignoring noise and allowing a large controller gain. When used in conjunction with the enhanced PID described in Section 2.2.10, the turning off of integral action for insignificant changes eliminates limit cycles from valve backlash and stiction and the need to retune the PID for large wireless default update rates or at-line analyzer cycle times. The threshold sensitivity would be configured into the PV signal and set slightly larger than the noise or A/D resolution (0.004 percent for a 16 bit A/D).

1.4 TYPICAL MODE SETTINGS

Table 1.1 is offered as an orientation. Tuning settings should be based on actual tests using software that can identify the loop dynamics. A great misuse of this table is to use the typical settings and not actually tune the PID for the application. The first setting outside of the parentheses is a typical value and the settings inside the parentheses give a typical range of values. A tuning setting that does not fit within the range is not necessarily wrong but should be reviewed for extenuating circumstances such as an incorrectly sized valve, incorrectly sized equipment (e.g., excessive or insufficient heat exchanger area), fouled equipment, or an exceptionally wide measurement calibration range compared to the normal process operating range. For near-integrating and true integrating processes, the PID gain used is typically much less than the maximum permitted per dynamics. To prevent the slow nearly undamped oscillations from violating the limit expressed in Equation 1.5e, the reset time must be accordingly set much larger than expected.

Table 1.1. Typical PID settings

Application type	Scan time (sec)	PID gain	Reset time (sec)	Rate time (sec)
Liquid flow	0.5 (0.2–1)	0.3 (0.2–0.8)	6 (1–12)	0 (0–2)
Liquid pressure	0.1 (0.02–0.5)	0.5 (0.2–1.0)	3 (1–6)	0 (0–2)
Tight liquid level*	5 (1.0–30)	10 (5–25)	600 (120–6,000)	0 (0–60)
Gas pressure (psig)*	0.2 (0.1–0.5)	5.0 (0.5–20)	300 (60–600)	3 (0–30)
Reactor pH*	1 (1.0–5)	1 (0.01–50)	120 (60–600)	30 (6–30)
Neutralizer pH*	2 (1.0–5)	0.1 (0.001–5)	300 (60–600)	60 (6–120)
Inline pH	1 (0.2–2)	0.2 (0.01–0.4)	30 (15–60)	0 (0–3)
Liquid reactor temperature*	5 (2.0–15)	5.0 (1.0–50)	600 (300–3,000)	120 (60–300)
Gas reactor temperature	1 (0.5 –2)	1.0 (0.5–5)	15 (5–30)	3 (1–6)
Inline liquid temperature	2 (1.0–5)	0.5 (0.2–2)	30 (15–60)	6 (3–15)
Distillation column temperature*	10 (2.0–30)	1.0 (0.2–10)	1200 (600–12,000)	120 (60–1,200)

*Denotes near or true integrating process where low PID gain will require a high reset time to prevent violation of the limit per Equation 1.5e and nearly undamped oscillations.

1.5 TYPICAL TUNING METHODS

The incredible range and impact of the size of the primary time constant to the dead time and the location of the disturbance have led to considerable disagreement. A switch to integrating process tuning rules when the primary time constant is much larger than the dead time and the realization most disturbances occur as a process input as load upset result in a convergence of rules when the objective is minimization of integrated error. Often the difference in factors settings quibbled about are less than the estimation error and variability in the dynamic terms (gains, dead time, and time constants). The claims by authors of the more than 100 tuning methods are often more a result of gamesmanship rather than reality. Some conclusions are based on tests where the dynamics are changed to make the control less stable. The most common test is to show how the Ziegler–Nichols method causes excessive oscillation for a 25 percent increase in process gain. The solution could have been to reduce the PID gain by 25 percent. The other test is to show excessive overshoot for a setpoint change. The solution would have been to simply add a setpoint lead-lag with the lag time equal to the reset time and the lead time equal to 25 percent of the lag time or use two degrees of freedom structure to eliminate the overshoot and still get to setpoint quickly.

The Ziegler–Nichols method and most other tuning methods before the 1980s were developed to minimize the peak and integrated error quantified in Chapter 3 for an unmeasured step disturbance on the process input (load upset). While this aggressive action is important in certain applications to prevent the activation of relief systems or initiation of a runaway condition, the robustness is generally insufficient and will not deal with practical problems and other objectives. The more recent tuning methods such as Lambda tuning focus on adding robustness, minimizing the effect of nonlinearities, interactions, and resonance and meeting other process objectives such as maximizing the absorption of variability per Section 1.4.12 for surge tank level, the coordination of loops for ratio control, and the consistency of a setpoint response of lower loops for cascade control and model predictive control.

Most of the control literature focuses on improving the response to a setpoint or a disturbance on the process output for moderate self-regulating processes. Gamesmanship tends to rule and algorithms and tuning methods are touted based on improvements without realization of the diversity of conditions and demands in actual industrial installations.

Lambda tuning rules are presented because of their flexibility in dealing with different processes, nonlinearities, objectives, and extenuating circumstances. Modifications are presented to the normal Lambda tuning rules to enable Lambda tuning to give a similar load disturbance rejection capability as the SCM when this is the criterion without interfering with the flexibility of Lambda tuning.

The SCM developed by the author is used to provide a benchmark of aggressive tuning settings that will minimize the peak and integrated error from load disturbances (process input disturbances) without causing an oscillatory response if the dynamics are perfectly known and constant. The SCM first estimates the ultimate period that has many diagnostic uses beyond the computation of tuning settings.

Other tuning methods can be used as long as the advocate uses good software to compute the settings and window of allowable controller gains for integrating and runaway processes is not violated. More important than the tuning rules is the actual tuning of the loop with good software including any change in settings by an experienced practitioner to address unknowns, nonlinearities, varying objectives, and extenuating circumstances. Consultants often tout their tuning rules when in actuality the process of intelligently tuning a loop was more important than the rules used. PID tuning totally set manually based on intuition is in most cases messed up.

Consultants have such pride and time invested in their tuning rules, the improvement in a loop is often more attributed to the rules than the software that gets settings based on identified dynamics and the intelligence used to modify the settings to deal with the conditions and meet the demands of industrial applications.

The use of Lambda rather than Lambda factors offers many important advantages. Consultants in the use of Lambda tuning for industrial processes do not enter a Lambda factor but an actual Lambda that is a closed loop time constant for self-regulating processes and an arrest time for integrating processes. Lambda typically ranges from one dead time to three dead times to meet application requirements. Limits are imposed on the reset and rate times to help maximize load rejection. Finally, processes with a time constant to dead time ratio greater than four are treated as near-integrating processes, and Lambda tuning rules for integrating processes are used.

All of the equations assume the maximum open loop self-regulating process gain or maximum open loop integrating process gain, minimum primary time constant, and maximum dead time and secondary time constant were evaluated at different operating conditions and worst case condition is used that results in the smallest gain and rate time and largest reset time. At low production rates, the dead time and open loop gain are often the largest for composition and temperature control of well-mixed liquid volumes. If these dynamic terms vary, adaptive tuning or scheduling of tuning settings is needed.

There is not enough space or time to cover all of the tuning rules effectively used in industry. Here we focus on seven major sets of tuning rules. The section concludes with a description of how to compute the arrest time to maximize the absorption of variability for level control in surge tanks and other volumes when the manipulated flow is the feed to a downstream unit operation. Since the nomenclature is extensive and necessary to a full understanding, we start with the nomenclature definition.

Nomenclature for Tuning Calculations:

A = cross sectional area of surge tank (m^2)
AR_{-180} = amplitude ratio at − 180 degrees phase shift (dimensionless)
a = gain factor in traditional open loop methods (0.4 to 1.0)
b = reset time factor in traditional open loop methods (0.5 to 4.0)
c = rate time factor in traditional open loop methods (0 to 1.0)
K_c = controller gain (dimensionless)
K_i = open loop integrating process gain (1/sec)
K_m = measurement gain (%/m for level)
K_p = process gain (m/kg for level)
K_v = valve or VSD gain (kg/sec/%)
K_o = open loop self-regulating process (steady-state) gain (%/%) (dimensionless)
K'_p = open loop runaway process gain (%/%) (dimensionless)
K_q = controller gain that causes quarter amplitude oscillations (dimensionless)
K_u = controller ultimate gain (dimensionless)
L = Ziegler–Nichols lag graphically estimated as intersection with original PV of a tangent to the inflection point of the PV open loop response (sec)
R = Ziegler–Nichols ramp rate sensitivity graphically estimated as slope of tangent to inflection point of the PV open loop response (%/sec per % PID output change)

N_m = measurement noise amplitude (%)
S_m = measurement threshold sensitivity limit (%)
S_v = valve stick-slip or resolution limit (%)
t_{fsr} = full-scale residence time (sec)
T_i = controller integral time (reset time) (sec)
T_d = controller derivative time (rate time) (sec)
T_q = quarter amplitude period (sec)
T_u = ultimate period (sec)
$\%CO$ = operating point controller output (%)
$\%CO_{Limit}$ = controller output limit (%)
$\Delta\%CO_{max}$ = maximum available change in controller output (%)
$\%PV_{Limit}$ = PV limit (%)
$\Delta\%PV_{max}$ = maximum allowable change in PV (%)
$\%SP$ = operating point setpoint (%)
ΔF_{max} = maximum change in valve or VSD flow (e.g., flow span) (m/sec)
ΔL_{max} = maximum change in level (e.g., level span) (m)
γ_f = gamma factor for IMC closed loop time constant or arrest time (dimensionless)
γ= gamma for IMC closed loop time constant or arrest time (sec)
λ_f = Lambda factor for closed loop time constant or arrest time (dimensionless)
λ= Lambda for closed loop time constant or arrest time (sec)
θ_o = total loop dead time (sec)
τ_o = primary time constant (open loop time constant—largest time constant in loop) (sec)
τ_s = secondary time constant (second largest time constant in loop) (sec)
τ_p = positive feedback process time constant (largest time constant in loop) (sec)
ω_n = natural frequency (critical frequency) (radians/sec)
ρ = liquid density in surge tank (kg/m^3)

1.5.1 LAMBDA TUNING FOR SELF-REGULATING PROCESSES

The Lambda tuning method for self-regulating processes is advocated by the author for use when the open loop time constant (primary time constant) is less than four times the total loop dead time. For self-regulating processes that are not near-integrating ($\tau_o < 4 * \theta_o$) we have the following series of equations.

The reset time is set equal to the open loop time constant. The reset time steadily decreases as the ratio of the time constant to dead time decreases from its maximum of four dead times if this tuning method is used when the open loop time constant is less than four times the total loop dead time. For systems with an open loop time constant much less than the dead time (dead time dominant), the PID action becomes essentially integral only due to a steady decrease in the controller gain and the reset time. This type of control helps deal with the abrupt almost step changes and noise in a dead time dominant system. Thus, the PID controller takes over the role of providing a gradual and smoothing response that is missing in the process. The best scenario is to have a large process time constant to fulfill this role, but for plug flow and sheet line processes this is not possible.

$$T_i = \tau_o \tag{1.6a}$$

A low limit is added to traditional equations to prevent a reset time smaller than 0.4 times the dead time for loops where the time constant is very small to provide some gain action.

$$T_i = Max\left[0.4 * \theta_o, \tau_o\right] \tag{1.6b}$$

$$K_c = \frac{T_i}{K_o * (\lambda_f * \tau_o + \theta_o)} \tag{1.6c}$$

In industrial applications, the Lambda factor multiplication of the open loop time constant is replaced with Lambda, the closed loop time constant ($\lambda = \lambda_f * \tau_o$), which is the time after the dead time to reach 63 percent of a setpoint change. The use of Lambda rather than Lambda factor provides the recognition that Lambda is set as a multiple of the dead time to minimize variability in the PID input or output and to provide the gain margin needed to deal with extenuating circumstances (e.g., interaction, inverse response, and resonance).

$$K_c = \frac{T_i}{K_o * (\lambda + \theta_o)} \tag{1.6d}$$

For maximum unmeasured disturbance rejection, a Lambda equal to the dead time is used:

$$K_c = 0.5 * \frac{\tau_o}{K_o * \theta_o} \tag{1.6e}$$

We will see in Chapters 8 and 9, a Lambda of three dead times minimizes the consequences of nonlinearities, inverse response, and resonance.

Normally a rate setting is not included as part of the Lambda tuning for self-regulating processes. If the primary time constant is greater than half the dead time, rate action may be beneficial. If a secondary time constant (next largest time constant) can be identified, the rate time is set equal to the secondary time constant. The rate time should be larger than half the dead time but not greater than one-fourth the reset time for an ISA Standard Form.

If the primary time constant is greater than half the dead time ($\tau_o > 0.5 * \theta_o$):

$$T_d = Min\left[0.25 * T_i, Max\,(0.5 * \theta_o, \tau_s)\right] \tag{1.6f}$$

1.5.2 LAMBDA TUNING FOR INTEGRATING PROCESSES

The Lambda tuning method for integrating processes is advocated by the author for use on self-regulating processes when the open loop time constant is greater than four times the total loop dead time (near-integrating processes) besides for use on true integrating and runaway processes to provide a faster return to setpoint for load upsets.

For PID control, the reset time is twice the arrest time plus the dead time.

$$T_i = 2 * \lambda + \theta_o \tag{1.7a}$$

For PI control, a low limit is added so that the reset time is not less than four dead times.

$$T_i = Max\left[4 * \theta_o, 2 * \lambda + \theta_o\right] \tag{1.7b}$$

$$K_c = \frac{T_i}{K_i * (\frac{\lambda_f}{K_i} + \theta_o)^2} \tag{1.7c}$$

In industrial applications, the Lambda factor multiplication of the inverse of the integrating process gain is replaced with Lambda, the closed loop arrest time ($\lambda = \lambda_f / K_i$), which is the time to stop an excursion for an unmeasured disturbance (time to peak error). The use of Lambda rather than Lambda factor provides the recognition that Lambda is set as a multiple of the dead time to minimize variability in either the PID input or output and to provide the gain margin needed to deal with extenuating circumstances (e.g., interaction, inverse response, and resonance) or is set as per Section 1.5.14 to provide the maximum absorption of variability.

The numerator uses the unlimited reset time.

$$K_c = \frac{2 * \lambda + \theta_o}{K_i * (\lambda + \theta_o)^2} \tag{1.7d}$$

For maximum unmeasured disturbance rejection, a Lambda equal to the dead time is used:

$$K_c = 0.75 * \frac{1}{K_i * \theta_o} \tag{1.7e}$$

We will see in Chapters 8 and 9, a Lambda of three dead times minimizes the consequences of nonlinearities, inverse response, and resonance.

The rate time should be greater than half the dead time but less than one-fourth the reset time.

$$T_d = Min\left[0.25 * T_i,\ Max\ (0.5 * \theta_o, \tau_s)\right] \tag{1.7f}$$

1.5.3 IMC TUNING FOR SELF-REGULATING PROCESSES

The IMC method was developed to provide a controller that is the inverse of the process dynamics. Similar to Lambda tuning, IMC tuning uses pole-zero cancellation theory. The IMC tuning rules vary with author, dead time approximation, and vintage. The IMC rules presented here are documented in the book *Advanced Control Unleashed* (Blevins et al. 2003). Many authors have offered improvements for different dynamics. In Sections 1.5.5 and 1.5.6, we include "Simple Analytic Rules for Model Reduction and PID Controller Tuning" (Skogestad 2003).

For self-regulating processes:

$$K_c = \frac{T_i}{K_o * (\gamma_f * \tau_o + 0.5 * \theta_o)} \tag{1.8a}$$

In industrial applications, the gamma factor multiplication of the open loop time constant is replaced with gamma, the closed time constant ($\gamma = \gamma_f * \tau_o$) for a setpoint change.

$$K_c = \frac{T_i}{K_o * (\gamma + 0.5 * \theta_o)} \tag{1.8b}$$

The reset time is set equal to the open loop time constant plus half the total loop dead time. By including the dead time, IMC tuning prevents the PID from using incredibly small reset times and small gains when the time constant is much less than the dead time ($\tau_o << \theta_o$). Thus, the IMC tuning behaves more like the modified Ziegler–Nichols tuning for dead time dominance. However, the reset time is even larger than optimum for systems where the time constant is much greater than the dead time ($\tau_o >> \theta_o$) resulting in an even slower recovery and consequently much larger integrated error from unmeasured disturbances. The user can take the same approach recommended for Lambda tuning, which is to use a near-integrator tuning when the time constant becomes four time larger than the dead time ($\tau_o > 4 * \theta_o$).

$$T_i = \tau_o + 0.5 * \theta_o \tag{1.8c}$$

For maximum unmeasured disturbance rejection, a gamma equal to the dead time is used:

$$K_c = 0.7 * \frac{\tau_o}{K_o * \theta_o} \tag{1.8d}$$

$$T_i = Max\ (2 * \Delta t_x, \tau_o + 0.5 * \theta_o) \tag{1.8e}$$

The IMC rate term is quite different in being proportional to the product of the open loop time constant and total loop dead time. The rate time should not be greater than one-fourth the reset time for an ISA Standard Form.

$$T_d = Min\ (0.25 * T_i, \frac{\tau_o * \theta_o}{2 * \tau_o + \theta_o}) \tag{1.8f}$$

1.5.4 IMC TUNING FOR INTEGRATING PROCESSES

The IMC tuning method is similar to the Lambda Tuning method for integrating processes except for the use of half of the dead time instead of the whole dead time in the denominator to calculate the controller gain.

$$K_c = \frac{T_i}{K_i * (\frac{\gamma_f}{K_i} + 0.5 * \theta_o)^2} \tag{1.9a}$$

In industrial applications, the gamma factor multiplication of the inverse of the integrating process gain is replaced with gamma, the closed loop arrest time ($\gamma = \gamma_f / K_i$) for an unmeasured disturbance.

$$K_c = \frac{T_i}{K_i * (\gamma + 0.5 * \theta_o)^2} \tag{1.9b}$$

The reset time is twice the arrest time plus the dead time.

$$T_i = 2 * \gamma + \theta_o \tag{1.9c}$$

For maximum unmeasured disturbance rejection, a gamma equal to the dead time is used:

$$K_c = 1.3 * \frac{1}{K_i * \theta_o} \tag{1.9d}$$

$$T_i = 3 * \theta_o \tag{1.9e}$$

If a secondary time constant (next largest time constant) can be identified, derivative action can be used with the rate time set equal to the secondary time constant. The rate time should not be greater than one-fourth the reset time for an ISA Standard Form.

$$T_d = Min\left[0.25 * T_i, \tau_s\right] \tag{1.9f}$$

1.5.5 SKOGESTAD INTERNAL MODEL CONTROL TUNING FOR SELF-REGULATING PROCESSES

Sigurd Skogestad recognized the problem with incredibly slow and fast integral action in IMC and Lambda tuning for processes with an extremely large and small primary time constant, respectively, because the reset time was set proportional to this time constant. The Skogestad Internal Model Control (SIMC) rules prevent the reset time from becoming exceptionally large or small. The rules end up with about the same controller gain as Lambda tuning for self-regulating processes. Here, we will use gamma (γ) employed by IMC tuning instead of the closed loop time constant (τ_c) used in the paper "Simple Analytic Rules for Model Reduction and PID Controller Tuning" (Skogestad 2003). The rules presented here are for PID control of a process with dead time and a large open loop time constant (τ_o) and a small secondary time constant (τ_s).

$$K_c = \frac{T_i}{K_o * (\gamma + \theta_o)} \tag{1.10a}$$

$$T_i = Min\left[\tau_o, 4 * (\gamma + \theta_o)\right] \tag{1.10b}$$

$$T_d = Min\left[0.25 * T_i, \tau_s\right] \tag{1.10c}$$

1.5.6 SIMC TUNING FOR INTEGRATING PROCESSES

For integrating processes, the SIMC reset time is about 270 percent larger than the reset time per Lambda and IMC tuning methods for maximum disturbance rejection where gamma is set equal to the dead time. The SIMC controller gain is about 30 percent smaller for this case. The product of the SIMC-gain and reset time is about twice as large as for Lambda tuning giving more margin to prevent slow rolling oscillations (Equation 1.5c).

$$K_c = \frac{1}{K_i * (\gamma + \theta_o)} \tag{1.11a}$$

$$T_i = 4 * (\gamma + \theta_o) \tag{1.11b}$$

$$T_d = Min\left[0.25 * T_i, \tau_s\right] \tag{1.11c}$$

1.5.7 TRADITIONAL OPEN LOOP TUNING

The following equations are for an ISA Standard Form PID controller. The limit on the rate time ensures the rate time is not larger than one-fourth the reset time. The a, b, and c factors are decreased as the ratio of time constant to dead time decreases. The minimum numbers in the nomenclature definition are for a time constant much less than the dead time (dead time dominant). Without the limits on the controller gain and reset time, the controller becomes exceptionally slow for dead time dominant loops.

$$K_c = Max\left[0.2*\frac{1}{K_o}, a*\left(\frac{\tau_o}{K_o*\theta_o}\right)\right] \tag{1.12a}$$

$$T_i = Max\left[0.4*\theta_o, b*\theta_o\right] \tag{1.12b}$$

$$T_d = Min\left[0.25*T_i, c*\theta_o\right] \tag{1.12c}$$

1.5.8 MODIFIED ZIEGLER–NICHOLS REACTION CURVE TUNING

The Ziegler–Nichols reaction curve method, unlike the ultimate oscillation method, is conducted with the controller in MAN. A step change is made in the manual controller output ($\Delta\%CO$) and a maximum ramp rate per percent change in controller output is estimated. The original paper shows this parameter R as being graphically estimated as the slope of a tangent line to the inflection point of a self-regulating process open loop response that goes to completion. The L parameter is used to denote a lag that is estimated as the time from the controller output change to the intersection of the tangent with the original PV. The official definition of a lag is any phase lag that can be due to a dead time or a time constant. In most publications today, lag time is used interchangeably with time constant whereas Ziegler–Nichols was using lag time as a dead time. The method is modified to provide a smoother than quarter amplitude response and to add some robustness by applying a 0.5 factor to the Ziegler–Nichols gain as shown in Equation 1.13a.

The parameter R is really the integrating process gain that can be measured online by passing a new $\%PV$ through a dead time block whose parameter is the total loop dead time to create an old $\%PV$ that is subtracted from the new $\%PV$ to create a delta $\%PV$. The maximum delta $\%PV$ divided by the dead time (θ_o) and the change in controller output is the maximum ramp rate per percent change in controller output.

$$K_c = 0.5*\frac{1}{R*L} \tag{1.13a}$$

$$T_i = 3*L \tag{1.13b}$$

The max ramp rate per percent change in controller output is the integrating process gain.

$$R = \frac{Max\left(\Delta\%PV/\Delta t\right)}{\Delta\%CO} = K_i \tag{1.13c}$$

The L parameter is the observed loop dead time.

$$L = \theta_o \tag{1.13d}$$

If we substitute the equations for the definition of *R* and *L* parameters, we end up with the Lambda tuning Equations 1.7d and 1.7e for integrating processes when Lambda is set equal to the dead time except the gain factor is different.

$$K_c = 0.5 * \frac{1}{K_i * \theta_o} \tag{1.13e}$$

$$T_i = 3 * \theta_o \tag{1.13f}$$

The integrating process gain for a near-integrating process is the open loop gain divided by the open loop time constant.

$$K_i = \frac{K_o}{\tau_o} \tag{1.13g}$$

If we substitute Equation 1.13g into Equation 1.13e we end up with Equation 1.13h that is the Lambda tuning Equation 1.6d for a self-regulating process for a 3:1 time constant to dead time ratio when Lambda is set equal to the dead time. The reset time is the same as Equation 1.13f since the time constant is equal to three times the dead time.

$$K_c = 0.5 * \frac{\tau_o}{K_o * \theta_o} \tag{1.13h}$$

1.5.9 MODIFIED ZIEGLER–NICHOLS ULTIMATE OSCILLATION TUNING

The Ziegler–Nichols ultimate oscillation tuning method is conducted with the controller in AUTO. If the loop is lined out, the controller is momentarily put in MAN and a step change is made in the controller output and the controller is immediately returned to AUTO. With the reset time at a maximum (more than 100 times greater than the dead time) and the rate time set to zero to give essentially a proportional-only controller, the controller gain is increased until there are equal amplitude oscillations. The controller gain that caused these oscillations is the ultimate gain and the period of the oscillations is the ultimate period. Generally, this technique is too exciting in that the loop is on the borderline of instability where the oscillations could rapidly grow in magnitude. Consequently, the manual quarter amplitude oscillation method or the relay auto tuner is preferred to get the ultimate gain and period. Here we look at how the Ziegler–Nichols ultimate oscillation method gives about the same results as the much more practical Ziegler–Nichols reaction curve method.

Starting with fundamental relationship that ultimate gain is the inverse of the product of the open loop gain and amplitude ratio at −180 degrees phase shift.

$$K_u = \frac{1}{K_0 * AR_{-180}} \tag{1.14}$$

For self-regulating single time constant processes, the amplitude ratio at −180 degrees is:

$$AR_{-180} = \frac{1}{\sqrt{1 + (\tau_o * \omega_n)^2}} \tag{1.15a}$$

$$K_u = \frac{\sqrt{1+(\tau_o * \omega_n)^2}}{K_o} \tag{1.15b}$$

Using natural frequency relationship to ultimate period $\omega_n = \dfrac{2*\pi}{T_u}$

$$K_u = \frac{\sqrt{1+(\tau_o * 2 * \pi / T_u)^2}}{K_o} \tag{1.15c}$$

For loop dominated by a large time constant ($\tau_o >> \theta_o \Leftrightarrow T_u << \tau_o$), the ultimate gain equation simplifies to:

$$K_u = \frac{2 * \pi * \tau_o}{K_o * T_u} \tag{1.15d}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$K_u = \frac{2 * \pi * \tau_o}{K_o * 4 * \theta_o} \tag{1.15e}$$

If we use the Ziegler–Nichols ultimate oscillation equations for a PID controller, we end up with the Ziegler–Nichols reaction curve method tuning except the reset factor which is two instead of three. The rate time for the ISA Standard Form is one-fourth the reset time. We can convert back and forth between a lag-dominant and near-integrator calculation for the controller gain by the use of Equation 1.13g to convert the dead time to time constant ratio to an integrating process gain. The method is modified to provide a smoother than quarter amplitude response and to add some robustness by applying a 0.5 factor to the Ziegler–Nichols gain as shown in Equation 1.13a.

$$K_c = 0.5 * 0.6 * K_u = 0.5 * 0.6 * \frac{2 * \pi * \tau_o}{K_o * 4 * \theta_o} = 0.5 * \frac{\tau_o}{K_o * \theta_o} \tag{1.16a}$$

$$T_i = 0.5 * T_u = 0.5 * 4 * \theta_o = 2 * \theta_o \tag{1.16b}$$

$$T_d = 0.125 * T_u = 0.125 * 4 * \theta_o = 0.5 * \theta_o \tag{1.16c}$$

Even if the Ziegler–Nichols ultimate oscillation method is not used, knowing the ultimate gain from using the relay tuner or using the equations in Chapter 4 offers knowledge of how close the loop is to instability. The gain margin is approximately the ratio of the ultimate gain to the current PID gain. This gain margin is extensively used to deal with not only changes in open loop gain but also dead time and time constants (see Chapter 7).

1.5.10 QUARTER AMPLITUDE OSCILLATION TUNING

The Ziegler–Nichols ultimate oscillation method has been heavily criticized for being too disruptive and potentially unsafe by requiring the user to create equal amplitude oscillations that put the loop on the verge of instability. The quarter amplitude method prevents excessive oscillations during the closed loop test to identify the loop dynamics and tuning settings by only pushing the loop to rapidly decaying quarter amplitude oscillations. This method also prevents mistaking limit cycles, which have equal amplitude oscillations as ultimate oscillations. The quarter amplitude period and controller gain are then used to approximate the ultimate period and gain, given some of the non-ideal effects. While the method is not as accurate as ultimate oscillation method, the error is usually well within the uncertainty of tuning settings due to nonlinearities. The controller gains are cut in half via a 0.5 factor to give a smoother response.

For an ISA Standard Form proportional-only (P only) controller:

$$K_c = 0.5 * 0.6 * K_u = 0.3 * K_u \tag{1.17a}$$

For an ISA Standard Form PI controller:

$$K_c = 0.5 * 0.4 * K_u = 0.2 * K_u \tag{1.17b}$$

$$T_i = 0.8 * \frac{T_u}{Min(4, 10 * (\frac{4 * \theta_o}{T_u} - 1)^2 + 1)} \tag{1.17c}$$

For an ISA Standard Form PID controller:

$$K_c = 0.5 * 0.6 * K_u = 0.3 * K_u \tag{1.17d}$$

$$T_i = 0.6 * \frac{T_u}{Min(4, 10 * (\frac{4 * \theta_o}{T_u} - 1)^2 + 1)} \tag{1.17e}$$

$$T_d = 0.8 * Max\left[0.0, 0.25 * (T_i - 0.5 * \theta_o)\right] \tag{1.17f}$$

Note that the derivative time (rate time) for an ISA Standard form must be less than the integral time (reset time) or instability will result from a reversal in controller gain. The Series form inherently prevented the effective rate time from becoming larger than the reset time (see Appendix K for more details).

The ultimate gain and period can be approximated as follows from the gain and period of quarter amplitude oscillations for industrial loops with a dead band or a resolution limit:

$$K_u = 1.5 * K_q \tag{1.17g}$$

$$T_u = 0.7 * T_q \tag{1.17h}$$

If the dead band or resolution limit in the valve or VSD and measurement is negligible, Equation 1.17h factor approaches one (quarter amplitude oscillation period approaches ultimate period).

1.5.11 SCM TUNING FOR SELF-REGULATING PROCESSES

The ultimate period can be estimated from the primary time constant and total loop dead time based on Bode plot results. The PID gain is limited to being greater than 20 percent of the inverse of the open loop gain to prevent too small a gain for dead time dominant loops. The settings computed here provide the most aggressive response to a load upset. In practice, the gain is reduced to provide a smoother and more robust response.

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o} \right)^{0.65} \right] * \theta_o \tag{1.18a}$$

For an ISA Standard Form PI controller:

$$K_c = Max\left(0.2 / K_o, 0.6 * \frac{\tau_o}{K_o * \theta_o} \right) \tag{1.18b}$$

$$T_i = 0.8 * \frac{T_u}{Min(4, 10 * (\frac{4 * \theta_o}{T_u} - 1)^2 + 1)} \tag{1.18c}$$

For an ISA Standard Form PID controller:

$$K_c = Max\left(0.2 / K_o, 0.8 * \frac{\tau_o}{K_o * \theta_o} \right) \tag{1.18d}$$

$$T_i = 0.6 * \frac{T_u}{Min(4, 10 * (\frac{4 * \theta_o}{T_u} - 1)^2 + 1)} \tag{1.18e}$$

$$T_d = Min\left[0.25 * T_i, 0.8 * Max\left[0.0, 0.25 * (T_i - 0.5 * \theta_o) + \tau_s \right] \right] \tag{1.18f}$$

1.5.12 SCM TUNING FOR INTEGRATING PROCESSES

The ultimate period can be estimated from the secondary time constant and total loop dead time based on Nyquist plot results. The PID gain is limited to being less than the valve stick-slip (S_v) divided by the difference between the measurement noise (N_m) and the measurement sensitivity limit (S_m) to prevent fluctuations in the PID output from exceeding the stick-slip causing excessive packing wear and high frequency disturbances. The divisor is limited to a 16 bit A/D convertor resolution of 0.003 percent. In older DCS with a 12 bit A/D, a resolution of 0.05 percent should be used in divisor.

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o} \right)^{0.65} \right] * \theta_o \tag{1.19a}$$

For an ISA Standard Form PI controller:

$$K_c = Min\left(\frac{S_v}{\max\left[(N_m - S_m), 0.003 \right]}, 0.6 * \frac{1}{K_i * \theta_o} \right) \tag{1.19b}$$

$$T_i = 0.8 * T_u \tag{1.19c}$$

For an ISA Standard Form PID controller:

$$K_c = Min\left(\frac{S_v}{\max\left[(N_m - S_m), 0.003\right]}, 0.8 * \frac{1}{K_i * \theta_o}\right) \tag{1.19d}$$

$$T_i = 0.6 * T_u \tag{1.19e}$$

$$T_d = Min\left[0.25 * T_i, 0.8 * Max\left[0.0, 0.25 * (T_i - 0.5 * \theta_o) + \tau_s\right]\right] \tag{1.19f}$$

1.5.13 SCM TUNING FOR RUNAWAY PROCESSES

The ultimate period can be estimated from the positive feedback time constant, secondary time constant and total loop dead time based on Nyquist plot results. The PID gain is limited to being greater than twice the inverse of the open loop runaway process gain ($K_p^{'}$) to prevent the process from going unstable due to insufficient feedback action. PI tuning is not offered since the omission of derivative action is not advisable.

$$T_u = 4 * \left[1 + \left(\frac{N}{D}\right)^{0.65}\right] * \theta_o \tag{1.20a}$$

$$N = \left(\tau_p^{'} + \tau_s\right) * \left(\tau_p^{'} * \tau_s\right) \tag{1.20b}$$

$$D = \left(\tau_p^{'} - \tau_s\right) * \left(\tau_p^{'} - \theta_o\right) * \theta_o \tag{1.20c}$$

For an ISA Standard Form PID controller:

$$K_c = Max\left(2.0 / K_o, 0.8 * \frac{\tau_p^{'}}{K_p^{'} * \theta_o}\right) \tag{1.20d}$$

$$T_i = 0.6 * T_u \tag{1.20e}$$

$$T_d = Min\left[0.25 * T_i, 0.8 * Max\left[0.0, 0.25 * (T_i - 0.5 * \theta_o) + \tau_s\right]\right] \tag{1.20f}$$

1.5.14 MAXIMIZING ABSORPTION OF VARIABILITY TUNING FOR SURGE TANK LEVEL

When the absorption of variability must be maximized, Lambda is chosen to be as large as possible. The most common occurrence is a level control application where the transfer is minimized of flow changes coming into the volume to flow changes going out of the volume. The flow changes coming in are absorbed as much as possible by allowing the level to change within operating limits, such as low and high alarm points. The Lambda integrating tuning method is used and the arrest time is as large as possible without causing violation of a level limit. The maximum arrest time Lambda depends upon the integrating process gain, the allowable change in the PV and the available change in the manipulated flow. The following equations show how to calculate Lambda for a generic application and then a surge tank level loop.

The maximum arrest time Lambda (λ) is the maximum allowable percent excursion ($\Delta\%PV_{max}$) divided by the maximum possible PV ramp rate. The maximum possible ramp rate is the PV rate of change per percent output change multiplied by the maximum available percent output change ($\Delta\%CO_{max}$).

$$\lambda = \frac{\left|\Delta\%PV_{\max}\right|}{\left|\dfrac{\frac{\Delta\%PV}{\Delta t}}{\Delta\%CO}\right| * \left|\Delta\%CO_{\max}\right|} \tag{1.21a}$$

Realizing that the integrating process gain is the PV ramp rate per percent output change:

$$\lambda = \frac{1}{K_i} * \frac{\left|\Delta\%PV_{\max}\right|}{\left|\Delta\%CO_{\max}\right|} \tag{1.21b}$$

An equivalent setpoint rate limit on the controller output (e.g., flow controller setpoint):

$$\left|\frac{\Delta\%CO}{\Delta t_{\max}}\right| = \frac{\left|\Delta\%CO_{\max}\right|}{\lambda} = K_i * \frac{\left|\Delta\%CO_{\max}\right|^2}{\left|\Delta\%PV_{\max}\right|} \tag{1.21c}$$

For a PV limit ($\%PV_{Limit}$) and corresponding CO limit ($\%CO_{Limit}$), we have:

$$\lambda = \frac{1}{K_i} * \left|\frac{\%PV_{Limit} - \%SP}{\%CO_{Limit} - \%CO}\right| \tag{1.21d}$$

The above calculation would be done for high and low operating limits and various setpoints. The smallest of the arrest times would be used in tuning.

We can obtain the more detailed requirements for surge level tank level control by computing the integrating process gain for level. The integrating process gain is the product of the valve, process, and measurement gains:

$$K_i = K_v * K_p * K_m \tag{1.22a}$$

The valve gain or VSD gain for a linear installed characteristic or flow loop is:

$$K_v = \frac{\Delta F_v}{\Delta\%CO} = \frac{\Delta F_{\max}}{100\%} \tag{1.22b}$$

The level process gain for mass flow is (omit density term for volumetric flow):

$$K_p = \frac{1}{A * \rho} \tag{1.22c}$$

The level measurement gain is:

$$K_m = \frac{\Delta\%PV}{\Delta L} = \frac{100\%}{\Delta L_{\max}} \tag{1.22d}$$

Substituting in the valve, process, and measurement gains, the integrating process gain is:

$$K_i = K_v * K_p * K_m = \frac{\Delta F_{\max}}{100\%} * \frac{1}{A * \rho} * \frac{100\%}{\Delta L_{\max}} = \frac{\Delta F_{\max}}{\Delta L_{\max}} * \frac{1}{A * \rho} \tag{1.22e}$$

The consequential arrest time for a level loop is:

$$\lambda = \frac{A * \rho * \Delta L_{\max}}{\Delta F_{\max}} * \frac{\Delta\%PV_{\max}}{\Delta\%CO_{\max}} \tag{1.22f}$$

An equivalent setpoint rate limit on the controller output (e.g., flow controller setpoint):

$$\left| \Delta\%CO / \Delta t \right|_{\max} = \frac{\left|\Delta\%CO_{\max}\right|}{\lambda} = \frac{\Delta F_{\max}}{A * \rho * \Delta L_{\max}} * \frac{\left|\Delta\%CO_{\max}\right|^2}{\left|\Delta\%PV_{\max}\right|} \tag{1.22g}$$

The computation of the arrest time can be significantly simplified if the dead time is assumed to be negligible. This is a reasonable assumption for surge tank level control because the total loop dead time is much smaller than the arrest time in Equations 1.7b and 1.7c for Lambda tuning of integrating processes.

$$K_c = \frac{T_i}{K_i * (\lambda)^2} \tag{1.22h}$$

$$T_i = 2 * \lambda \tag{1.22i}$$

If we substitute Equation 1.22i into 1.22h and cancel out Lambda in the numerator by one of the Lambdas in the denominator, we have a simpler Equation 1.22j for the PID gain:

$$K_c = \frac{2}{K_i * \lambda} \tag{1.22j}$$

$$K_c = \frac{2}{\frac{\Delta F_{\max}}{\Delta L_{\max}} * \frac{1}{A * \rho} * \frac{A * \rho * \Delta L_{\max}}{\Delta F_{\max}} * \frac{\Delta\%PV_{\max}}{\Delta\%CO_{\max}}} = 2 * \frac{\Delta\%CO_{\max}}{\Delta\%PV_{\max}} \tag{1.22k}$$

The first expression in Equation 1.22f can be simplified to Equation 1.22l if we define a full-scale residence time (t_{fsr}) as the maximum volume ($A * \rho * \Delta L_{\max}$) for the maximum span of the level measurement divided by the maximum flow capacity of the valve or VSD ($\Delta F_{\max}$). This full-scale residence time term is not to be confused with the actual residence time (operating volume divided by operating flow rate) used in process calculations in chemical engineering and throughout the rest of this book that is particularly important for computing back mixed time constants or plug flow transportation delays and the time for reaction conversion.

$$t_{fsr} = \frac{A * \rho * \Delta L_{\max}}{\Delta F_{\max}} \tag{1.22l}$$

We get Equation 1.22m using the full-scale residence time (t_{fsr}) by substitution as per Equation 1.22l into the first expression in Equation 1.22f for Lambda:

$$\lambda = t_{fsr} * \frac{\Delta\%PV_{\max}}{\Delta\%CO_{\max}} \tag{1.22m}$$

We have the simple Equation 1.22n from the result of Equation 1.22k for the PID gain:

$$K_c = 2 * \frac{\Delta\%CO_{\max}}{\Delta\%PV_{\max}} \tag{1.22n}$$

We can solve Equation 1.22n for the maximum desired change in the PV (level) divided by the maximum desired change in manipulated flow:

$$\frac{\Delta\%PV_{\max}}{\Delta\%CO_{\max}} = \frac{2}{K_c} \tag{1.22o}$$

If we substitute Equation 1.22o into 1.22n we end up with the simple Equation 1.22p for the reset time:

$$T_i = 2 * t_{fsr} * \frac{\Delta\%PV_{\max}}{\Delta\%CO_{\max}} = 2 * t_{fsr} * \frac{2}{K_c} = t_{fsr} * \frac{4}{K_c} \tag{1.22p}$$

1.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (Form and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in setpoint (ΔSP) or load flow at the process input (ΔF_L). Table 1.2 summarizes the test conditions.

In Table 1.2, the "Delay" is the total loop dead time in a FOPDT approximation. "Delay" is the sum of the process dead time, automation system delays and lags, and secondary process time constant. The "Lag" is the primary process time constant. The "Lag" is a positive feedback process time constant when a runaway process is designated. For runaway processes, the "Open Loop Gain" is increased so the process has enough muscle to deal with load upsets. The positive feedback time constant is increased by the same factor as this gain to provide effectively the same open loop integrating process gain (0.01 1/sec) used in the tests for near and true integrating processes. The secondary time constant was one second, the PID execution time 0.5 seconds, the valve delay was 1.2 seconds, the process model execution time was 0.25 seconds, and the measurement update rate was 1 second. As per equations in Chapters 4, 5, and 6 these dynamics contribute about three seconds to the total loop dead time. No analog input filter or PID input filter was used in these tests.

The ultimate period of a self-regulating process varies between two times to four times the dead time for a primary time constant to dead time ratio approaching zero and infinity, respectively. Most integrating and runaway processes have an ultimate period slightly larger than four dead times.

The SCM tuning is used to set the maximum gain, minimum reset time, and maximum rate time settings to achieve maximum disturbance rejection. These settings represent a benchmark but are generally too aggressive and do not address other objectives. In general there is always a tradeoff between performance for unmeasured load upsets and robustness (ability for response to be non-oscillatory for changes in dynamics). The plots generally show that a doubling of gain will cause excessive oscillations. While the gain could have been halved to provide more robustness, we will see in later chapters that by using the near-integrating process identification and Lambda as a multiple of the dead time, Lambda tuning provides the flexibility to deal with nonidealities and diverse process objectives of industrial processes.

When Lambda or IMC tuning is used in Figures 1.10a,b,c,d, the Lambda and gamma, respectively are set equal to the total loop dead time as noted in the PID Tuning column to provide an aggressive response to unmeasured load upsets.

Table 1.2. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
1.5a, b, c	Dead time dominant	1 dimensionless	20	2	$\Delta F_L = 10\%$	SCM
1.6a, b, c, d, e, f	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	SCM
1.7a, b, c, d, e, f	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	SCM
1.8a, b, c, d, e, f	Runaway	4 dimensionless	10	400	$\Delta F_L = 20\%$	SCM
1.9a	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	SCM
1.9b	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	SCM
1.9c	Runaway	4 dimensionless	10	400	$\Delta F_L = 20\%$	SCM
1.10a	Dead time dominant	1 dimensionless	20	2	$\Delta F_L = 10\%$	$\lambda = \theta_o$ $\gamma = \theta_o$
1.10b	Moderate self-regulating	1 dimensionless	20	20	$\Delta F_L = 20\%$	$\lambda = \theta_o$ $\gamma = \theta_o$
1.10c	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	$\lambda = \theta_o$ $\gamma = \theta_o$
1.10d	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	$\lambda = \theta_o$ $\gamma = \theta_o$

The ISA Standard Form is used and unless otherwise noted the Structure was PI on error D on PV. The external reset feedback option using actual valve position readback as the BKCAL signal did not play a role in these tests since the valve was relatively fast.

In all of these tests, the integrated error for a step disturbance at the process input (load upset) as detailed in Section 3.2 is proportional to the reset time and inversely proportional to the PID gain. For a response that does not oscillate across the setpoint, the integrated error has a consistent sign and is thus the same as the integrated absolute error. The integrated absolute error can be seen on a trend chart as the total area between the setpoint and PV.

1.6.1 PERFORMANCE OF TUNING SETTINGS ON DEAD TIME DOMINANT PROCESSES

The peak error for step load upsets to dead time dominant processes is essentially the same as if there was no feedback control. As seen in Figures 1.5a, b, c, the best that feedback control can do is to reduce the offset from setpoint.

Figure 1.5a shows a proportional-only controller (P on error no I or D) does not do much to reduce the offset because the PID gain is small due to dead time dominance. A doubling of the

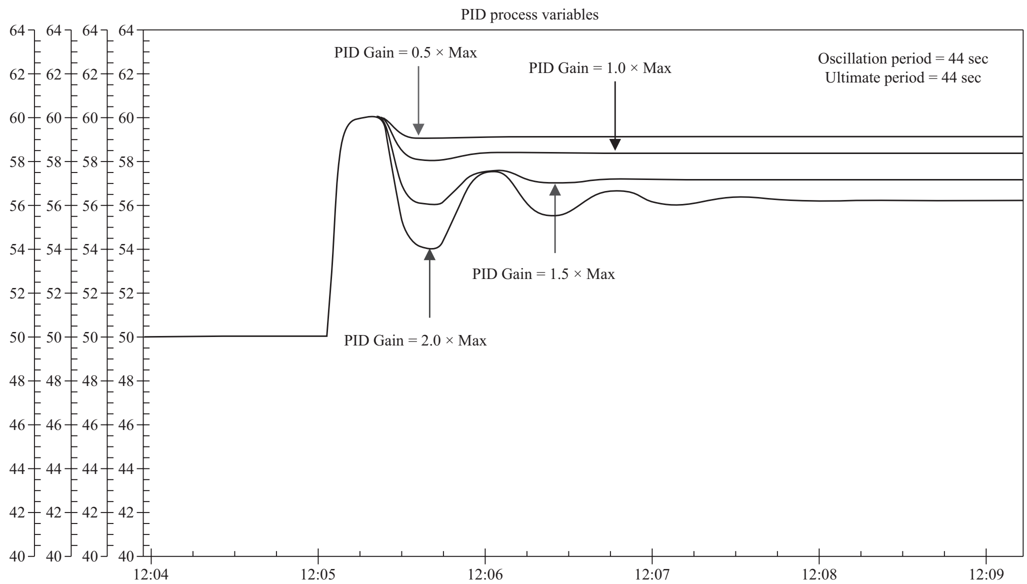


Figure 1.5a. Effect of PID *gain* in *proportional-only* controller on load response for *dead time dominant* process.

controller gain causes an oscillation as expected but the amplitude decays quickly. The period of the oscillation is close to the ultimate period that is slightly greater than two times the dead time for this dead time dominant process. Thus, the ultimate period can be found by increasing the reset time by a factor of 1,000 to have essentially a proportional-only controller and then increasing the PID gain until you get an oscillation.

In Figure 1.5b, the addition of integral action repeats the proportional mode often enough to quickly eliminate the offset. However, a simple halving of the PID gain shows a much more protracted recovery time. A doubling of the PID gain causes much larger amplitude oscillations than those for the proportional-only PID with a period almost 50 percent larger than the ultimate period. The addition of integral action puts the loop closer to instability for a PID tuned to minimize load upsets. 99.9 percent of controllers have integral action.

If the reset time is decreased to be 40 percent of the minimum, an oscillation develops as seen in Figure 1.5c that decays but the period is nearly twice the ultimate period. The distinguishing feature of a PID tuned with too small a reset time is an oscillation period much larger than the ultimate period. You need to know whether the process is dead time dominant to know the ultimate period is about two times the dead time as opposed to the more common case of an ultimate period close to four times the dead time. Fortunately, dead time dominant processes are not main stream and are mostly seen in extruders and sheet lines or fast processes with an at-line analyzer. In other processes where the process time constant is less than the process dead time from a transportation delay, the sensor time constant is often larger than the process dead time (e.g., electrode and thermowell lag greater than transportation delay).

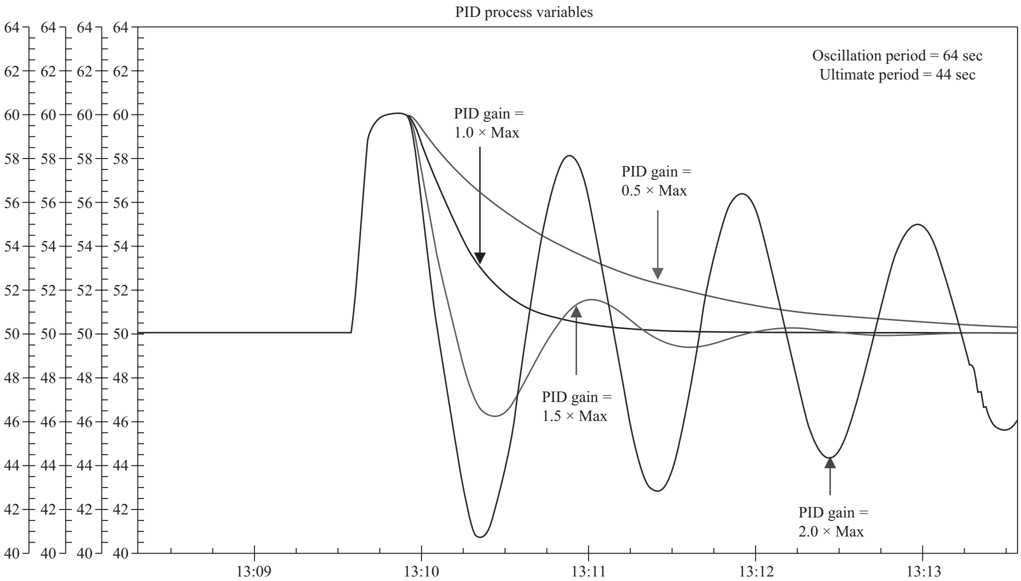


Figure 1.5b. Effect of PID *gain* in *PI* controller on load response for *dead time dominant* process.

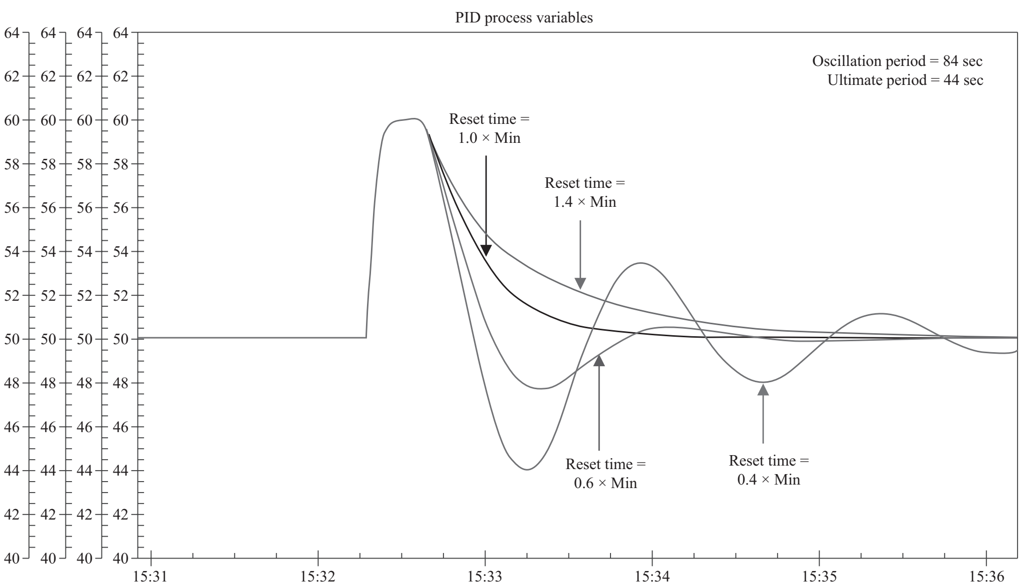


Figure 1.5c. Effect of *reset time* in *PI* controller on load response for *dead time dominant* process.

1.6.2 PERFORMANCE OF TUNING SETTINGS ON NEAR-INTEGRATING PROCESSES

The peak error for step load upsets to near-integrating processes significantly decreases as the PID gain is increased. Since the PID gain can be high for these processes, the PID is capable of reducing the peak error by a factor of 10 or more. The load upset had to be increased to 20 percent in order to show an appreciable error. Since the period of oscillation is double that of the dead time dominant process, the total dead time was halved for tests to integrating and runaway processes to retain a comparable time frame.

Figures 1.5a and 1.6a shows a Proportional-only controller (P on error no I or D) can significantly reduce the offset as the gain is increased to the maximum allowed. Many near-integrating processes have a ratio of primary time constant to dead time that is 10 to 100 times larger than tested leading to an offset that is a small fraction of a percent for the large load upset of 20 percent used in this test. However, achieving minimum peak error and offset, puts the loop close to the edge of stability. Doubling the PID gain leads to oscillations that are of nearly equal amplitude. The period of the oscillation is close to the ultimate period.

The addition of integral action makes the response oscillatory for all the settings shown in Figure 1.6b. A doubling of the PID gain causes oscillations that grow in amplitude, a clear sign of instability. Some tuning rules favor increasing the reset time as factor of dead time unless derivative action can be used.

The addition of derivative action makes the response smoother as seen in Figure 1.6c for the maximum allowable PID gain setting. The response becomes oscillatory when the PID gain setting is doubled but the oscillations are faster and amplitude is much smaller than for the case where derivative was not used. For a 50 percent increase in PID gain, the approach

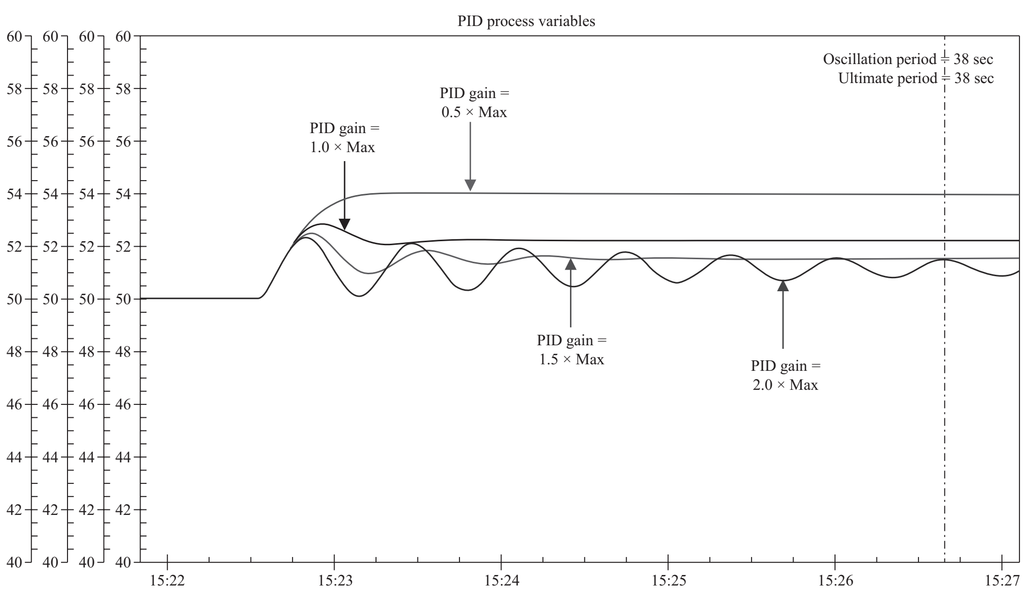


Figure 1.6a. Effect of PID *gain* in *Proportional-only* controller on load response for *near-integrating* process.

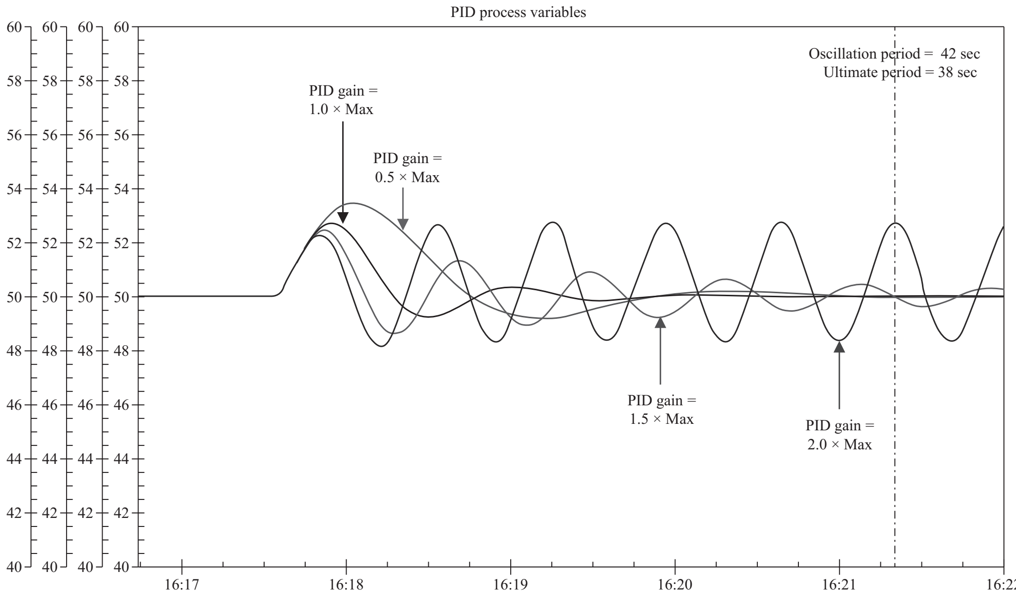


Figure 1.6b. Effect of PID *gain* in *PI* controller on load response for *near-integrating* process.

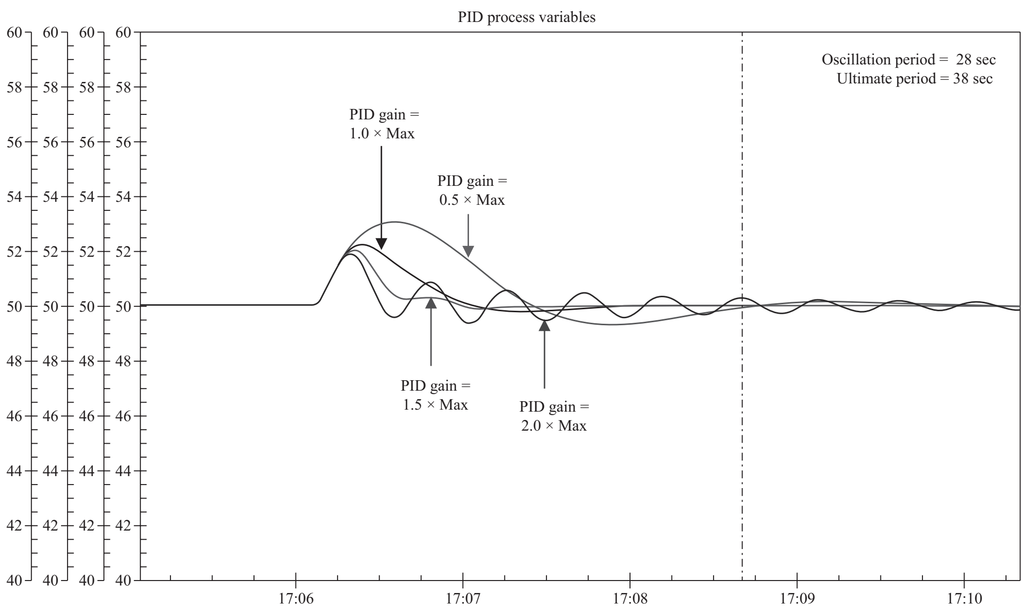


Figure 1.6c. Effect of PID *gain* in *PID* controller on load response for *near-integrating* process.

to setpoint momentarily falters which is indicative of too much proportional action. A curious situation develops as the PID gain is decreased for a controller with integral action. The response develops a slow oscillation in both Figures 1.6b and 1.6c but is slower in Figure 1.6c because the response for maximum allowable gain is smoother. In Section 1.6.5, test results

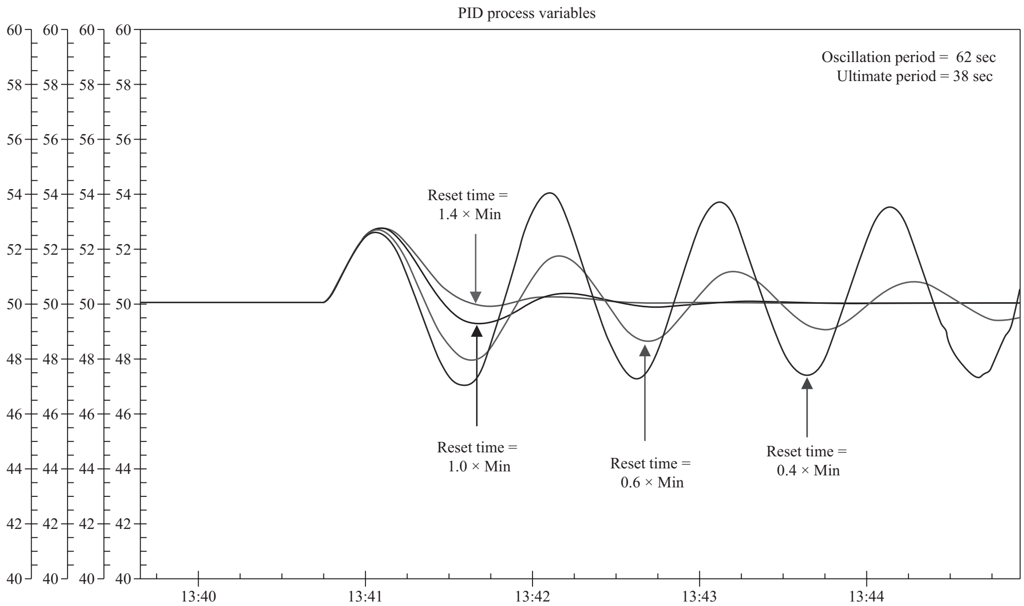


Figure 1.6d. Effect of *reset time* in *PI* controller on load response for *near-integrating* process.

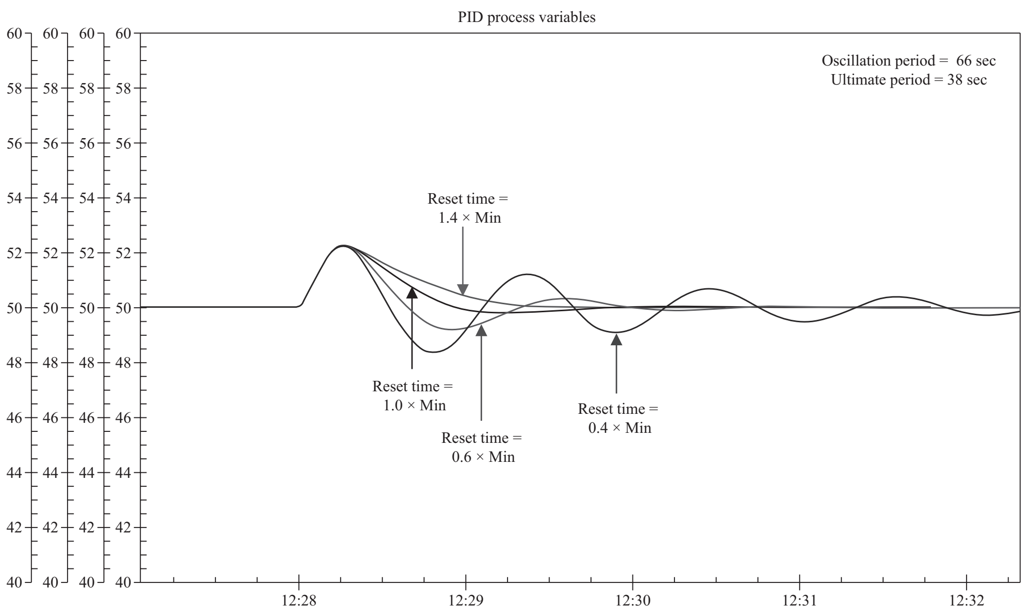


Figure 1.6e. Effect of *reset time* in *PID* controller on load response for *near-integrating* process.

will show this oscillation gets worse as the PID gain is decreased, a perplexing situation for most practitioners who were taught that a high PID gain is the cause of oscillations. If integral action is used, there is a window of allowable gains. To make the window larger, the reset time must be increased.

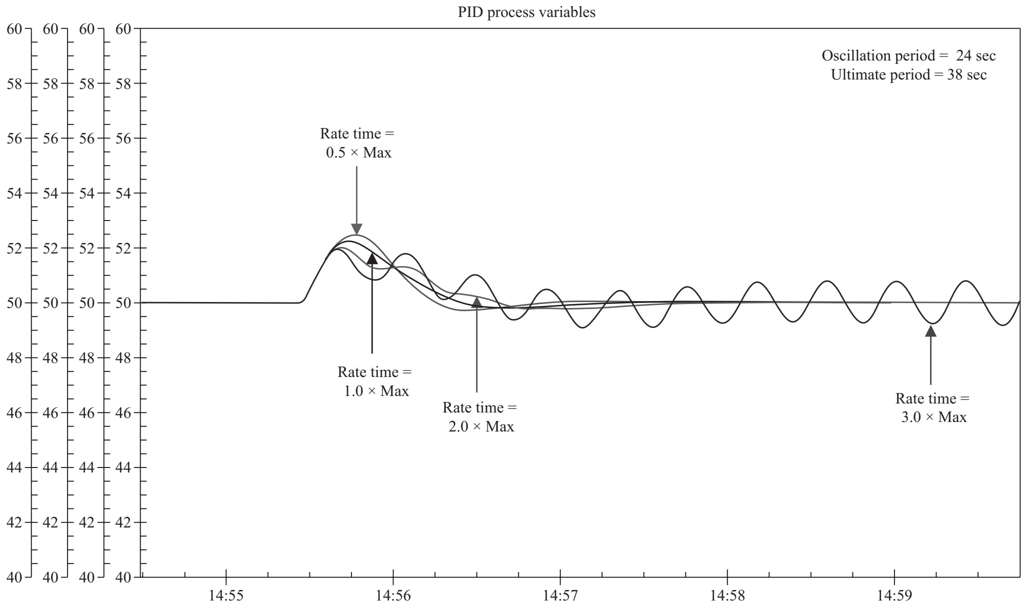


Figure 1.6f. Effect of *rate time* in *PID* controller on load response for *near-integrating* process.

The near-integrating process is more sensitive than the dead time dominant process to reset time as seen in Figure 1.6d in that unacceptable oscillations start sooner as the reset time is decreased and is noticeable even in the recommended minimum setting.

Figure 1.6e shows the addition of derivative action counteracts the effect of integral action to help stabilize the process, lessening the sensitivity to a decrease in reset time similar to what occurred in Figure 1.6c in terms of sensitivity to an increase in PID gain.

Figure 1.6f shows that a doubling of rate time causes the PV to start oscillating in the approach to setpoint from too much derivative action. For a tripling of the rate time, the oscillations start to grow. The oscillation period is about 40 percent less than the ultimate period, which is a simple diagnostic for too large of a rate time setting.

1.6.3 PERFORMANCE OF TUNING SETTINGS ON TRUE INTEGRATING PROCESSES

Figure 1.7a shows the response of the proportional-only controller for a true integrating process is about the same as for a near-integrating process except for larger amplitude of the oscillations when the PID gain is doubled. Proportional-only controllers or proportional plus derivative controllers are used on true integrating processes where the process response can only go in one direction. A one direction (unidirectional) response occurs when there is not a split range to provide a response in the opposite direction. Examples are batch vessels whose temperature is raised with steam (no coolant) or lowered with coolant (no steam), neutralizers whose pH is lowered with an acid (no base) or raised with a base (no acid), or concentrations of a product are increasing in the batch from processes involving evaporation, crystallization, or nonreversible

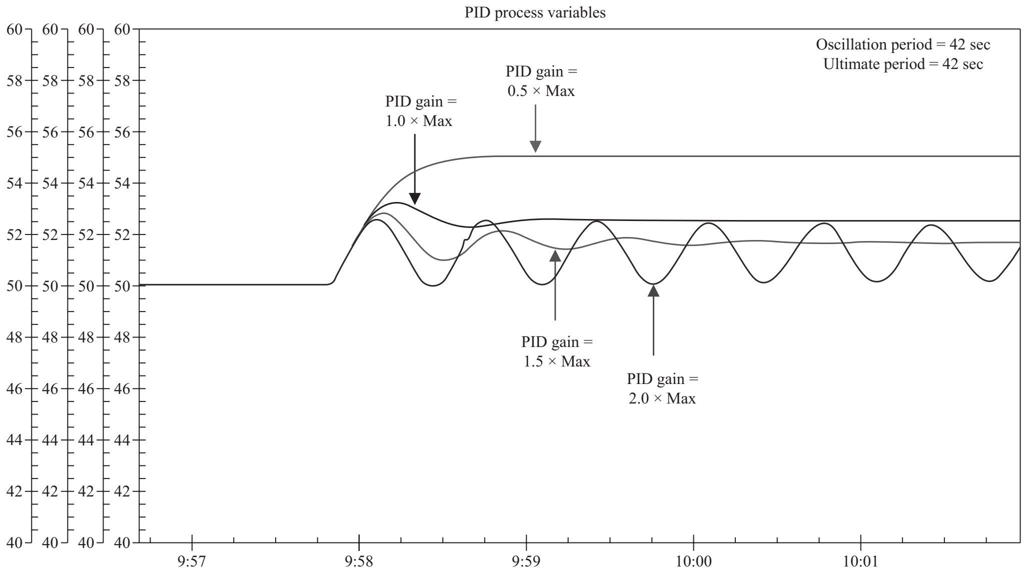


Figure 1.7a. Effect of PID *gain* in *proportional-only* controller on load response for *true integrating* process.

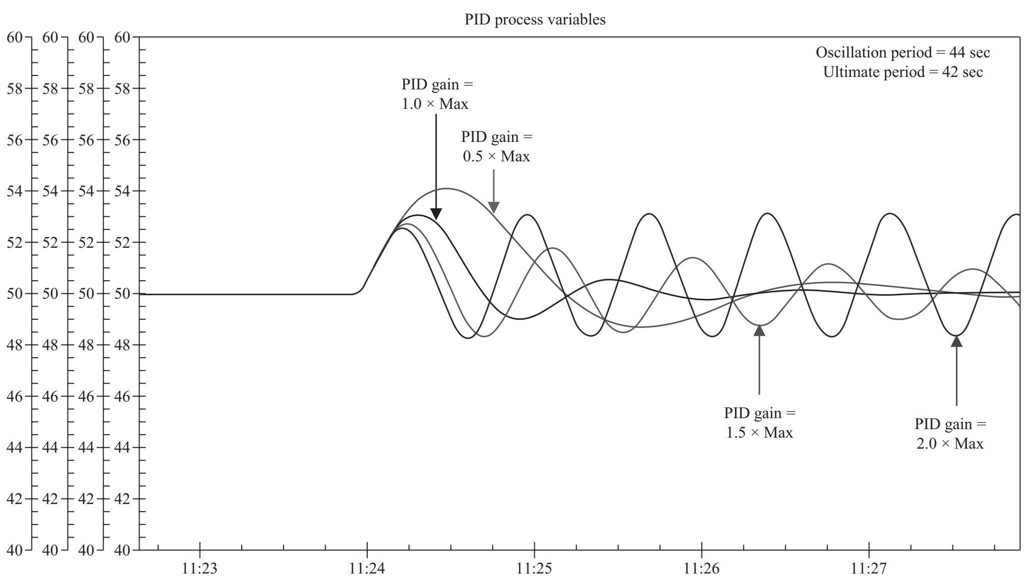


Figure 1.7b. Effect of PID *gain* in *PI* controller on load response for *true integrating* process.

reactions. The conversion of the controlled variable to the slope of the batch profile creates the opportunity for a response in both directions enabling the use of integral action to provide more repeatable batch profiles. Chapter 15 will discuss this and other techniques for innovative fed-batch control strategies that provide more predictable and efficient batches.

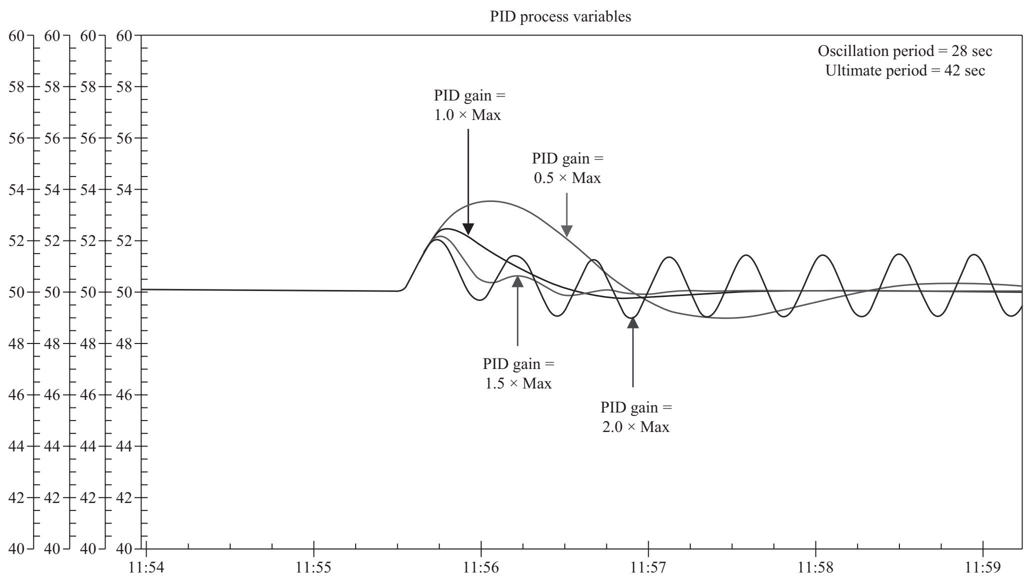


Figure 1.7c. Effect of *PID gain* in *PID* controller on load response for *true integrating* process.

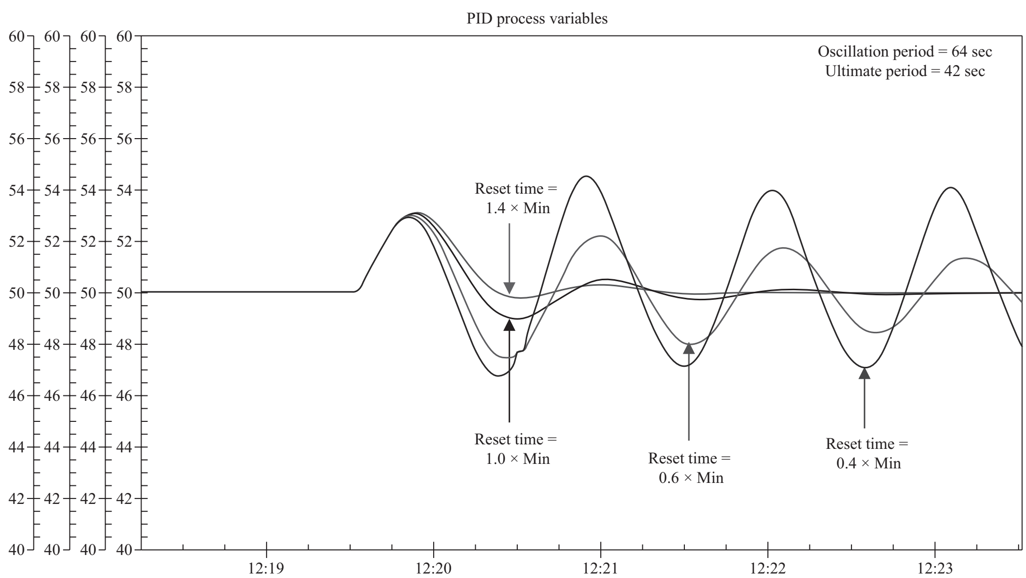


Figure 1.7d. Effect of *reset time* in *PI* controller on load response for *true integrating* process.

Figures 1.7a through 1.7f show a similar effect of tuning settings except the consequences are more severe for the true integrating process compared to the near-integrating process even though the open loop integrating process gain is the same.

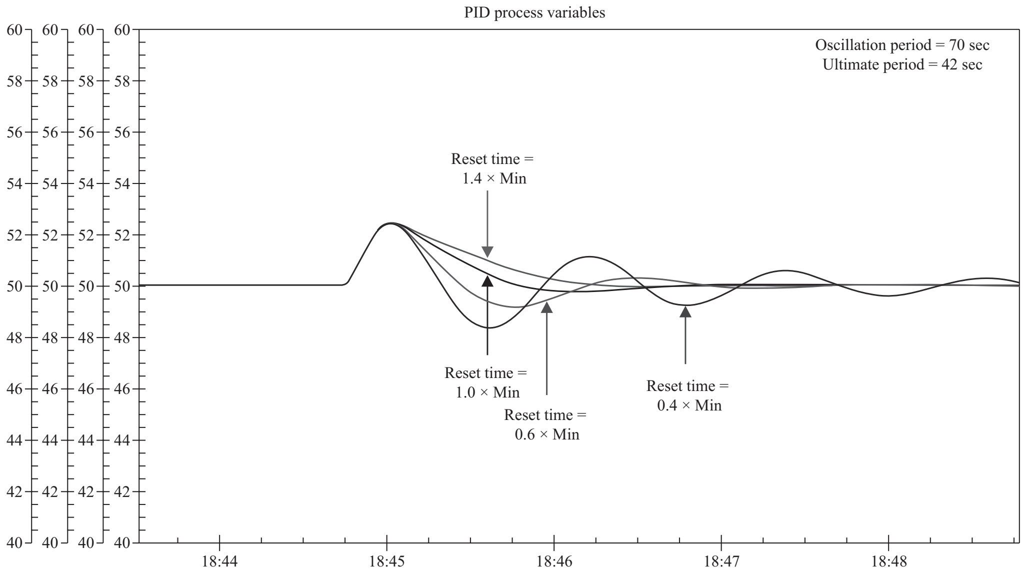


Figure 1.7e. Effect of *reset time* in *PID* controller on load response for *true integrating* process.

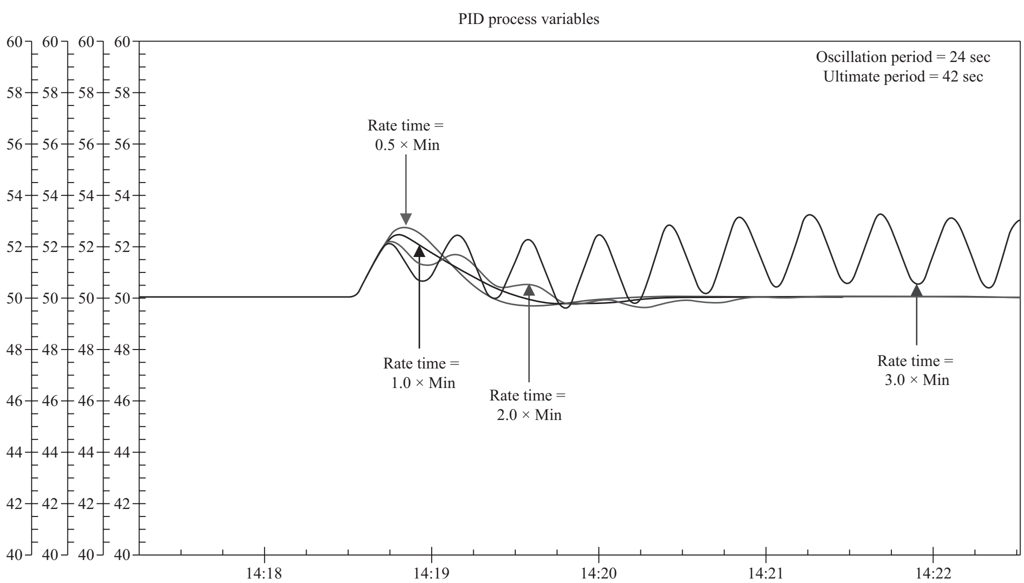


Figure 1.7f. Effect of *rate time* in *PID* controller on load response for *true integrating* process.

1.6.4 PERFORMANCE OF TUNING SETTINGS ON RUNAWAY PROCESSES

Figures 1.8a through 1.8f show the consequences of deviations from the best PID settings are more severe for runaway processes than true integrating processes even though the open loop integrating process gain is the same. Figure 1.8f shows a rather unexpected behavior for

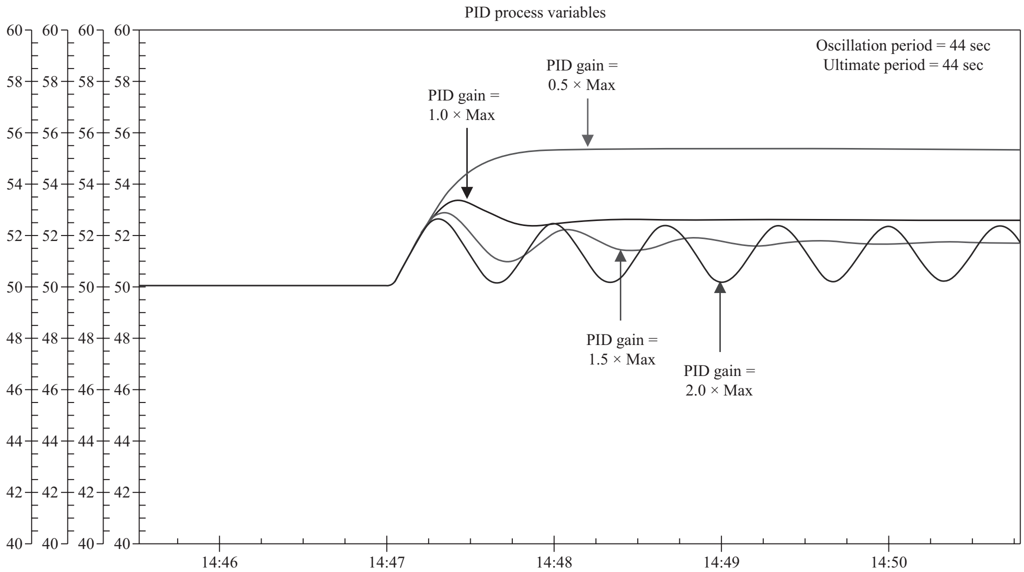


Figure 1.8a. Effect of PID *gain* in *Proportional-only* controller on load response for *runaway* process.

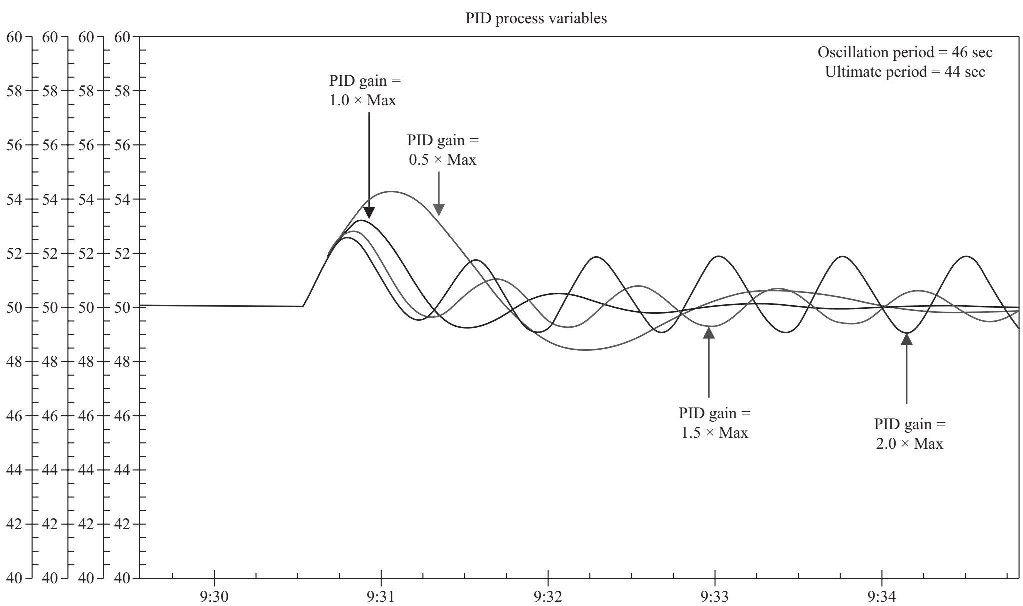


Figure 1.8b. Effect of PID *gain* in *PI* controller on load response for *runaway* process.

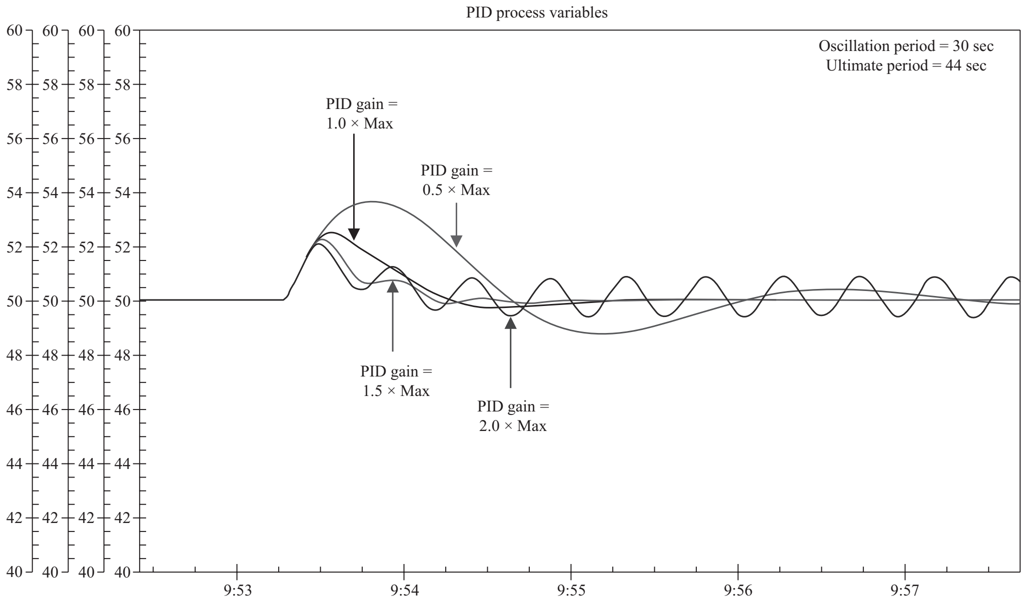


Figure 1.8c. Effect of PID *gain* in *PID* controller on load response for *runaway* process.

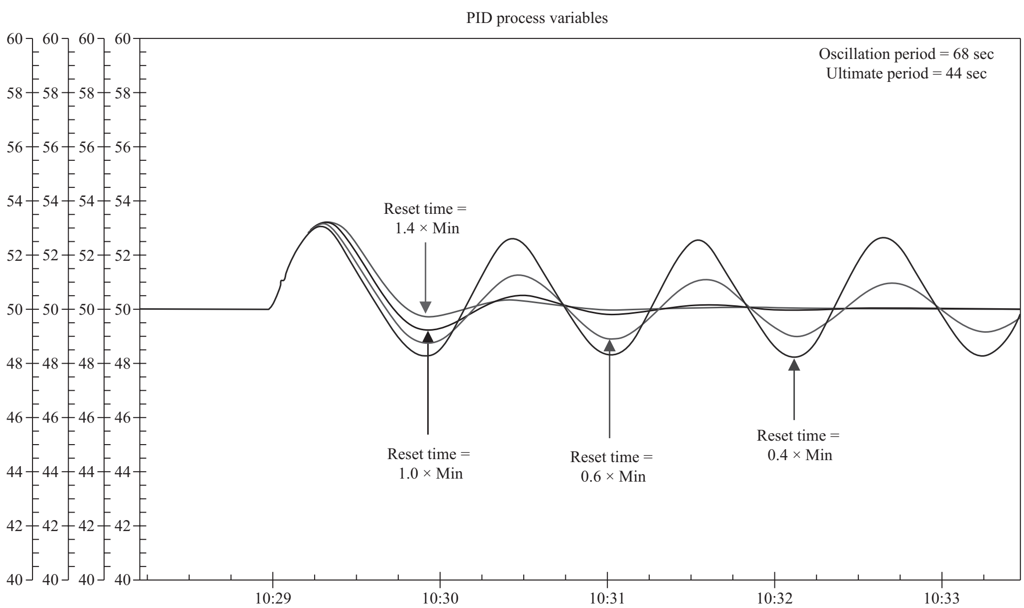


Figure 1.8d. Effect of *reset time* in *PI* controller on load response for *runaway* process.

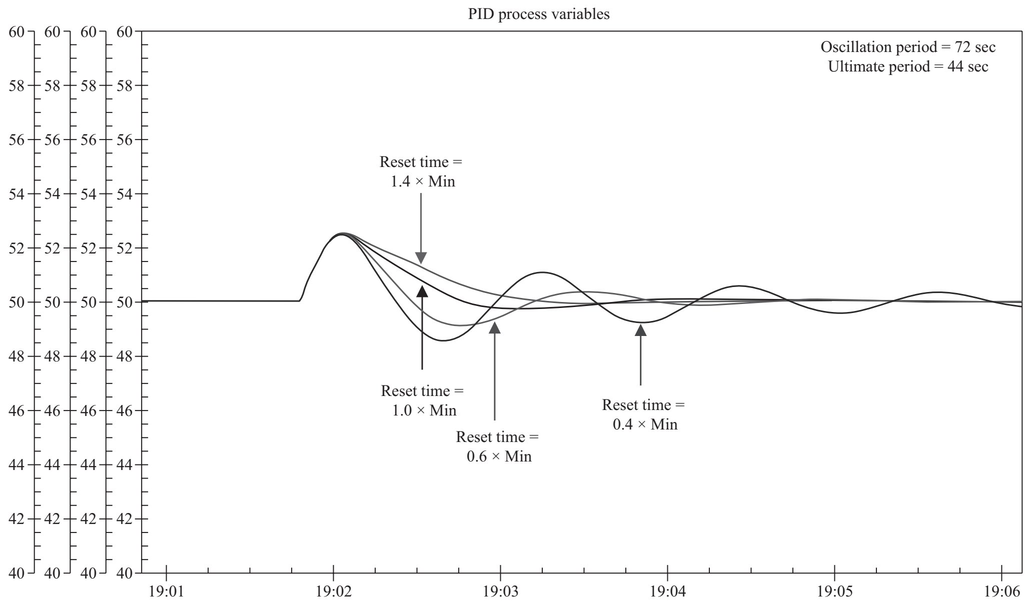


Figure 1.8e. Effect of *reset time* in *PID* controller on load response for *runaway* process.

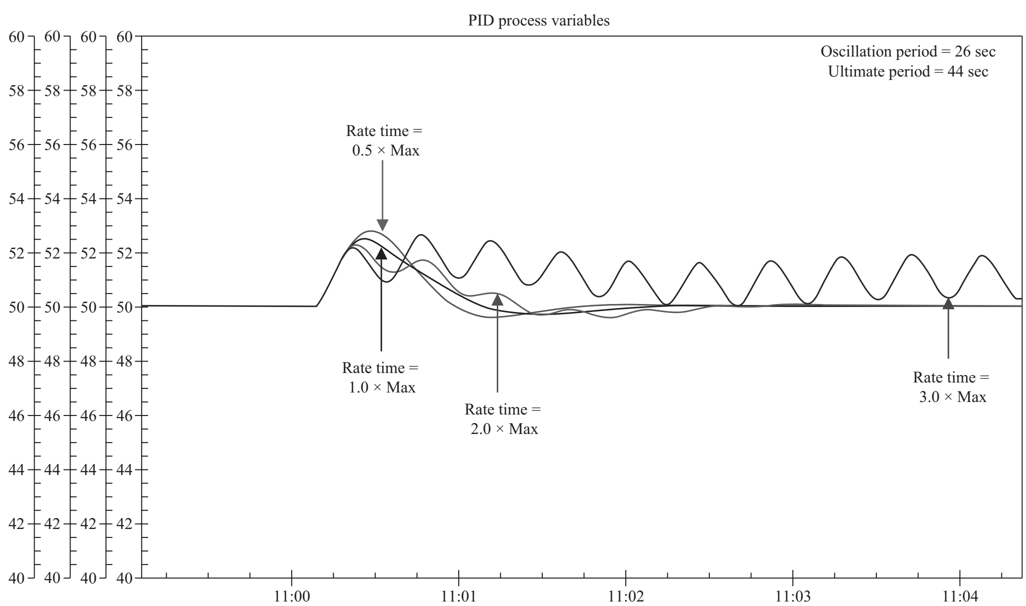


Figure 1.8f. Effect of *rate time* in *PID* controller on load response for *runaway* process.

a rate time that is triple the maximum in terms of the growing oscillation moving away from the setpoint. Apparently the negative feedback action is not sufficient to overcome the positive feedback action in the process to keep the oscillation centered on the setpoint.

1.6.5 SLOW OSCILLATIONS FROM LOW PID GAIN IN INTEGRATING AND RUNAWAY PROCESSES

If the PID gain is decreased below the maximum and the reset time is not proportionally increased (Equation 1.5c), near-integrating, true integrating, and runaway processes will start to oscillate. Figures 1.9a, b, c show that the oscillations become quite severe if the gain is decreased to be 10 percent of the maximum. The oscillations will always decay but the period and amplitude increases as the gain is decreased. For a PID gain that is 10 percent of the maximum, the oscillation period approaches 10 times the ultimate period. The incredibly large period (e.g., 40 times the dead time) is an excellent diagnostic clue to the problem.

Since the maximum PID gain on these processes can exceed values way beyond the comfort zone of operations and even process control practitioners (e.g., 100), values less than 10 percent of the maximum are quite common. If the common concept of decreasing the gain will help reduce oscillations is employed, the gain setting gets even lower. The result is often a PID gain setting that is one percent of the maximum resulting in nearly undamped oscillations.

Some of the issues of using a high PID gain can be readily addressed. The use of setpoint rate limits on the PID and the AO block can slow down the changes in the PID output so as not to upset operations or other loops. The PID must have external reset turned on when the PID output goes to another PID with setpoint rate limits or to an AO with setpoint rate limits to prevent the PID output changing faster than a downstream block can respond.

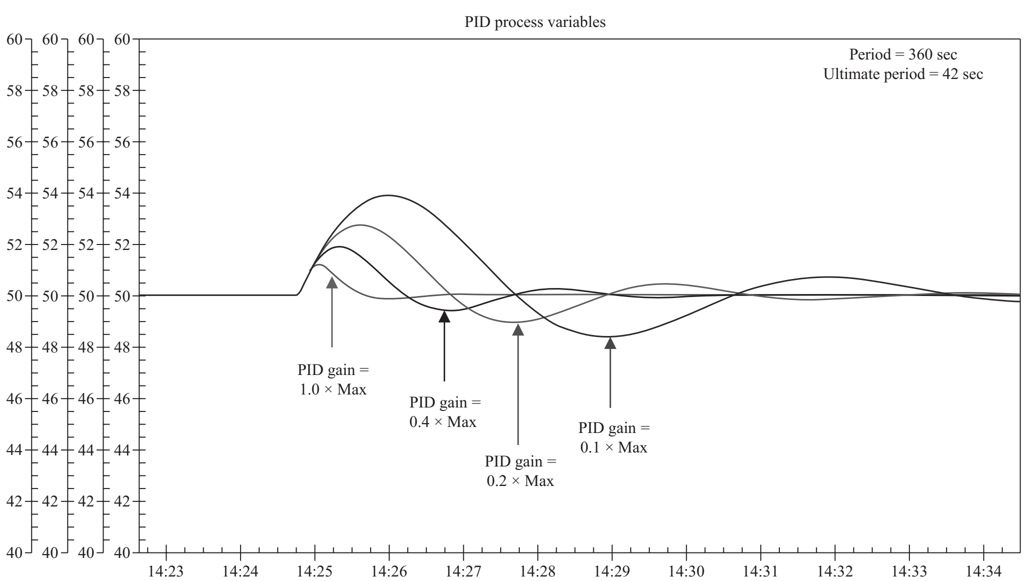


Figure 1.9a. Effect of *low PID gain* in PID controller on load response for *near-integrating* process.

If the PID gain cannot be increased, the reset time must be proportionally increased. This has led to the observation that most integrating and runaway processes benefit from increasing the reset time by a factor of 100 or more.

The Figures 1.9a, b, c show that the problem gets worse as the process loses internal feedback. The oscillations are more severe as you move from near-integrating to true integrating to

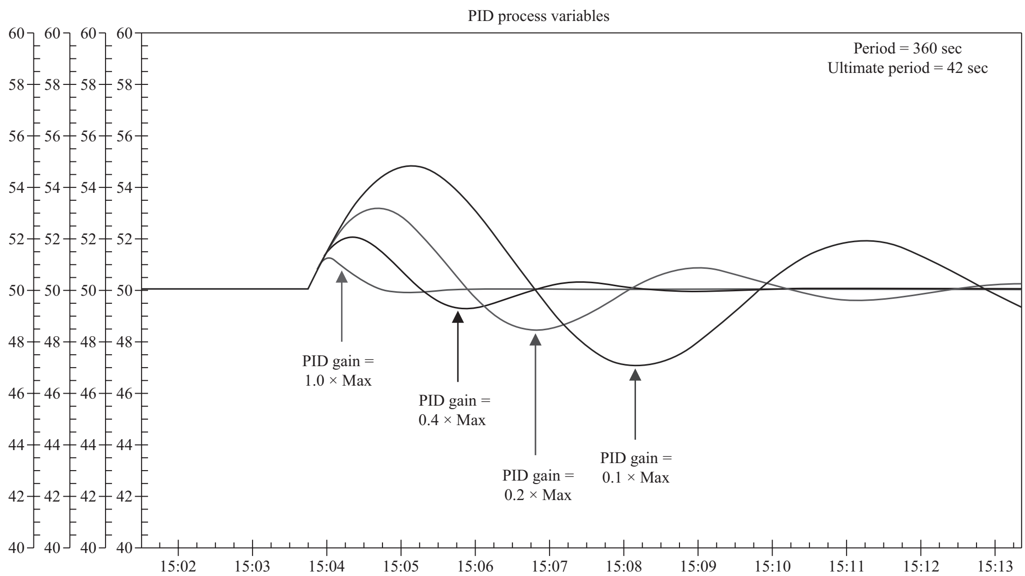


Figure 1.9b. Effect of *low PID gain* in PID controller on load response for *true integrating* process.

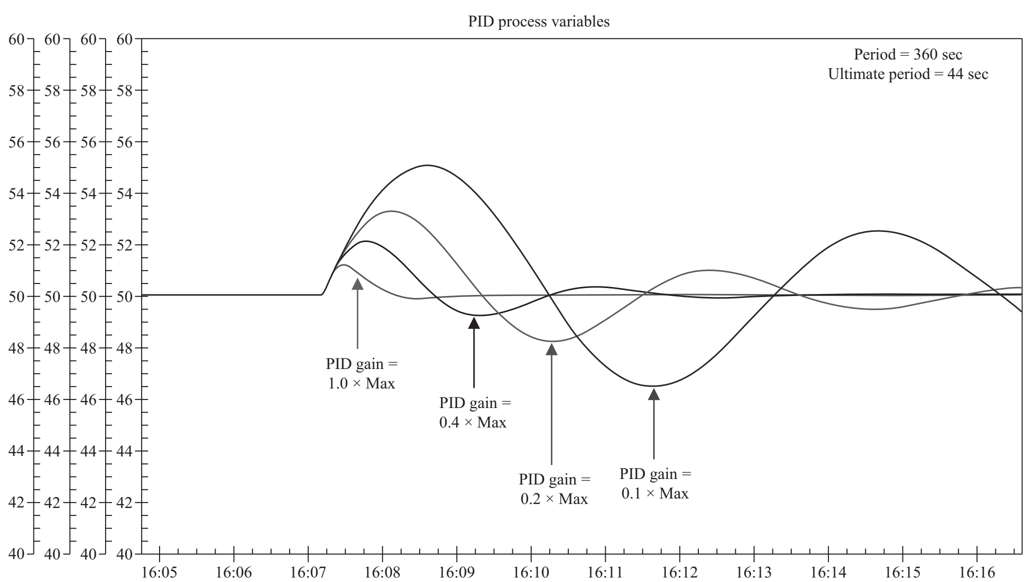


Figure 1.9c. Effect of *low PID gain* in PID controller on load response for *runaway* process.

runaway. This potential problem has led to some users going to a structure of PD on error and no I. If a structure does not have integral action, the bias (controller output when the PV is at setpoint) must be set properly.

1.6.6 PERFORMANCE OF TUNING METHODS ON VARIOUS PROCESSES

Figures 1.10a, b, c, d show how four major tuning methods perform on four major types of processes. The Lambda and gamma settings are at the minimum suggested value of one dead time. Lambda tuning switches to using integrating process rules for near-integrating processes as recommended by the Lambda tuning founder Bill Bialkowski. The unified methodology in this book advocates the switch to integrating tuning rules when the primary process time constant to dead time ratio is larger than four which is a smaller ratio than originally envisioned. The IMC and SIMC tuning could have benefited from this practice.

The low limit to reset time depicted in Equations 1.6b and 1.7b and the use of derivative for self-regulating processes or the limits placed on rate time in Equations 1.6f and 1.7f are not part of the standard Lambda tuning rules and were not consequently used in the tests. The performance of Lambda tuning for load disturbances in the test results would have been close to that for the SCM if the modified equations were used and Lambda was decreased to be about 0.6 times the dead time for self-regulating processes.

For dead time dominant processes, proportional action exceeds the integral action in IMC tuning because the integral time depends upon both the primary time constant and the dead time causing a faltering in the approach to setpoint seen in Figure 1.10a. For Lambda and SIMC tuning, the proportional action is lower than reset action since the reset time is so small from being set equal to the primary time constant. The SCM tuning has significantly more proportional action due to a low gain limit imposed in the tuning method but is fortuitously in balance with the integral action. This balance may not be true for other primary time constant to dead time ratios.

For the moderate self-regulating process, Figure 1.10b shows the methods give similar results except the peak error is larger for the Lambda and SIMC methods. A closed loop time constant (Lambda or gamma) set equal to 0.6 times the dead time, would have given a peak error comparable to that of the SCM method. As the primary time constant becomes much larger than the dead time, the Lambda and SCM methods provide the fastest recovery as seen in Figure 1.10c. The IMC and SIMC have a protracted response taking about five times longer to reach setpoint as a result of the reset time being dependent upon the large primary time constant. Surprisingly for the true integrating process in Figure 10.1d, the IMC response becomes oscillatory. On the other hand, the SIMC response is slow and the peak error large. The performance for a runaway process is not shown because there are no IMC or SIMC tuning rules. The Lambda tuning rules for integrating processes is a practical solution. The SCM tuning rules for runaway processes is based on knowledge of the positive feedback time constant and open loop runaway process gain. While it is theoretically possible to calculate these dynamic terms from first principle relationships, in practice integrating process tuning rules are used based on the maximum initial ramp rate of the process in the right direction for a setpoint change. Consequently, the results of SCM and Lambda tuning on a runaway process would be about the same as seen for true integrating processes. For runaway processes, it is critical that Lambda is not be set much beyond the largest possible dead time because of the danger of a runaway from insufficient negative feedback action.

Too much importance should not be placed on the performance seen in the figures because tuning settings need to be adjusted to deal with nonlinearities, interactions, resonance, and other objectives such as coordination and a consistent response to the setpoint being manipulated by upper loops or an advanced control system (e.g., Model Predictive Controller). Chapter 2 will use the Lambda tuning method in a unified methodology to better meet the requirements in a broad spectrum of applications.

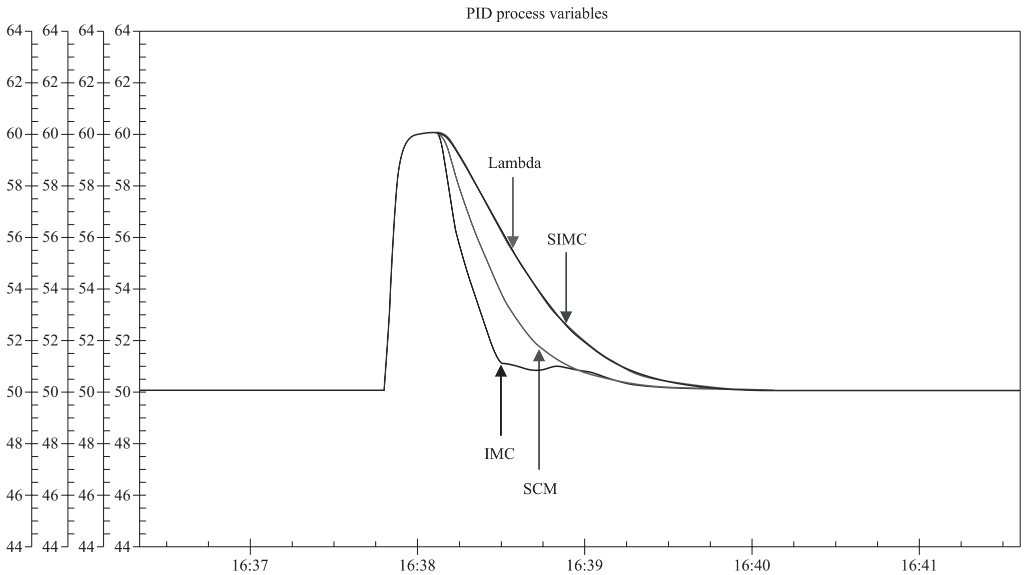


Figure 1.10a. Effect of *tuning methods* in *PI* controller on load response for *dead time dominant* process.

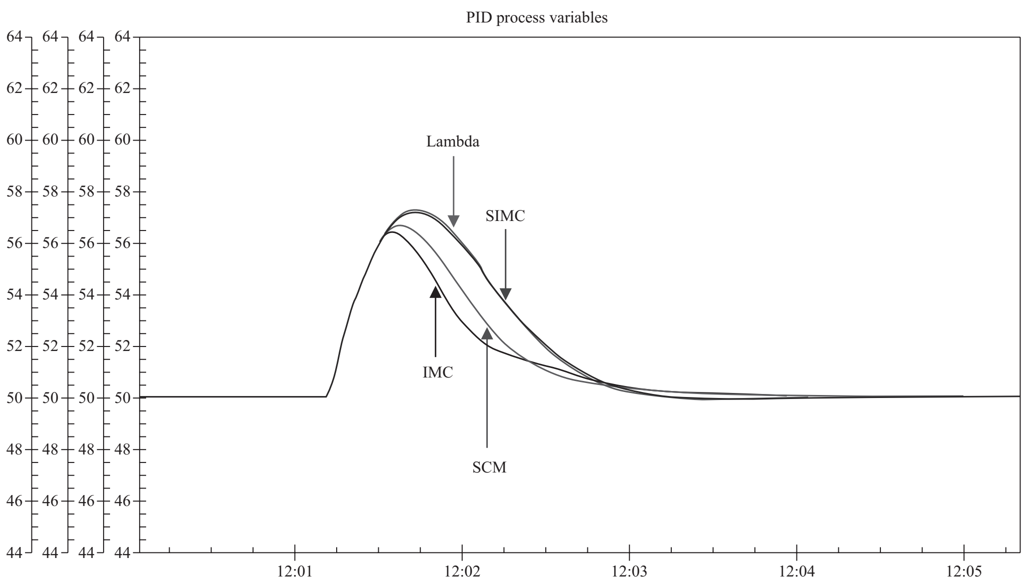


Figure 1.10b. Effect of *tuning methods* in *PID* controller on load response for *moderate self-regulating* process.

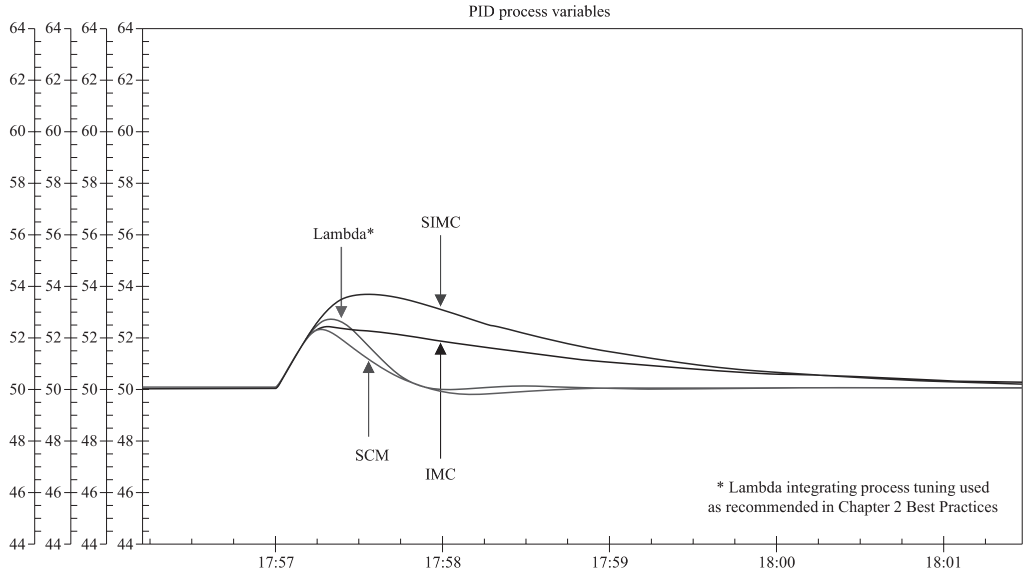


Figure 1.10c. Effect of *Tuning Methods* in *PID* controller on load response for *near-integrating* process.

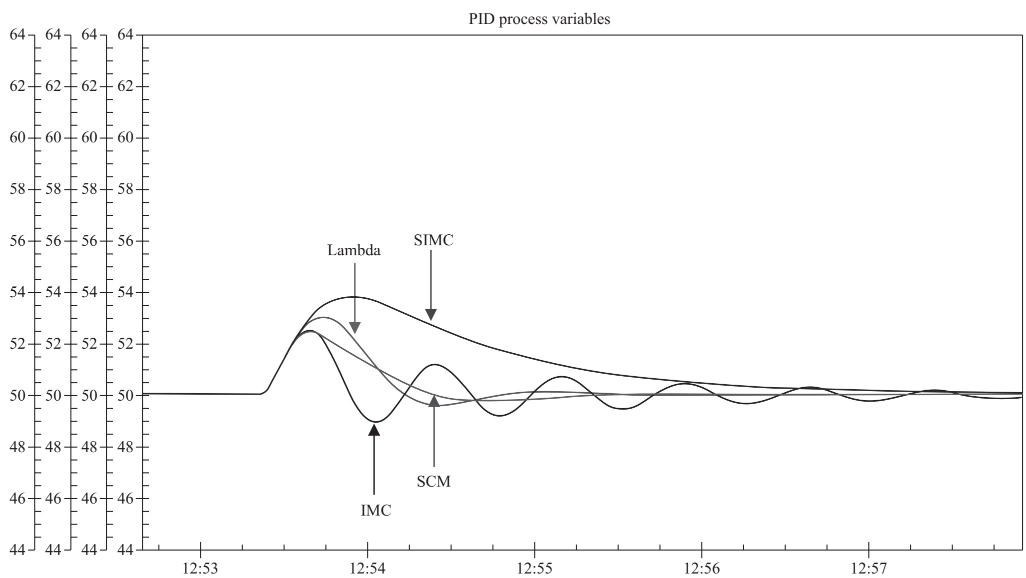


Figure 1.10d. Effect of *tuning methods* in *PID* controller on load response for *true integrating* process.

KEY POINTS

1. The abrupt action from the proportional and derivative modes is needed for loops with a slow gradual response (large primary time constant) to recover from upsets and to reach setpoints faster.
2. The smooth action from the integral mode is needed for loops with a fast abrupt response (small primary time constant) to avoid amplifying noise and upsetting other loops by sudden PID output changes.
3. Too much integral action is found in slow loops because of human impatience, lack of recognition of delayed effect of actions due to dead time, digital displays of setpoints and PVs, and trend recordings with dumb time scales.
4. Too much integral action is found in near-integrating, true integrating, and runaway loops because the reset time was not proportionally increased to cancel out the decrease in PID gain. The discomfort with steps in the PID output from setpoint changes results in gain setting orders of magnitude too low that require the reset time to be orders of magnitude higher to keep the product of the PID gain and reset above the low limit (Equation 1.5c). The lack of understanding that oscillations much slower than ultimate period grow as the PID gain is decreased for these processes, causes users to further decrease the gain.
5. Negative feedback is needed for process control. The less negative feedback there is in the process, the more there needs to be in the PID controller. Proportional mode action provides the most immediate and aggressive negative feedback.
6. The process feedback action becomes more problematic as you go from near-integrating to true integrating (zero process feedback) to runaway (positive process feedback).
7. A PI controller provides the tightest control for load upsets in loops with extremely fast response times (e.g., polymer pressure and incinerator pressure loops with VSD or surge loops with fast valves) by the use of fast integral action.
8. A PID controller provides the tightest control for load upsets in loops with extremely large primary process time constants (near-integrating), a true integrating, or runaway response. The derivative mode provides some anticipatory action. The rate time is set equal to the secondary time constant. If the secondary time constant is not identified, the rate time is set equal to half the loop dead time.
9. Most tuning settings started out addressing moderate self-regulating processes.
10. Most of the disagreement in tuning stems from the lack of understanding of the implications of the size of the primary process time constant and the fact that most disturbances in the process industry are on the process input as load upsets rather than on the process output as shown in the control literature.
11. The tuning for a load disturbance can be tested by momentarily putting the PID in MAN, making a step change in the PID output, and then putting the PID in AUTO. The step size should be about the same size as a typical total correction in the PID output to deal with disturbances but at least five times larger than measurement noise, valve backlash, and any resolution limit. The dead time can be estimated as the time it takes from the step change in PID output till a noticeable change in the PID PV.
12. For a load disturbance, if the return to setpoint is more than twice as slow as the initial excursion, the reset time is too large. If the PV oscillates with a period much greater than six times the dead time, the reset time is too small. For a setpoint change, an overshoot of the new setpoint is probably caused or aggravated by too small a reset time.

13. For a load disturbance, if the return to setpoint oscillates with a period between three and six times the dead time, the PID gain is too large. For a setpoint change, a hesitation (faltering) in the approach to the new setpoint is indicative of too large a PID gain.
14. For a load disturbance, if the return to setpoint oscillates with a period less than three times the dead time, the rate time is too large. For a setpoint change, an oscillation in the approach to the new setpoint is indicative of too large a PID rate time.

For the diagnostics in key points 12 to 14, we are assuming the ultimate period is about four times the dead time, which consequently excludes dead time dominant loops or integrating and runaway loops with large secondary time constants.

CHAPTER 2

Unified Methodology

2.1 INTRODUCTION

The proportional-integral-derivative (PID) controller is the workhorse of the process industry. The modern PID has an extensive set of underutilized features often due to the lack of understanding of the role and possibilities. Here, we provide a concise summary of the options and parameters and how they can be used for process control improvement. We conclude with a step-by-step list of solutions in a unified methodology to get the most out of your PID controller. Since the PID can only be as good as the measurements and final control elements, the essential aspects of these components are included in the methodology.

2.1.1 PERSPECTIVE

The PID dates back about 100 years. For over 50 years, the PID controller, like other instrumentation, was pneumatic. The implementation restrictions of using bellows, nozzles, flapper, links, and levers resulted in a Series Form and the positive feedback implementation of integral action in the PID. The interesting part of this story is that the positive feedback implementation had inherent advantages lost by many controllers in the transition to the electronic and digital age.

In the Series Form, the derivative action comes first. The output from the derivative mode is the input for the proportional and integral modes. In the pneumatic implementation, a positive feedback bellows for integral action acting on a flapper was balanced by a negative feedback bellows for proportional action. An adjustable three way valve with an exhaust to atmosphere provided the proportional band (PB) tuning setting. An adjustable needle valve in combination with the bellows volume gave a filter time constant that was the reset time tuning setting. The PB setting translated to the percent change in process variable (PV) or setpoint required to cause a full scale (100 percent) change in PID output. Some manufacturers converted to PID gain by dividing 100 percent by the PB. The reset time setting was the time required to cause a repeat of the proportional mode contribution to the controller output (e.g., minutes per repeat). Some manufacturers inverted this setting (e.g., repeats per minute).

When electronic analog controllers came on the scene, the Series Form was retained but the implementation of the integral mode drastically changed. Designers saw that an integrator could now be used for the integral mode as depicted in control theory text books. Most manufacturers went with an integrator rather than a positive feedback filter for integral action. The positive feedback implementation was confusing, causing the mind to go in circles, tracing integral mode. What these suppliers did not realize is that a valuable feature of external reset feedback was lost in the change to an integrator. This feedback is the key to suppressing oscillations from a wide variety of sources. The feature is also important for effective implementation of override control by stopping integral action when the manipulated variable is not responding or is not available.

The first systems for digital implementation of the PID largely copied electronic analog implementation. The Distributed Control Systems (DCS) developed in the 1980s typically used the Series Form and the integrator implementation of integral action.

In the modern DCS, the user can choose the ISA Standard Form, eliminating the interaction between modes associated with the Series Form when the derivative mode is used. The Standard Form eliminates the need to apply interaction factors to see the effective contribution from each PID mode in the time response. What some users do not realize is that this interaction inherently protects the user from an effective rate time becoming larger than the effective reset time. The tuning rules for the Series Form sometimes had the rate time equal to the reset time. If you set the rate time equal to or larger than the reset time in the ISA Standard Form, severe oscillations can develop from the interplay of the integral and derivative modes for most chemical process systems.

The modern DCS offers many different PID structures to set or moderate the proportional and derivative action on error and PV. The structures also enable the elimination of a mode (e.g., Proportional plus Derivative; no integral). In the analog controller days, a different structure often required a different model number PID.

2.1.2 OVERVIEW

Most electronic analog controllers and early vintages of DCS retained the Series Form but lost the positive feedback implementation of integral action important for external reset feedback. The modern DCS now uses the ISA Standard Form and offers external reset feedback. *For safety besides performance, the person doing the DCS implementation needs to be aware of the implications of PID Forms and integral mode design particularly in migration projects.*

The user has a large number of PID structures as configuration choice along with setpoint filters and rate limits. The blocks for control functions, such as split range and signal characterization, and the expression capability in a calculation block allow the user to achieve incredible flexibility by a simple configuration change. Ideas can be developed and tested in a matter of hours. The limit to innovation is largely the creativity of the configuration engineer or technician. This chapter seeks to open the mind to extensive possibilities with enough guidance to make the implementation practical.

2.1.3 RECOMMENDATIONS

1. Use the ISA Standard Form and remember to convert settings from Series Form as per Appendix K if derivative action is used.
2. Use the PID Two Degrees of Freedom (2DOF) structure where the beta and gamma setpoint multipliers are adjustable for the proportional and derivative modes, respectively,

to set the fraction of mode action on error instead of PV. The beta and gamma factors are set at about 0.5 to minimize rise time and overshoot. Alternately, the structure of "PI on error and D" on PV (beta = 1, gamma = 0) can be used with a lead-lag on the setpoint where the lag time is set equal to the reset time and the lead is one-fourth the lag time. The setpoint lead-lag method gives a step change in output but eliminates the additional kick and oscillation in the output associated with rate action in the 2DOF structure. If rise time is not important, a setpoint filter can be used where the filter time is set equal to the reset time or beta and gamma factors can both be set equal to zero in 2DOF to minimize overshoot. Exception: A PID with a one direction (unidirectional) response (PV can only increase or decrease) needs a "PD on error, no I" structure or the PV will overshoot and be unable to return to setpoint.

3. Use output tracking, remote output, or manual mode and a filtered rate limited full throttle response for the shortest possible time to reach and settle out at a setpoint or to coordinate loops to make a startup or fed-batch smoother or faster.
4. Use output tracking, remote output, or manual mode to give fast preemptive action to prevent an abnormal situation (e.g., compressor surge and Resource Conservation and Recovery Act [RCRA] pH violation).
5. Use a feedforward summer to preemptively mitigate fast large disturbances and to decouple interactions that cannot be sufficiently downplayed by tuning.
6. Use a lead-lag and dead time block to provide dynamic compensation of the feedforward so the feedforward correction arrives at the same point and time in the process as the disturbance or interaction.
7. For flow feedforward, provide an external ratio and bias station so the user can set the desired ratio and see the actual ratio of the manipulated flow to feed flow.
8. Use external reset feedback to prevent oscillations from slow manipulated variables (e.g., slow secondary loops, valves, or variable speed drives [VSDs]), deadband or resolution limits (e.g., valves with stiction or backlash, VSDs with deadband, and I/O cards with less than 16 bits or a wide signal range), and to enable a slow approach to optimums and fast getaway for abnormal conditions by the use of setpoint rate limits on the manipulated variable. The rate limits provide an essential advantage seen in Model Predictive Control (MPC) where the size of the move (change in manipulated variable) is limited for each execution by a tuning parameter called move suppression (e.g., penalty on move). Normally, the MPC move suppression is fixed for both directions (increase and decrease). The PID can provide directional move suppression for optimization and protection.
9. Use Up and Down rate limits on the setpoint of an Analog Output (AO) block and PID to provide directional move suppression, enabling a cautious approach to an optimum and a fast getaway to prevent abnormal operation.
10. Use an enhanced PID described in this book to prevent oscillations from discontinuous updates (e.g., at-line and offline analyzers and wireless devices).
11. Use setpoint feedforward to minimize rise time when the PID gain is relatively low (e.g., product of PID gain and open loop gain is less than one).
12. If a fast valve position readback for external reset feedback or the enhanced PID with a threshold sensitivity limit to screen out noise is not available, use integral deadband to suspend integral action when PV is within the desired range to stop oscillations from stick-slip and backlash.

2.2 PID FEATURES

The modern PID controller can have over 50 options or parameters for PID control with another 50 options or parameters for auto tuning and adaptive tuning. Most of the parameters are for monitoring the function of PID features. Many of the remaining parameters that are adjustable only come into play when special options, such as nonlinear and feedforward control, are employed. While suppliers may get concerned about the PID being viewed as overly complex, users appreciate the visibility and flexibility afforded realizing that the options and parameters are empowering. Nothing was more frustrating to me during my 33 years as user than the inability to turn on and off, observe, and adjust the functionality of a feature. Developers who remove user access thinking they have addressed all possibilities do not realize the incredible differences between the food and beverage, oil and gas, pharmaceutical, pulp and paper, power, petrochemical, and specialty chemical industries.

The parameters that must be set for basic control include control action (direct or reverse), mode, output action (increase-open or increase-close), output limits, anti-reset windup limits, setpoint limits, gain, reset, and rate. Note that output and anti-reset windup limits in newer systems have tended to move from being in percent of scale to being in engineering units of the PID output scale. Not knowing the units and setting the proper values of these parameters can lead to a disabled process and possibly unsafe conditions.

While most of this book is on the proper setting of gain, reset, and rate parameters, Chapter 1 discussed some of the other basic control parameters, and in the following section and chapters the use of options and parameters are covered that enable the PID to provide an incredible range of solutions.

2.2.1 PID FORM

PID controllers developed before this century used the Series Form also known as the interactive algorithm where the derivative mode is computed first in series with proportional and integral modes. The Series Form minimized the cost and complexity of analog controllers. Early digital controllers retained this Form to give the same feel in tuning and enable the use of the same tuning settings. Figure 2.1 is the block diagram for a Series Form with a conventional integral mode. Pneumatic and many electronic controllers actually used a positive feedback implementation of the integral mode as described in Section 2.2.2 on external reset feedback. The conventional integral mode depicted in Figure 2.1 principally appeared with the advent of digital controllers where an integrator could easily be implemented, coordinated, and limited.

In the Series Form, the derivative mode result is the input to the proportional and integral modes (See Appendix K difference equations). The advantage of the Series Form is that interaction factors prevent the effective rate time from becoming greater than effective reset time that could cause instability. As the rate time setting increases, the effective gain and rate time decreases and the reset time increases. The primary disadvantage of the Series Form is the lack of understanding of the interaction factors detailed in Appendix K. There is also greater magnification of noise from multiplicative effect of the derivative and proportional in series especially if there is poor measurement resolution and no signal filter or derivative filter.

Most modern PID controllers use the ISA Standard Form also known as the Ideal Form. The contribution from each of the modes is computed in parallel as shown Figure 2.2, the block

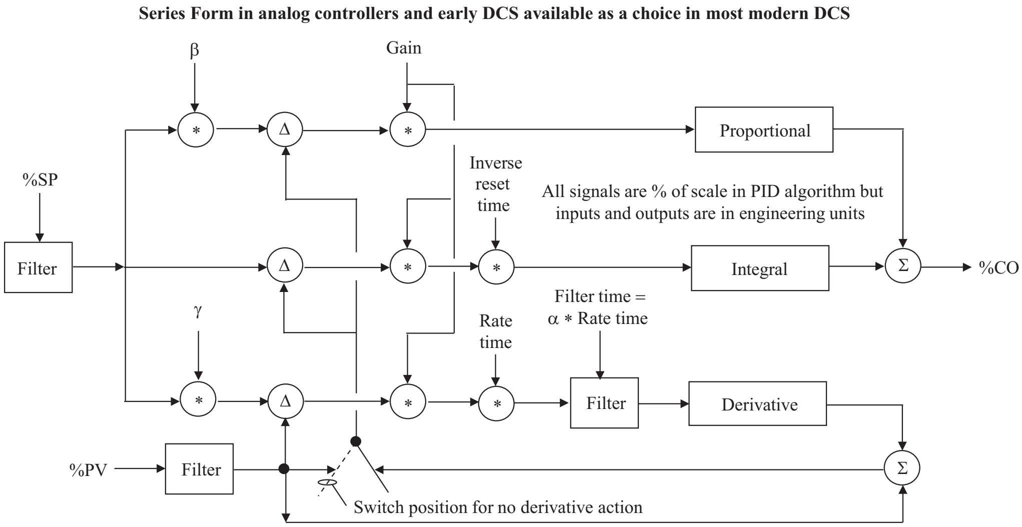


Figure 2.1. Series form PID with conventional integral mode.

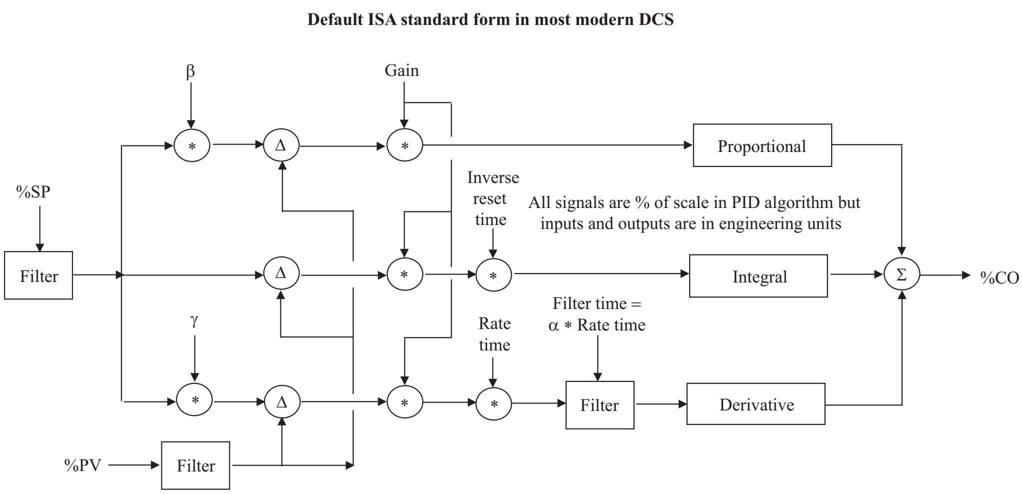


Figure 2.2. ISA standard form PID with conventional integral mode.

diagram for the ISA Standard Form with a conventional integral mode. If the rate time is zero, the ISA Standard Form and the Series Form are effectively the same if the implementation of the integral mode is identical.

The ISA Standard Form eliminates interactions between the modes in the time domain leading to the name noninteracting. Often not recognized are the interactions between the modes in the frequency domain of ISA Standard Form. The Laplace transform of the ISA Standard and Series algorithm reveal interaction and no interaction, respectively, between the terms. The label *noninteracting* is technically inappropriate in that while the ISA Standard Form is noninteracting in the time domain, this Form is interacting in the frequency domain. Correspondingly the Series algorithm is interacting in the time domain but noninteracting in the frequency domain. Both Forms have a gain multiplier in the contribution from each mode, which is not the case for the Parallel Form.

In the Parallel Form, the contribution from each mode is computed in parallel but the gain tuning setting only affects the proportional mode (see Figure K.3 in Appendix K). The tuning setting for Parallel integral and derivative modes are sometimes called integral and derivative gains rather than reset and rate times. Since there is no multiplication by the gain setting for the integral mode, the units of repeats per minute or seconds per repeat are meaningless. The units for integral and derivative gains may not even be given.

The Parallel Form is rarely used in industrial systems but is commonly seen in control theory textbooks. Consequently the Parallel Form is mistakenly programmed into dynamic models for control studies and operator training systems causing potentially disastrous results if tuning settings are moved from simulations to industrial systems since the difference in integral and derivative settings can be quite dramatic. For PID controllers with high gain settings, the Parallel Form reset and rate tuning settings could be an order of magnitude or more larger than what was intended.

Appendix K shows the equations to convert tuning settings between Series, ISA Standard, and Parallel Forms. The user should use these equations paying close attention to units because some proportional mode tuning settings are a PB with units of percent instead of a dimensionless gain. The integral mode tuning settings can be in repeats per time unit or the inverse that is time units per repeat for the Series and ISA Standard Form. For this inverse, the setting is simply given in time units instead of time units per repeat. The derivative mode setting is in time units. The time units for the integral and derivative mode tuning settings can be minutes or seconds. In this book, the ISA Standard Form is assumed and used to generate the test case results. The proportional mode tuning setting is a dimensionless gain. The integral mode tuning setting is seconds per repeat (simply given as a reset time in seconds). The derivative mode tuning setting is a rate time in seconds.

The beta and gamma factors in the block diagrams are setpoint weights for the proportional and derivative modes, respectively, discussed in Section 2.2.3 in terms of setting PID structure. The alpha factor in the derivative mode is the fraction of the rate time that becomes the time constant of a filter applied to the derivative mode input. If the filter is missing or the factor is zero, the derivative mode output can be a full scale spike for a step change in the mode input. Alpha factors range from 0.1 to 0.2 with a typical default value of 0.125.

2.2.2 EXTERNAL RESET FEEDBACK

External reset feedback is a feature made possible by the positive feedback implementation of integral action. As described in the Perspective, pneumatic controllers did not have integrators but could provide a filter by the use of a bellows (capacitance) and a restriction (resistance) (e.g., needle valve) where the adjustment of the restriction set the filter time constant. To create integral action, the output of the PID controller is fed back through a filter and added to proportional mode output creating positive feedback. The natural question is how does the algorithm prevent the continual ramp of the PID output from positive feedback? When the error is zero, the output of the proportional mode is zero. This means the filter output must equal the filter input. The external reset path hence reaches a steady state stopping positive feedback. The literature shows a block diagram of external feedback in a Series Form with a PID on error

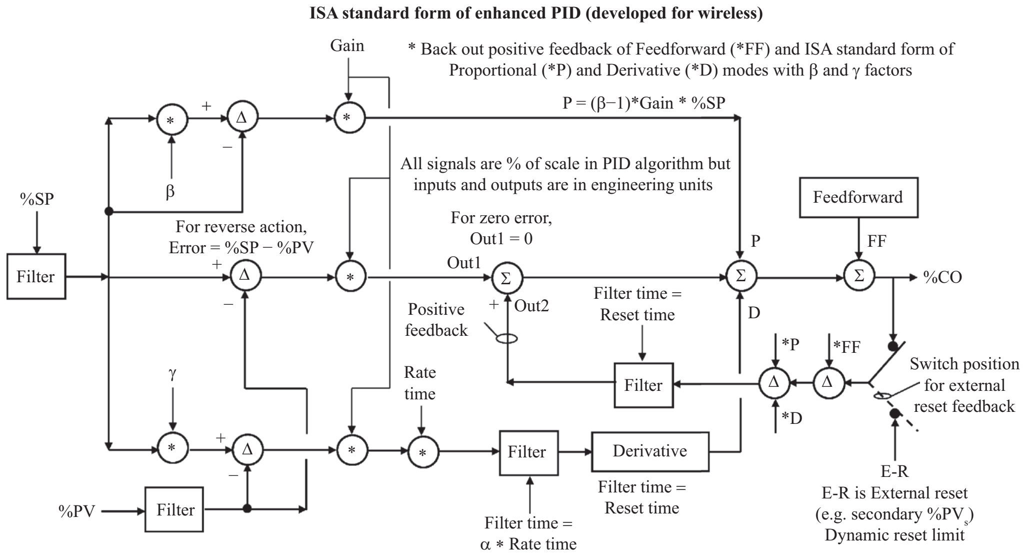


Figure 2.3. ISA standard form PID with positive feedback integral mode enabling external reset feedback (dynamic reset limit).

structure because this was the original implementation. In these diagrams, derivative and then proportional mode are computed in series with the positive feedback added to proportional mode output. The depiction for the ISA Standard Form becomes complicated particularly as you try to show the computation of modes in parallel with the effect of beta and gamma factors. Figure 2.3 is a back engineered best effort to depict how signals would be computed and added to the PID output and then subtracted from the positive feedback filter input (*P, *D, and *FF) to account for PID Structure and feedforward action. To include the effect of the beta factor in the proportional mode but not the integral mode, the term P = (1 – β) * Gain * %SP is added to the output to provide the summation of modes.

The real power of the positive feedback implementation comes into play when the input to the filter does not come from the PID output but from whatever the PID is manipulating. When the switch is placed in the external reset feedback position, the signal is typically setup to be a PV from a secondary loop in a cascade control system, an AO block, a VSD, or control valve. In some controllers this option is termed *dynamic reset limit*. By using the PV of what the PID is manipulating as external reset feedback, the PID output cannot change faster than the PV responds. Thus, oscillations are inherently prevented from a secondary loop that is slower than the primary loop or the use of rate setpoint limits in an AO or secondary loop to provide move suppression. Furthermore, if the lower PV is not responding, integral action will stop. Thus, the PID output will freeze for a broken wire, loss of wireless communication, a measurement that does not change due to a threshold sensitivity limit or resolution limit, a coated sensor, or a periodic result (e.g., at-line analyzer or off-line analyzer), or a control valve that does not move due to plugging, backlash, or stiction. The stopping of integral action can prevent limit cycles from dead band, stick-slip, threshold sensitivity limits, and resolution limits for many types of applications. Unnecessary crossings of the split range point can be minimized by directional move suppression and the elimination of limit cycles. Finally, external reset feedback enables an enhanced PID where the PID algorithm is only executed when the measurement changes with the exponential response of the filter computed based on the elapsed time since the last measurement update. The enhanced PID enables at-line and even off-line analyzers to be used for closed loop control without having to retune the PID controller. All of these benefits that originate from external reset feedback will be discussed and test results given at the corresponding sections in this book.

2.2.3 PID STRUCTURE

PID structure choices use beta (β) and gamma (γ) set point weighting factors shown in Figures 2.1–2.3. A controller that has both factors adjustable has a "two degrees of freedom" *(2DOF) structure.* Other structures have the β and γ factors set equal to zero or one. The user can also omit a mode entirely to get P-only, I-only, ID, and PD control with various assigned factors. PI control is achieved by simply setting the derivative (rate) time to zero. In general, the user must not set the controller gain equal to zero in an attempt to get I-only or ID control or set the integral (reset) time to zero in an attempt to get P-only or PD control. Note that the use of P-only or PD control requires additional choices of how to set the bias and its ramp time. Table 2.1 lists eight choices offered.

Structure 1 (PID action on error) provides the fastest approach to a new setpoint by virtue of a step from the proportional mode and the kick from the derivative mode in the controller output from the setpoint change as seen in Figure 1.1 in Chapter 1 that provides a clear view of the contribution of each mode. A large step change from the proportional mode is the key to reducing the rise time (time to reach setpoint whether the PV is increasing or decreasing) in near-integrating, true integrating, and runaway loops. For small setpoint changes and low controller gains where the step change in the PID output from the proportional mode is small, the kick instigated by derivative action can help get through significant valve backlash and stick-slip to get the valve moving. The kick appears to be a spike on trend charts with a large time spans. The abrupt change in output is often seen as disruptive by operators when they make setpoint changes. If the burst of flow through the control valve does not affect other users of the process or utility fluid, the kick is more of a psychological than a process concern. The kick can be made smaller by decreasing the γ factor. At any rate, the reduction in rise time (time to reach setpoint) from derivative action on error is marginal for good control valves or higher controller gains and larger setpoint changes.

Table 2.1. Major PID structure choices

(1) PID action on error (β = 1 and γ = 1)
(2) PI action on error, D action on PV (β = 1 and γ = 0)
(3) I action on error, PD action on PV (β = 0 and γ = 0)
(4) PD action on error, no I action (β = 1 and γ = 1)
(5) P action on error, D action on PV, no I action (β = 1 and γ = 0)
(6) ID action on error, no P action (γ = 1)
(7) I action on error, D action on PV, no P action (γ = 0)
(8) 2DOF (β and γ adjustable 0 to 1)

Structure 2 (PI action on error, D action on PV) is the structure most often used. Structure 2 eliminates the kick from derivative action for a setpoint change by setting gamma to zero ($\gamma = 0$). The increase in rise time going from structure 1 to 2 is negligible for the more important loops, such as column and vessel temperature where derivative action is used. The step in the output from the proportional mode on a setpoint change is large because of the high controller gain. Increases in process gain or dead time will increase the overshoot unless the controller gain is decreased accordingly. If the elimination of setpoint overshoot is much more important than rise time, then structure 3 may be best.

Structure 3 (I action on error, PD action on PV) eliminates overshoot but with quite a sacrifice in speed of approach to the setpoint. For bioreactors where the prevention of overshoot for pH and temperature setpoint changes is of paramount importance and increases of cycle time of even an hour in a batch that has a fixed cycle time of 10 days is unimportant, structure 3 is a simple and effective solution. Structure 3 is also used for plug flow and gas unit operations, because the overshoot of the PID output besides the PV is important and the rise time is fast anyway because the primary process time constant is small. When rise time, overshoot, errors from fast load upsets must all be minimized, a setpoint lead-lag or structure 8 enables the use of aggressive tuning and the means for optimizing the setpoint response.

Structure 4 (PD action on error, no I action) is used on processes adversely affected by integral action. The temperature control of severely exothermic polymerization reactors use structure 4 because integrating action in the controller increases the risk of a runaway. If integral action is used, the reset time should be increased by a factor of 10 for these positive feedback processes to be safe. Users may not be aware of this requirement leading to overshoot that can trigger a runaway. The bias for structure 4 is set equal to the normal PD controller output when the PV is at setpoint.

Structure 4 is used for total dissolved solids (TDS) control of boiler drums and vessel level control to eliminate the slow reset cycles from too small of a reset time or too small of PID gain. For drums, the boiler blow down may be discontinuous, making control of the TDS integrating response more difficult with integral action. For reactors, the increase in level from proportional-only control for a decrease in reactant feed flow provides a more constant residence time. However, the setting of the bias in these applications is confusing to the user.

Structure 4 is used on batch processes that respond in one direction (unidirectional). For example, in bringing a batch pH up to a setpoint by the addition of a base reagent where the base is not consumed in a reaction, the batch will only respond in the direction of increasing pH. The pH will overshoot setpoint if integral action is used. If split ranging is added with an acid reagent, there will be some wasted reagent due to cross neutralization of reagents and limit cycling across the split range point from stiction that is greatest near the closed position. For structure 4 and a single reagent, the bias is set for zero reagent addition when the PV is at setpoint.

Structure 5 (P action on error, D action on PV, no I action) is used for the same reasons as structure 4. As with structure 2, the kick from rate action for setpoint changes is eliminated. The use of structure 4 instead of 5 may not offer any advantage because the value of the kick is marginal for these processes since the PID gain is usually high.

Structure 6 (ID action on error, no P action) is used for valve position controllers (VPC) to eliminate the interaction with process controller whose valve position is being optimized. The VPC could be optimizing the coarse adjustment from a large control valve in parallel with a small control valve manipulated as fine adjustment by the process controller or many of the optimization opportunities discussed in Section 12.8. VPC can optimize utility supply pressure or temperature to minimize energy use or optimize feed rate to maximize production rate. The tuning of this VPC is problematic. Tuning rules often cited are: The reset time should be larger than 10 times the product of the gain and reset time of the process controller and 10 times the residence time of the process to eliminate interaction. The reliance on slow integral action makes the VPC unable to prevent the process controller from getting into trouble for large fast disturbances. Feedforward action can be added to help, but a more flexible and easier to tune solution is to use an enhanced PID with external reset feedback and AO setpoint rate limits to provide directional move suppression (discussed in Sections 2.2.2 and 2.2.10).

Structure 7 (I action on error, D action on PV, no P action) is used for the same reasons as structure 6. As with structure 2, the kick from rate action for setpoint changes is eliminated. Since there is no PID output step change to get through backlash and stiction, structure 6 instead of 7 may help by reducing rise time for valves with poor precision.

Structure 8 is used to provide a balance between a fast rise time and minimal overshoot. The performance of this structure for various beta and gamma factors is compared in Section 2.6 Test Results to a setpoint lead-lag where the lead and lag times are varied.

2.2.4 SPLIT RANGE

When a PID controller must manipulate multiple flows, split range control is often used where the PID output is split into several ranges each dedicated to a particular flow by means of a Splitter block. The PID output may manipulate the flow by directly throttling a control valve or by manipulating the setpoint of a flow controller in a cascade control system. The most common split range setup consists of a PID output with a split range point of 50 percent to make a transition from throttling one valve to throttling another valve. If the control valves have the opposite effect, the first valve closes as the output goes from 0 to 50 percent and the second valve opens as the PID output goes from 50 to 100 percent. The first valve is often chosen to be the one that should fail open. Consider three common examples of vessel temperature, pH, and pressure control. The coolant, acid, and vent valve closes as the PID output goes from 0 percent to the split range point. The heating, base, and nitrogen valve opens as the PID output goes from the split range point to 100 percent.

The split range point introduces a severe discontinuity and nonlinearity. Friction and consequently stick-slip is greatest near the closed position, changes in installed valve flow characteristic is most severe, and the changes in process dynamics most dramatic. When switching from steam to coolant, there is also a change in phase creating the possibilities of droplets and bubbles going to and from steam. The result is a tendency to oscillate across the split range point.

Here are some split range guidelines:

1. Use control valves with the least backlash and least stiction particularly near closed position. Since precise throttling valves by design do not provide tight shutoff, use an automatic on-off valve in series coordinated with the throttle valve to provide isolation when a transition across the split range point is sustained.
2. Allocate a sufficient percentage of the system pressure drop as the control valve drop to make the installed flow characteristic more linear.
3. If the pressure drop across the control valve does not appreciably change with flow (e.g., small frictional piping losses resulting in a large valve pressure drop to total system drop

ratio), use a linear trim otherwise use an equal percentage inherent flow characteristic to provide the lowest valve gain near shutoff and thus the lowest magnification of stick-slip and deadband.

4. If a flow loop is manipulated by a PID for split range control, make sure the flow measurement has sufficient rangeability to prevent noise or erratic signal at low flow. If necessary, switch to direct throttling of the valve at low flow.
5. Use a split range point that compensates for the difference in open loop gains seen by the PID when manipulating different streams. For example, if the acid valve has four times the effect of a base valve in changing pH, the split range point should be 80 percent so the acid valve and base valves stroke for a 80 and 20 percent change in PID output, respectively. For open loops gains that change with setpoint, a bumpless recalculation of the split range point would be needed just before the loop is switched to auto with the new setpoint or right after the setpoint change is made with the loop in auto. Valve positions must be held during the recalculation.
6. Use adaptive tuning to correct for changes in process dynamics seen by the PID when manipulating different streams.
7. Eliminate split range control where possible. Consider a VPC instead of split range control when simultaneously throttling multiple streams does not waste energy or raw materials. When throttling a small and large valve on the same stream, a VPC achieves sensitivity besides the rangeability sought by split range control and eliminates the discontinuity from switching from one valve to another. Maximizing the use of low cost streams such as air instead of oxygen for dissolved oxygen control, waste reagents for pH control, and waste fuels for furnace temperature control is best done by a VPC.
8. Use an enhanced PID with a threshold sensitivity limit, and setpoint rate limits on manipulated flow to provide directional move suppression that prevents unnecessary crossings of the split range point.

2.2.5 SIGNAL CHARACTERIZATION

Signal characterization first appeared in pneumatic positioners to help linearize installed characteristics by the use of a cam. To change the characterization required swapping out cams or even custom cutting cams. The linearization was very rough and was often lost when a positioner was replaced in the middle of the night.

The signal characterizer block in today's control systems allows the user to readily set 21 pairs of inputs and outputs to provide a piecewise linear fit to almost any type of nonlinearity. The added visibility and flexibility makes this feature a viable option to compensate for any known gain nonlinearity. Compensation for the installed flow characteristic of a valve is best done by putting a signal characterizer on the PID output.

The most powerful application of signal characterization is in pH control to create a linear reagent demand controller. The characterizer converts the pH signal to a 0 to 100 percent reagent demand PV. The input to the characterizer is the X axis of the titration curve (ratio of reagent to influent flow) and the output of the characterizer is the Y axis (pH). The pairs of inputs and outputs are more closely packed in the normal operating region on the titration curve. A second characterizer can be added on the output of the first characterizer to form a cascade of characterizers that provide a finer resolution of nonlinearities. Both the pH measurement and setpoint must be passed through identical characterizers since the PV scale is now percent reagent demand. An interface may need to be created for the operator to enter the setpoint in pH. The actual pH should be displayed and trended.

The inputs and outputs of signal characterizers must be clearly displayed and operations and maintenance must be educated as to the translation of the signals to ensure that problems with the measurements and valves can be diagnosed and fixed. The slope of the characterizer output must always have the same sign (monotonically increasing or decreasing) because this is the sign of the process action. Local reversals in the sign of the slope will cause a limit cycle across the reversal. The problem is similar to buzzing that occurs from dips in a pump head curve or compressor characteristic curve approaching the zero slope point at low flow.

Knowledge of nonlinearities is never complete, but an imperfect signal characterization is much better than no characterization. The signal characterizer can do a great job of handling the changes in gain over a large range with as much resolution as desired by a cascade of characterizers. The signal characterizer does not eliminate the need for adaptive tuning but frees up the adaptive software to focus on unknown changes in plant dynamics. See Section 9.2.3 for more details on the use of signal characterizers.

2.2.6 FEEDFORWARD

Feedforward control is particularly effective when a disturbance at the process input is large, fast, and precisely measured. While theoretically a feedforward multiplier is more effective for changes in the slope of a plot of the manipulated flow versus the measured disturbance, for many practical reasons such as scaling and nonlinearity problems introduced by a multiplier, a feedforward summer is used for feedforward control especially on vessels and columns. On well mixed volumes, the effect of an increase in process gain with a decrease in ratio is canceled out by the increase in process time constant negating the conventional argument for a feedforward multiplier. Also, the size and effect of an offset is much greater than a slope error at a given operating point. The bias correction has a long history of success in correcting the trajectories in MPC and inferential measurements.

The most common feedforward signal is flow because the changes in feed flows are faster than changes in composition and temperature and the flow is the most common process input. A flow feedforward summer is implemented via an external ratio and bias station so the operator can set the desired ratio and see the actual ratio. On startup, the operator may run on ratio control with the process controller in manual until the equipment reaches operating conditions. The initial flow ratios can be set based on the process flow diagram (PFD). The PFD can be put live online on the operator graphics to show the actual ratios used. The technique also enables a plant wide feedforward system to rapidly increase and decrease production rates by simultaneous increases in feed and manipulated flows to maintain the ratios on the PFD.

The correction provided by the feedforward signal must arrive in the process at the same point and same time but with opposite sign as the effect of the unmeasured disturbance. Feedforward is not beneficial if the disturbance is on the process output because the correction is delayed by the process dynamics. The exception is a process with minimal dead time and a process lag that could be compensated for by the lead setting of a lead-lag applied to the feedforward. See Section 12.2 for more details and test results on the dynamic compensation of feedforward signals.

Here are some feedforward control guidelines:

1. Use a measurement of the largest abrupt load disturbance for feedforward control, usually a feed flow.
2. If the measurement becomes inaccurate or erratic at low loads (e.g., low flows), freeze the feedforward signal or make a smooth transition to a computed feedforward measurement based on actual valve position.
3. Use a feedforward summer where the PID provides a plus or minus correction of sufficient range for the worst case scenario (e.g., plus or minus 50 percent).
4. Display for the operator the most instructive computed and actual feedforward parameter and give the operator the ability to adjust the feedforward gain. For the most common application of flow feedforward (e.g., flow ratio control), the desired and actual ratio of manipulated flow to feed flow are displayed.
5. Enable the operator to run on feedforward control without feedback control (e.g., feedforward active in manual) on startup or when the measurement is not representative of what is really happening in the process or when the measurement is out of service (particularly important for columns and analyzers).
6. Adjust the feedforward gain and dynamic compensation so the correction arrives at the same point at the same time as the load upset with the opposite equal effect. Error on the side of under correction and late correction to avoid creating an inverse response where the initial process response is opposite of the true process response to the load upset, confusing the PID controller.
7. For flow ratio control of feed streams, ratio each stream flow setpoint to a leader flow controller setpoint to achieve desired reaction or blend stoichiometry. To maximize coordination for production rate changes, tune each feed flow PID with the same closed loop time constant or use equal PID setpoint filter times.

2.2.7 DECOUPLING

Decoupling is not used much in PID control unless you consider the flow feedforward signal is a half-decoupler. Chapter 10 discusses how to reduce interaction by the preferential pairing of the PID manipulated and controlled variables and the PID tuning to provide separation of dynamics. The chapter moves on to describe decoupling the interactions between two PID controllers by the simple addition of feedforward signals if the interaction cannot be reduced by pairing or tuning. MPC can do a better job than PID at handling complex dynamics and multiple interactions.

2.2.8 OUTPUT TRACKING AND REMOTE OUTPUT

Output tracking and the remote output mode are used to coordinate the PID output with a sequential operation, associated with an abnormal condition, batch operation, startup, or shutdown. The output is set and held until the condition or operating state has changed. Section 12.3 offers more details on applications than is offered here.

For surge prevention, a drop in suction flow precipitous enough to indicate the start of surge or large enough to put the operating point close to the surge curve warning of an impending surge triggers the rapid opening of the anti-surge valve by the use of output tracking or remote output mode. For an impending surge, the PID output may be rapidly incremented open rather than stepped immediately to a large open position to reduce the upset to users of the compressed gas. A similar strategy is used for the prevention of a RCRA violation from a waste stream pH approaching 2 or 12 pH, the RCRA pH limits. These strategies are called open loop backups because they take preemptive fixed actions when feedback control cannot deal with the abnormal operation. Feedback control is restored as soon as possible by returning the PID to its last mode.

Process engineers like to optimize the timing and magnitude of feeds to a batch operation as the result of process and operating experience as a precursor to fed-batch control. Process engineers may use a similar strategy for starting up continuous unit operations. In some cases, the strategy is necessary until the process has reached an operating point where the PV such as temperature is representative of the state of the process and within the rangeability of limits of the measurements. The use of timely intelligent PID outputs that can be continuously improved is much better than the manual outputs that are by definition off in timing and nonrepeatable even for the same shift of operators. However, at some point the PID needs to be released to automatic, cascade, or remote cascade more so that feedback action can compensate for the inevitable unknowns and disturbances. Some process engineers have difficulty particularly in pharmaceutical operations relinquishing control of feeds to a PID controller. The classic case is where process engineers develop a sophisticated sequencing of air or oxygen or both flows instead of releasing these flows to dissolved oxygen control in a bioreactor.

The time to reach setpoint can be dramatically reduced and overshoot minimized in near-integrating and true integrating processes by preemptively setting the PID output to the limit (usually a high limit) and holding the output at the limit for one dead time after the PV predicted one dead time into the future has reached setpoint. The PID output is filtered and rate limited to prevent an excessive upset to utility and raw material systems. This strategy can reduce the startup time and batch cycle time for attaining desired temperature, pH, and compositions in well mixed liquid vessels.

2.2.9 SETPOINT FILTER, LEAD-LAG, AND RATE LIMITS

A setpoint filter is used to reduce setpoint noise and overshoot. If the filter time is set equal to the reset time setting, the result is a PD on PV and I on error controller or equivalently a 2DOF structure with beta and gamma factors both zero. However, the use of a filter on a secondary (inner or lower) loop setpoint seriously deteriorates the performance of cascade loop by eliminating the immediate response to the demands of the primary (outer or upper) loop. The use of Lambda tuning to keep closed loop time constants of secondary loops coordinated for blending and maintaining stoichiometric ratios in reactions is recommended instead of set point filters to better account for differences in loop dynamics.

A setpoint filter on local setpoint changes made by an operator can eliminate abrupt movement of the controller output. Level and temperature controllers on liquid volumes can have a high PID gain. Step changes in the controller output upset operators and in some cases other upset loops manipulating flows from the same source (e.g., header).

For many plug flow and gas unit operations, such as the reactors, heat exchangers, and furnaces seen in refineries and petrochemical plants, abrupt movement in the PID output upsets other loops especially if there is heat integration where streams are recycled to recover heat. For setpoint changes, a smooth change in the PID output with no overshoot of the final resting value is critical. Rise time is inherently small because of the relatively small primary time constant. For these applications, a “PD on PV and I on error” structure or a setpoint filter is beneficial.

A setpoint rate limit can accomplish the same purpose of eliminating abrupt changes in the PID output with the added benefit of the ramp rate being set differently depending upon whether the setpoint change is up or downscale. Setpoint rate limits can be set in AO or in PID blocks. These setpoint rate limits provide directional move suppression that is extremely useful for preventing unnecessary crossings of the split range point and offering a fast getaway and slow approach to an optimum for VPC and surge control. A positive feedback integral mode with external reset feedback of the PV associated with the setpoint is needed to prevent oscillations from a PID output changing faster than a valve, VSD, or secondary loop can respond. Setpoint rate limits do not prevent overshoot because of the sharp transition from ramp to no ramp when the final setpoint is reached.

A setpoint lead-lag can minimize rise time and overshoot. The recommended settings are a lag time equal to the reset time and a lead time that is a fraction of the lag time (e.g., lead = one-fourth the lag time). A setpoint lead-lag may not be a standard feature in PID block. The 2DOF structure can accomplish the same objectives. Test results show the effect of various setpoint lead-lag and 2DOF settings.

Here are some setpoint response guidelines:

1. For plug flow and gas flow operations and bioreactors, use a “PD on PV and I on error” structure or a setpoint filter time equal to the reset time to provide a smooth setpoint response with no overshoot in the PV or PID output.
2. For mixed liquid chemical unit operations such as reactors and columns, use a setpoint lead-lag or 2DOF structure to optimize the tradeoff between rise time and setpoint overshoot.
3. For protecting against abnormal operation (e.g., compressor surge), preventing environmental violations (e.g., RCRA pH), reducing unnecessary crossings of the split range point causing wasted utilities or raw materials (e.g., heating and cooling or acid and base reagents) and optimizing process (e.g., VPC), use setpoint rate limits on manipulated AO or PID block with an enhanced PID to provide directional move suppression.
4. Do not use a setpoint filter on a secondary (inner or lower) loop PID in a cascade control. For flow ratio control, put the filter on the computed setpoint before feedback correction to reduce flow measurement noise without slowing down cascade control system.

2.2.10 ENHANCED PID FOR WIRELESS AND ANALYZERS

The enhanced PID was developed to handle the discontinuous updates from wireless transmitters and digital valve controllers (digital valve positioners). The PID algorithm is suspended until there is a setpoint, feedforward, or PV change. The PID is executing at the same rate as if the field devices were wired so there is no additional delay in detecting a change. PID algorithms were developed for analyzers that attempt to accomplish a similar purpose by synchronizing the PID execution with the analyzer cycle time. There is one update of the PID output for each analyzer result. However, this *synchronized sampling control* adds a delay to the PID response for setpoint and feedforward changes that is half of the cycle time. The dead time from the enhanced PID is half of the PID execution rate which can be set fast enough to rapidly capture setpoint and feedforward changes. If there is no update due to a communication or an analyzer failure, the enhanced PID simply stops executing and recovers smoothly when the changes in measurement signals resume. *Synchronized sampling control* may continue to execute with a frozen PV.

When a change in setpoint, feedforward, or PV occurs, the PID computes the proportional mode and an integral and derivative mode contribution based on the elapsed time since the last update.

The computation of the integral mode contribution is a first order exponential response of the change in input to the filter in the positive feedback implementation of the integral mode. The exponential response calculation uses the elapsed time.

The computation of the derivative mode also uses the elapsed time between updates so that changes are properly spread out over the actual time interval rather than being considered to having occurred in the PID execution time. For a wireless update every 10 seconds and a PID execution rate of 1 second, the kick in the PID output from the derivative mode is reduced by a factor of ten.

The use of an enhanced PID and a small threshold sensitivity limit to screen out noise can inherently stop limit cycles from backlash and stiction and can eliminate the need for retuning the PID for the additional delay from the use of wireless devices and analyzers. As the delay from discontinuous updates becomes larger than the 63 percent response time of the process of a moderate self-regulating or dead time dominant process, the PID gain can be increased to be as large as the inverse of the maximum open loop steady state gain. See Section 12.4 for more details on the use of the enhanced PID for closed loop control with at-line and off-line analyzers.

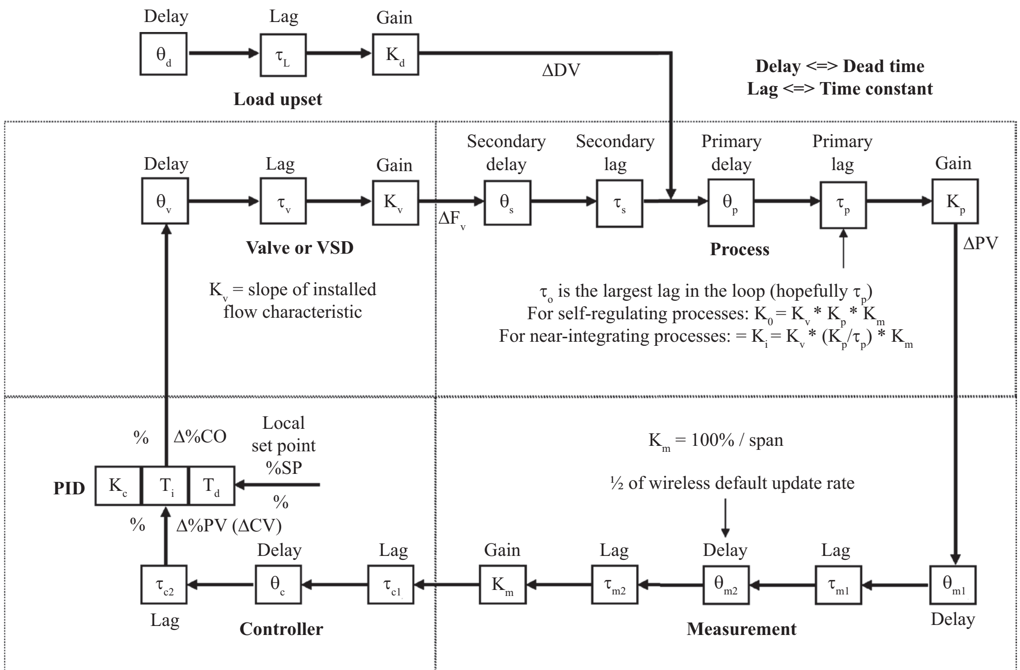
2.3 AUTOMATION SYSTEM DIFFICULTIES

The ideal automation system enables the user to exactly see and manipulate the process. The automation system becomes inconspicuous enabling a focus on the process. Unfortunately the automation system dynamics, rangeability, precision, and accuracy are factors in achieving process objectives. An awareness of the what, when, where, and why of automation difficulties that are limiting loop performance is the first step. The next step addressed in the next section and in Chapter 13 is determining if the peak or integrated error (IE) for load upsets or the rise time or overshoot for setpoint changes is affecting process efficiency or capacity.

The block diagram in Figure 2.4 shows the major sources of delays, lags, and gains in a PID control loop manipulating a control valve or VSD (e.g., variable frequency drive). The block diagram uses the shorter notation of lag for time constant and delay for dead time commonly used in practice.

The literature focuses on process dynamics. The open loop response is stated in terms of a process gain, process time constant, and process dead time misleading users into not realizing the effect of valve and VSD, measurement, and controller scaling and dynamics. The literature also does not adequately treat integrating and runaway process dynamics.

This book uses the terms open loop gain, total loop dead time, primary time constant and secondary time constant. Figure 2.4 plays a key role in understanding these terms and whether a particular block in the diagram is causing a loop performance problem. What surprises is the effect of deadband, resolution, and sensitivity limits not only in creating limit cycles but also in altering the observed open loop gain and total loop dead time.



First Order Approximation: $\theta_o \cong \theta_v + \theta_s + \theta_p + \theta_{m1} + \theta_{m2} + \theta_c + Y^*\tau_v + Y^*\tau_s + Y^*\tau_{m1} + Y^*\tau_{m2} + Y^*\tau_{c1} + Y^*\tau_{c2}$

Figure 2.4. Types and locations of dynamics in the PID control loop.

2.3.1 OPEN LOOP GAIN PROBLEMS

For open loop gain, this book makes the distinction as to open loop self-regulating process gain, open loop integrating process gain, and open loop runaway process gain. The open loop gain regardless of the type of process is the product of the manipulated variable gain, ratio gain, process gain, and measurement gain. The open loop gain is identified as the percent change in the PV divided by the percent change in controller output ($\Delta\%PV/\Delta\%CO$). For integrating processes, the response is a ramp rate giving units of %/sec per % PID output change simply stated as 1/sec. For self-regulating and runaway processes, the open loop gain is dimensionless since the percent units cancel out. The runway process gain is more academic than practical since open loop tests run the risk of a runaway. The process response of runaway processes and self-regulating processes with a large primary time constant is so slow, a near-integrating process classification is useful because the process gain and primary time constant are rolled into an integrating process gain identified from the maximum PV ramp rate early in the process response.

If the manipulated variable is the setpoint of a secondary PID in a cascade control system, the gain is the secondary PID scale span in engineering units divided by 100 percent. Here the manipulated variable gain is linear and constant and no problem. If the manipulated variable is the signal to a control valve or VSD, the manipulated variable gain is a valve or VSD gain that is the slope of the installed flow characteristic with the Y axis in flow units and the X axis in percent signal. The slope is highly non-linear for a valve especially for a low ratio of valve pressure drop to the total system pressure drop. The slope is much more linear for a VSD except at low flows when the ratio of static head to total head approaches one.

Flow loops have a unity process gain unless the engineering units for flow of the valve and measurement are different in the open loop gain calculation. The process gains for other loops are nonlinear varying with operating conditions. The open loop gain for most temperature and composition loops is inversely proportional to the feed flow. The process gain for pH and conductivity control is the slope of the plot versus reagent to feed flow ratio and the plot of conductivity versus concentration, respectively. The process gain for level is inversely proportional to the product of fluid density and cross sectional area for mass flow control.

In distillation column control, temperature is an inferential measurement of composition. The error in the temperature measurement translates to a smaller error in composition as the measurement sensitivity increases. The column tray for the temperature control should be chosen that provides the largest change in temperature for both increases and decreases in the manipulated flow to feed flow ratio (e.g., distillate to feed flow). This tray selection provides the best measurement sensitivity (least composition measurement error) and prevents a process gain that is too small and too nonlinear.

In pH control, the measurement sensitivity is extreme because of the exponential relationship between pH and hydrogen ion concentration. The strategy here is to choose a setpoint that is on the flattest portion of the titration curve to prevent an excessive process gain magnitude and nonlinearity and measurement noise from less than perfect mixing.

The measurement gain is simply 100 percent divided by the measurement scale range. Narrowing the calibration span can increase accuracy since the measurement error is often expressed as a percent of span but a smaller span increases the measurement gain decreasing the maximum allowable PID gain.

The observed open loop gain is less for valve or VSD deadband or resolution. This book will follow the ISA standard for valve response testing and use a resolution limit instead of stick-slip. A resolution limit is effectively a slip equal to the stick. The term also enables the inclusion of VSD resolution resulting from the number of bits in the analog to digital convertor (A/D) input card for the speed command signal.

The observed change in PV is on an average the change in PID output minus half of deadband or resolution limit multiplied by actual open loop gain. The reduction in observed open loop gain increases as the change in output approaches the deadband. However, at times there can be no observed effect. The observed dead time from deadband can match the actual dead time for changes in the same direction. The observed dead time for stick slip can match the actual dead time for changes that are multiples of the stick-slip.

Here are some guidelines to avoid open loop gain problems:

1. Size a control valve or VSD so that the change in slope is less than 4:1 on the installed flow characteristic, the operating range is above the low flow limit (e.g., >10 percent stroke or speed), and below the high flow limit (e.g., <50 percent stroke for rotary and 80 percent stroke for sliding stem valves).
2. Provide a valve pressure drop to system pressure drop ratio that is large enough (e.g., >0.2) and pump static head to total head pressure ratio that is small enough (e.g., <0.2) to prevent an excessively nonlinear valve or VSD.
3. Select a column tray temperature with the greatest sensitivity.
4. Select a pH setpoint on a titration curve with the least sensitivity.
5. Size heat transfer surface areas so the operating points on a plot of temperature versus manipulated to feed flow ratio does not have too large a slope at low ratios and too small

of a slope at high ratios for the operating range of temperatures in all streams. Take into account fouling and the increase in the heat transfer coefficient with flow.

6. Avoid the use of horizontal drums for distillate receivers to prevent the extreme changes in integrating process gain from the change in cross sectional area and a process sensitivity that is less than the measurement sensitivity at half full.

2.3.2 TIME CONSTANT PROBLEMS

The primary time constant is the largest time constant in the loop. Hopefully, the largest time constant is in the process downstream of where load disturbances enter so that these are slowed down giving the PID more time to catch up. If the time constant in the valve or VSD is large, the correction and recognition of changes in the process are slowed down.

Fouled electrodes and thermowells can have an excessive sensor time constant (lag) from a decrease in the mass transfer and heat transfer coefficients, respectively. Pressure and differential pressure transmitters on compressor control can become too slow by simply increasing the transmitter damping setting. The process time constant for surge systems and for liquid pressure control is less than a second.

The second largest time constant in the loop is termed the secondary time constant. For temperature control, the process heat transfer surface lag is the principal source although a thermowell with a coating or large annular clearance can cause a larger secondary time constant. The secondary time constant becomes an equivalent dead time in a first order plus dead time approximation, which is sufficient for moderate self-regulating processes. The effect of a secondary time constant increases as the process self-regulation decreases. For near-integrating, true integrating, and runaway processes, the identification of the secondary time is advisable because of the significant deterioration in loop performance and the advantage of setting the rate time equal to the secondary time constant.

If the measurement lag becomes much larger than the process time constant, an insidious situation develops where the trend charts may look smoother because the measurement is showing an extremely attenuated (filtered) view of the actual process oscillations. The controller gain may even be able to be increased further misleading the user into thinking the measurement lag is beneficial. The key is the loop period increases due to the increase in loop dead time from the conversion of the primary process time constant to dead time.

The size of the primary time constant relative to the total loop dead time is the main source of disagreement on PID tuning and the loop objectives. The tuning rules and need for gradual versus abrupt PID output action is dramatically different.

Most of the literature is dealing with a moderate self-regulating process where the primary time constant and total loop dead time are in the same ball park. These responses are typically found in inline liquid and gas unit operations with negligible back mixing. For these processes, the time to reach setpoint is relatively fast and the degree of filtering of abrupt changes is negligible. Consequently, the PID tuning and structure is chosen to give a gradual response of the PID output with no overshoot of the final resting value.

In contrast, large well mixed liquid volumes have a large primary time constant or an integrating response that dramatically slows down the approach to setpoint and effectively filters out abrupt changes. As a result, the PID tuning and structure is chosen to provide a large and immediate response of the PID output through proportional and derivative action on setpoint changes. Overshoot of the final resting value of the PID output (termed overdrive) is necessary for setpoint changes and load disturbances, particularly for integrating and runaway processes.

2.3.3 DEAD TIME PROBLEMS

The total loop dead time is the sum of all the delays in the loop plus the equivalent dead times from time constants smaller than the primary time constant or an identified secondary time constant. Section 4.2 details how to compute the fraction of small time constants and equal large time constant converted to dead time. Interactive time constants in series should be converted to noninteractive time constants in series as shown in Appendix I. For a large number of equal time constants in series, which is the case for distillation column temperature response where each tray is a time constant, the dead time is about one-eighth of the summation of the interactive time constants.

The transportation delay to a sensor can cause an excessive measurement dead time. The delay is the volume between the connection to the process and the sensor divided by the flow rate in this line. Wireless measurements can introduce excessive dead time in fast loops by a slow default update rate. At-line analyzers and especially off-line analyzers can cause excessive dead time in almost any loop from the sample transport and processing, analysis cycle time, and multiplex time.

The pre-stroke dead time and equivalent dead time from slewing rate from large actuators can be excessive particularly for pressure and surge control. Here, the solution is to put a volume booster on each output of the positioner with the bypass valve opened just enough to prevent high frequency oscillations (e.g., >0.5 cps). The oscillations are the result of the positioner output looking into a booster volume that is much smaller than the actuator volume. Diaphragm and piston actuators have one and two positioner outputs, respectively. If a booster is used instead of a positioner on butterfly valves, a positive feedback situation is created with the booster that can lead to unstable operation including the unsafe slamming closed of a valve that as per the PID output should be open.

There is an additional dead time observed for closed loop control that is the valve or VSD deadband or stick-slip divided by the rate of change of the PID output, or the measurement threshold sensitivity and resolution limits divided by the rate of change of the PV. The additional dead time from the valve is not seen for step changes in the PID output that are larger than the deadband or stick-slip, which is the case for most open loop tests and setpoint changes with proportional action on error instead of PV. The additional dead time from the measurement is not observed for a step change in the actual PV that is larger than the measurement, the threshold sensitivity, and resolution limits. A step changes in the PID output change multiplied by the open loop gain that is not appreciably filtered by a process time constant can cause a rapid change in the PV. While this is not a step change, the change can be fast enough to exceed the threshold sensitivity or resolution limit in a time interval short enough to reduce the observed dead time. We will see in test results that the tuning compensation is more in terms of increasing the reset time than decreasing the PID gain.

2.3.4 LIMIT CYCLE PROBLEMS

A limit cycle occurs when there are one or more integrators in the loop and a resolution limit, threshold sensitivity limit, or a stick-slip. Deadband requires two or more integrators to create a limit cycle. The integrators can be in the process (e.g., level, batch temperature, batch concentration …), or in the automation system (e.g., integral action in positioner, secondary PID, or primary PID). Limit cycles exist in every loop but are not recognizable if the amplitude is less than the exception reporting or compression of the data historian or the amplitude of the measurement noise. Limit cycles are generally not a problem until the PV amplitude is larger than 0.1 percent. The amplitude of the limit cycle is the valve or VFD deadband divided by the PID gain or the valve stick-slip or measurement resolution or threshold sensitivity limit multiplied by the open loop gain. The amplitude of the limit cycle from dead band can be reduced by increasing the PID gain. The improvement can be impressive because the PID gain is probably much lower than possible from the expectation that reducing the PID gain would reduce cycling.

2.3.5 NOISE PROBLEMS

The sensor design, location, and operating conditions can result in excessive measurement noise. Differential head meters and vortex meters are especially sensitive to velocity profile and hence the upstream and downstream piping configuration. Most flow meters get noisy as you approach the low flow limit.

Bubbles cause measurement noise in pH and dissolved oxygen control. pH electrodes are also especially vulnerable to noise from electromagnetic interference and concentration fluctuations. Droplets cause temperature sensor noise, a particular problem in desuperheaters.

2.3.6 ACCURACY AND PRECISION PROBLEMS

Bias errors are generally more problematic than span errors because they are more frequent and cause an offset between the actual setpoint and the desired setpoint. A span error shows up as a change in open loop gain. An offset is not introduced by a span error unless there is a large setpoint change. The change in open loop gain is usually insignificant compared to the other process nonlinearities.

Operations may gradually compensate for the offset through experience by changing the setpoint, but the correction is always late and inaccurate. A control loop driving the setpoint can correct for an offset faster than it is changing. Thus, offsets in positioners and secondary loops are not important except in terms of data analytics and inferential measurements and material and energy balances, the basis for high fidelity simulations and PFDs.

Poor precision is more of a problem than poor accuracy for process control and analysis in that the dynamic error cannot be corrected by operations and only partially corrected by upper loops. Precision is seen as the standard deviation in the measurement for a constant PV. The source of poor precision is sensor repeatability, sensor interferences, and threshold sensitivity and resolution limits. Precision can be corrected to the degree an upper loop can correct the average of oscillations if the oscillations are consistent and sufficiently attenuated by liquid volumes. Precision problems are too transient and confusing for operations and process engineers to make manual corrections.

2.4 PROCESS OBJECTIVES

Increasingly we need to be able to focus our efforts with the biggest economic impact as companies and processes are pushed to be more profitable. The recognition of the automation professional beyond just installing systems depends upon the ability to find, obtain, and document ways to make a plant more productive. Here, we show how process objectives can be translated to loop performance and the effective use of PID features.

2.4.1 MAXIMIZE TURNDOWN

As inventories are minimized and market demands fluctuate, the need increases for the ability to be able to change production rates rapidly and smoothly. A plantwide feedforward control system is used to keep important PVs such as temperature and composition close to setpoint as feed flow rates are simultaneously changed in the complete train of unit operations.

How low the plant can be turned down depends upon the rangeability of the flow meters, control valves, and VSDs. Flow feedforward depends upon the ability of these automation components to respond precisely and smoothly over the complete range of plant operation.

Most flow meters have a low fluid velocity limit. The actual rangeability matches the stated rangeability by the supplier when the user's maximum flow matches the maximum flow of the meter size. This coincidence is rare especially since line size meters are commonly specified. Thus, the actual rangeability achieved is considerably less than advertised.

The signal for many flow meters gets erratic and noisy below the low flow limit. For vortex meters, the stated rangeability is about 12:1. The consequences of going below the vortex meter low limit are quite severe in that the signal drops out to zero unless some special software is used to freeze the signal. For differential head meters, the differential pressure transmitter input for a 4:1 turndown drops to one-sixteenth of the maximum due to the square root relationship. The rangeability can be extended from 4:1 to 8:1 or more by the use of a second low range transmitter, and a stronger signal to noise ratio from a smaller beta ratio and a more uniform velocity profile. Magnetic and Coriolis flow meters can have a rangeability of 50:1 and 200:1, respectively, offering the greatest plant turndown.

The stated rangeability for control valves and VSDs has little practical value. The actual rangeability depends upon the installed flow characteristic having a reasonable slope (not too steep or too flat) and a sensitive smooth response near the closed position and minimum speed. For valves, an excellent rangeability (e.g., 200:1) is attained for an equal percentage inherent characteristic with a sufficient valve drop to system pressure drop ratio (e.g., >0.2) and a small deadband from backlash (e.g., <0.1 percent) and a small stick-slip from stiction (e.g., <0.05 percent). For VSDs, a comparable rangeability is possible for a low static head to total head ratio (e.g., <0.1) and minimum dead band setting (e.g., <0.2 percent) and an excellent signal resolution (e.g., <0.1 percent). The standard speed input cards for some VSD have a resolution of only 0.35 percent.

2.4.2 MAXIMIZE SAFETY AND ENVIRONMENTAL PROTECTION

The peak error for load disturbances is the loop performance parameter most relevant to preventing an excursion from causing a compressor surge, activating a relief device, triggering a safety instrumentation system trip, or exceeding a pH limit. In some cases, a preemptive action may need to be taken by the means of an open loop backup described in Section 2.2.8 via use of the output tracking or remote output modes. The use of directional move suppression and an enhanced PID noted in Section 12.6 can provide the fast getaway for abnormal operation when VPC is used for optimization. The goal of safety and environmental protection overrides the temporary loss in process efficiency from ensuring the reaction is never too small.

2.4.3 MINIMIZE PRODUCT VARIABILITY

For a more intelligent metric, only errors of the key PV(s) associated with product quality that exceeds a margin and thus indicative of product being off-spec should be integrated. For liquid volumes with some degree of mixing, the loop performance parameter IE is an indicator of the accumulated amount of off-spec product since short term fluctuations are averaged out by the back mixing in these liquid volumes. For gas unit operations and plug flow systems (e.g., plug flow reactors and static mixers), extruders, and sheet lines, the integrated absolute error (IAE) is a better indicator of total product off spec since short term plus and minus fluctuations are not averaged out and immediately appear in the product.

The metric most often seen in the literature depicts the benefit as shown in Figure 13.1a from the reduced variability in a statistical distribution of the key PV. Often an accompanying benefit can be realized by eliminating the offset shown in Figure 13.1b inadvertently imposed by operations due to an overreaction to variability and a lack of process understanding. The area of the distribution that is beyond the limit is a measure of the amount product off-spec. The distribution should be computed as close to the final product quality stream so that the averaging effect of intervening volumes is included.

A power spectrum analyzer can help find the source of the variability. The frequency with the greatest power in the key PV is searched for in power spectrums throughout the plant. The furthest upstream spectrum showing the same frequency is generally the culprit. The loop initiating the problem can be confirmed by putting the suspect PID in manual. Often the problem is traced back to a level controller oscillating due to violation of Equation 1.5e from too low a PID gain or too small a reset time.

If the amplitude is variable, the problem is most likely a tuning problem. The period of a variable amplitude oscillation relative to the size of the loop dead time is indicative of the cause. For most loops, the ultimate period is about four times the total loop dead time. If the oscillation period is about three to four dead times, four to five dead times, and six to eight dead times, the problem is most likely overly aggressive derivative, proportional, and integral tuning settings, respectively. If the oscillation is much greater than eight dead times, the problem could be due to violation of Equation 1.5e on a near-integrating, true integrating, or runaway process. For this last case, decreasing the PID gain makes the problem worse.

If the amplitude is fixed, the oscillation is most likely a limit cycle caused by excessive dead band, stick-slip, or insufficient resolution or threshold sensitivity. Other possibilities include a frequent periodic disturbance, such as the discharge from a centrifuge, and resonance. Section 8.4 goes into much more detail on the sources of oscillations.

2.4.4 MAXIMIZE PROCESS EFFICIENCY AND CAPACITY

Less off-spec can be taken as increase in capacity by an increase in production rate or as an increase in process efficiency by a reduction in raw material feed rate for the same production rate. If the off-spec is recycled or treated for disposal, there is also a lower processing cost (e.g., energy use) for a decrease in off-spec.

Operating closer to a spec limit can often save on raw material and energy use. In Section 13.3, a few choice examples are given with significant benefits.

The simple configuration addition of an enhanced PID as a VPC can improve process efficiency by lowering energy costs and increase process capacity. Section 12.6 offers many examples with guidance as to the setup of the VPC.

An increase in capacity can occur from greater on-stream time. The improvement can be significant for the processes that are difficult to startup. Fewer trips can be achieved by a smaller peak error as noted in Section 2.4.2 and by better measurement reliability. Middle signal selection of three measurements will inherently eliminate a single measurement failure of any type causing a false trip or appreciably disrupting a PID.

For batch operations, an increase in process capacity can be obtained from a shorter batch cycle time or a larger concentration at the batch end point. Chapter 13 will cover the opportunities for Fed-Batch operation where key feeds are being manipulated by PID control rather than simply sequenced on and off.

2.5 STEP-BY-STEP SOLUTIONS

We start with the steps for the selection and installation of the field measurements and control valves or VSDs to enable the control system to meet plant objectives (Steps 1 to 6). The measurement is the essential window into the process and manipulating a flow is the essential way to affect a process.

We continue with the steps for designing control strategies. Simple rules of thumb will be offered for setting up cascade, composition, flow, level, pH, pressure, and temperature control systems (Steps 7 to 12). Examples of common unit operations will help provide understanding of the steps involved. The objective is to set the stage for tuning and configuration of the PID to do the best job possible.

We conclude with a unified approach to tuning that enables a common and simplified method for setting PID tuning parameters (Steps 13 to 25). Key features can be used to eliminate the need for retuning to deal with different dynamics and objectives.

1. *Select the measurement method that has the best repeatability and threshold sensitivity and the least noise and drift.* For temperature, resistance temperature detectors have order of magnitude better specs in all of these performance criteria than thermocouples if vibration or temperature is not too high. Coriolis meters have a similar order of magnitude performance advantage over other flow meters on streams of variable composition and can be the right choice in smaller lines where lifecycle cost is considered. For larger line sizes where differential head meters are the economic choice, use *meter runs* (orifice, flow tube, or venturi meter as an integral part of a straight run of pipe with known minimal roughness with flow lab calibration test results from meter supplier) and dual transmitters to reduce noise and extend rangeability. Radar level measurements have an order of magnitude performance advantage and are the right choice for accurate inventory or tight mass balance control despite changes in process composition if the dielectric constant is not too low. Use pH electrodes with spherical or hemispherical bulbs, glass to

meet worst case process conditions (e.g., high temperature), and replaceable liquid reference junctions. Use smart transmitters and minimize the use of sample lines, impulse lines and capillary systems.

2. *Select the measurement location that provides the most representative indication of the PV of interest, the greatest sensitivity, fastest response, and the least noise.* For temperature sensors, this means a liquid velocity of at least 0.5 fps or a gas velocity of 50 fps. For pH electrodes, the goal is a liquid velocity of at least 5 fps and no bubbles. For column temperature control, the best tray is generally the one with the largest temperature change for both an increase and decrease in reflux to feed flow ratio. The sensor is best located in the liquid on the tray rather than on the vapor space above the tray for a faster sensor response.
3. *Select and design the control valve or VSD to provide the best installed flow characteristic and the least deadband and best resolution.* Understand that the valve or VSD (e.g., variable frequency drive) gain is proportional to the slope of the installed flow characteristic. Recognize based on information in Chapter 7 that the installed flow characteristic gets more nonlinear as the size of the valve drop decreases and pump static head increases compared to the total system pressure drop. Use software to compute and plot the installed flow characteristic. For a control valve, a diaphragm actuator, digital positioner, equal percentage characteristic and sliding stem valve with a valve drop that is at least 25 percent of the system pressure drop offers the best performance. For a variable frequency drive, a pulse width modulated inverter, an internal speed control system, *inverter duty* totally enclosed fan cooled (TEFC) or totally enclosed water cooled (TEWC) motors depending on process temperature, XPLE jacketed cables, separation from instrumentation cables, and isolation transformers as needed to prevent EMI, high resolution signal input cards, deadband less than 0.1 percent in drive setup, and a static head that is less than 10 percent of the system pressure drop.
4. *Tune the valve's digital positioner or drive's speed control to provide a fast nonoscillatory response for minimum offset without integral action.* If integral action must be used, set an integral deadband to suppress a limit cycle. There are some isolated examples where the cycling will help a temperature loop where the process volume averaging of the fast oscillations helps achieve a greater precision. The advantage of tighter temperature control may outweigh the premature wearing out of valve packing.
5. *Ask yourself the questions in the checklists* given in the appendices of the ISA book *101 Tips for a Successful Automation Career to* make sure you are covering all the bases in the design and the implementation of the field instrumentation and final control elements. If you want a greater understanding, consider buying or borrowing the ISA book *Essentials of Modern Measurements and Final Elements in the Process Industry*.
6. *Add a flow measurement to every important process and utility stream to enable a secondary flow lower loop for cascade control.* A secondary flow loop isolates pressure disturbances, and nonlinearities of the installed characteristic of control valve and VSDs from the control of a higher PV such as composition and temperature. The flow measurement enables flow feedforward control and the possibility of changing continuous unit operation production rates by moving plant flows in unison per the PFD. The flow measurements also enable closing material and energy balances leading to better process knowledge eliminating uncertainties from pressure flow relationships. For info on the many benefits of flow measurements see *InTech* articles "Advances in Flow and Level Measurement Enhance Process

Knowledge, Control" (McMillan 2011) and "Feedforward Control Enables Flexible, Sustainable Manufacturing" (McMillan 2012). A word of caution: Control valves and VSD normally have a greater rangeability than a differential head or vortex meter. When this occurs, a calculated flow based on the installed characteristic should be substituted for the measurement flow before the signal becomes too noisy or in the case of the vortex meter drops out. A bias to the pressure drop used in the flow calculation should be automatically set to provide a smooth transition from measured to computed flow.

7. *Realize that the input determining composition, pH and temperature loop response is a flow ratio.* Plot these PVs versus the ratio of the manipulated flow to the feed flow. For vessel and particularly reactor temperature control, set up a secondary coil or jacket temperature lower loop to isolate utility system nonlinearities and disturbances from the primary temperature loop. Design the utility system to keep the coil and jacket flow constant to prevent a low heat transfer coefficient, fouling, and high dead time and high process gain at low flow. Use direct blending of utility streams and coil or jacket streams rather than heat exchangers to make the secondary loop faster. Avoid changes in phase in the coil and jacket. Use steam injectors to create hot water without bubbles eliminating the need to make a transition between cooling water and steam.
8. *Recognize that the input determining level and pressure loop response is the difference in flow entering and exiting the volume.* Pressure control in general needs to be tight since pressure changes translate to flow changes in control valves and VSD and determine equilibrium relationships with temperature and affect the driving force for mass transfer (e.g., vaporization of volatile components and absorption of dissolved gases such as oxygen and carbon dioxide). Pressure also affects the reaction rate of gas reactants. Level control may or may not need to be tight. Tight level control may be necessary for tight residence time (volume/flow) in continuous reactors and crystallizers where the level setpoint changes with production rate to keep the ratio of volume to total feed flow constant. Tight level control is also important for recycle of distillate to column as reflux from a distillate receiver for self-regulation of reflux in the column and recycle of recovered reactants to a reactor self-regulation of concentrations in recycle system. On the other hand, many plant oscillations can be traced to a level control loop where tight level control was unnecessary and the level loop is excessively changing feeds to downstream equipment. The classic examples are surge tank levels and distillate level controllers, manipulating distillate flow to downstream users rather than reflux to a column. Simply decreasing the level controller gain without increasing the reset time may create slow oscillations from the violation of Equation 1.5e.
9. *Choose the controlled variable that is affected the most by the PV really trying to be controlled, typically a composition.* For distillation columns, this rule corresponds to a tray temperature that shows the largest change to both an increase and decrease in the pertinent flow ratio. For acid or base concentration control, this rule translates to the use of pH for an acid or base concentration less than 0.1 normality and conductivity for a higher concentration.
10. *Choose the manipulated variable that causes the greatest effect on the controlled variable.* For column temperature control, a change in the reflux/feed ratio most often has a larger effect than a change in the steam/feed ratio. For columns separating low concentrations of light (lower boiling point) components leading to low distillate flow, the distillate receiver manipulates reflux flow rather than distillate flow. For columns separating

low concentrations of heavy (higher boiling point components), the sump level controller manipulates steam flow rather than bottom flow. Unfortunately this may introduce inverse response in the level control but if the bottom flow is too low to provide enough muscle to deal with range of operating conditions, the choice to manipulate steam is right to always assure the ability to control level.

11. *Choose the manipulated variable that is least affected by disturbances given sufficient effect on the controlled variable.* The classic example is the pressure and flow control of a pipeline with two control valves. The more important control variable is flow since this is what ultimately affects the downstream equipment. The flow loop should manipulate the control valve with the largest pressure drop so that pressure upsets have the least effect. This corresponds to the smallest valve given it can handle the maximum flow, which is counter intuitive. Stroking either valve will provide the flow range so here the choice is determined by which valve minimizes the effect of common disturbances, which for flow is pressure. The pressure controller manipulates the larger valve. The pressure controller should be tuned for fast aggressive action as noted in step 8 while the flow controller should be tuned for slow smooth action. Pressure regulators did this naturally by providing high gain and fast action. Put pressure measurement and control in the control room for visibility, adjustability, and diagnostics. Make sure the control valve pre-stroke dead time and slewing rate and PID execution rate are fast enough and mostly proportional action with a high PID gain is used for tight pressure control.
12. *Ask yourself the questions in the checklists* given in the appendices of the ISA book *101 Tips for a Successful Automation Career* to make sure you covering all the bases for temperature and pH control. For much more extensive information checkout the ISA books *Advanced pH Measurement and Control* and *Advanced Temperature Measurement and Control.*
13. *Set the output limits to keep the manipulated setpoints in the desired operating range.* For VSDs, set the process PID low output limit so the speed cannot cause the discharge head to approach the static head in order to prevent excessive sensitivity to pressure and to prevent reverse flow. In general, set the anti-reset windup limit to match the output limit. If the output scale is engineering units, the output limits and anti-reset windup limit must be based on the output scale range and units.
14. *Choose the best structure for your application.* Generally the best choice is a structure with PI on error and D on PV although a structure of PID on error may be useful for small setpoint changes to help get through valve deadband. For a unidirectional response (e.g., batch heating or neutralization), use a structure such as “P on error and D on PV” and “PD on error” so that there is no integral action. For a highly exothermic reaction, you might want this structure to help prevent a runaway from integral action.
15. *Minimize the signal filter time.* A noise filter should be just large enough to keep the controller output fluctuations from exceeding the resolution limit or deadband of the final control element so that the valve or VSD does not respond to noise.
16. *For near-integrating, true integrating, and runway processes, use the Lambda integrating process tuning rules.* To maximize the transfer of variability from the PV to the manipulated variable, set the Lambda (arrest time) equal to the maximum dead time and use the largest integrating process gain for all possible operating conditions in the tuning. Set the rate time equal to the secondary time constant. Use a low limit of four times the dead time for the reset time for PI control and half the dead time for the rate time for PID control. Consult

Table 2.2 on how to increase Lambda to deal with nonlinearities, non-idealities, and other objectives. For example, to maximize the absorption of variability (e.g., surge tank level) compute the minimum arrest time from Equations 1.21d, 1.22f, or 1.22m for all possible operating conditions. If you decrease the PID gain, proportionally increase the PID reset time to prevent slow rolling oscillations. For runaway processes, the PID gain must not be less than the inverse of the open loop gain. When changing from the Series Form to the ISA Standard Form in newer DCS, convert the tuning settings based on the differences in Form. In the ISA Standard Form the rate time should not be greater than one-fourth the reset time. Be sure to realize and convert any difference in Form and tuning setting units between different PID vintages and suppliers. You must take into account the units and inverse relationship between gain and PB and reset time in seconds and minutes per repeat. Not addressing these differences can result in settings off by orders of magnitude.

17. *For moderate self-regulating and dead time dominant processes, use the Lambda self-regulating tuning rules.* To maximize the transfer of variability from the PV to the manipulated variable set the Lambda (closed loop time constant) equal to the maximum dead time and use the largest process gain and smallest time constant for all possible operating conditions in the tuning. If the dynamics are exactly known and there are no extenuating circumstances, the peak error for fast disturbances can be minimized by a Lambda approaching half the dead time. Consult Table 2.2 on how to increase Lambda to deal with nonlinearities, nonidealities, and other objectives. Schedule tuning settings preferably with an adaptive tuner with the specified Lambda to deal with changes.
18. *Turn on external reset feedback.* Make sure the external reset feedback signal is correctly propagated back to the PID (e.g., BKCAL signal) especially to account for split range, signal characterizer, or signal selector blocks on the PID output.
19. *For final control elements that are slow or that have deadband or resolution limit, use a fast readback of the valve position or variable frequency drive speed as the external reset feedback.* The external reset feedback of the actual PV response will prevent a burst of oscillations from the PID output changing faster than the final control element can respond. The external reset will also prevent the limit cycles from deadband, threshold sensitivity, and resolution limit.
20. *For cascade control, use the PV of the lower loop as the external reset feedback.* The external reset feedback of the actual PV response will prevent a burst of oscillations from the PID output changing faster than the secondary loop can respond.
21. *For setpoint rate limits, use the PV of the AO block or lower loop as the external reset feedback to prevent the need to retune the PID.* Add setpoint rate limits to minimize the interaction between loops and to provide directional move suppression to enable a fast getaway for abnormal conditions and a slow approach to optimum particularly useful for pH, surge, and VPC. For VPC, use an enhanced PID developed for wireless with a threshold sensitivity limit to ignore insignificant changes in the valve position to be optimized.
22. *For wireless devices and at-line and off-line analyzers, use the enhanced PID to prevent the need to retune the PID for self-regulating processes.* The performance advantage of the PID increases as the update time increases relative to the process response time. If the update time is much larger than the 63 percent process response time (dead time plus primary time constant), the PID gain can be as large as the inverse of the maximum open loop self-regulating process gain.

Table 2.2. Lambda tuning solutions and suggestions to achieve different objectives

Objective	Solution	Assumptions	Suggestion
Maximize absorption of process variability	Minimize changes in manipulated flow	Integrating process	λ computed per Eq. 1.21d, 1.22f, or 1.22m
Minimize resonance	Minimize amplification	Load oscillation	$\lambda > 5*\theta_o$
Prevent oscillation from equal % valve	Increase gain margin to 9 or more	Valve gain change less than 6:1	$\lambda > 5*\theta_o$
Prevent oscillation from inverse response	Increase gain margin to 6 or more	Inverse response lead time < θ_o	$\lambda > 3*\theta_o$
Prevent oscillation from nonlinear lag	Increase gain margin to 9 or more	Self-reg. lag change less 10:1	$\lambda > 5*\theta_o$
Coordinate multiple flow loops	Make all loops respond equally slow as slowest	Sensor and valve sizing consistent	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_{lower}$
Minimize peak error in runaway process	Maximize PID gain	Cascade control Rx to jacket	$\lambda = 0.5*\theta_o$ $T_d > 0.5*\theta_o$
Minimize peak error to prevent relief	Maximize PID gain	Linear vent valve and constant SP	$\lambda = \theta_o$
Minimize amplitude from valve dead band	Maximize PID gain	Linear or adapted control loop	$\lambda = \theta_o$
Prevent violation of cascade rule	Make Lambda ratio > 5 preferably by making lower loop faster	Linear or adapted control loops	$\lambda_{upper} > 5*\lambda_{lower}$
Reduce interaction between loops	Make Lambda ratio > 5 preferably by making faster loop faster	Linear or adapted control loops	$\lambda_{slower} > 5*\lambda_{faster}$

For a given Lambda and open loop dynamics, gain margin can be estimated as the ratio of the ultimate gain to PID gain used. Chapter 4 provides the equations for the ultimate gain for self-regulating, integrating, and runaway processes. A Lambda equal to the dead time provides a gain margin of three that means the loop does not become unstable until the open loop gain increases by 300 percent. Additionally, the closed loop response remains non-oscillatory for increases in the open loop gain less than 50 percent.

The implied dead time can also be computed from Lambda and the actual process dead time via equations developed in Chapter 5. The allowable extra dead time that can be introduced by a controller, filter, wireless measurement, or analyzer without appreciably increasing the peak or IE is the implied dead minus the current dead time.

Table 2.2 provides guidelines on how to set Lambda to address different application difficulties and objectives other than minimization of the IE from large fast unmeasured disturbances at the process input. In each case, the solution and suggestion are based on an accurate knowledge and use of the maximum total loop dead time and integrating process tuning rules for near and true integrating, and runaway processes. In order to deal with dead time uncertainty the associated Lambda must be increased in proportion to the possible dead time positive error (e.g., Lambda increased by 50 percent for a potential 50 percent increase in dead time). For the runaway processes, the reactor to jacket temperature cascade loop dynamics are assumed to be fixed by minimization of excursions via aggressive PID action so the heat of reaction is relatively constant.

For the coordination of multiple flow loops for maintaining stoichiometry for changes in feed demand for composition or production rate control, the Lambda of each of the flow loops ratioed is set equal to the Lambda of the slowest flow loop (e.g., flow loop with the greatest dead time). The leader flow setpoint is manipulated by the upper control loop. The follower flow loop setpoints are ratioed to the leader flow loop setpoint. If feed flow disturbances (e.g., pressure disturbances) are more of a problem than feed composition disturbances, each flow loop is individually tuned for the tightest flow control, and a single filter is added to the leader flow setpoint to provide simultaneous setpoint changes that the slowest flow loop can achieve.

2.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in setpoint (ΔSP) or load flow at the process input (ΔF_L). Table 2.3 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The effect on a near-integrating process setpoint response for the 2DOF structure gamma and beta settings is studied in Figures 2.5a through 2.5d and for setpoint lead and lag settings in Figures 2.6a through 2.6d. The effect on the load response of various processes for valve deadband from backlash is studied in Figures 2.7a through 2.7i and for valve resolution from stick-slip in Figures 2.8a through 2.8i.

The goal in the setpoint tests was to see the relative effectiveness of the 2DOF structure and the setpoint lead-lag in minimizing rise time (time for PV to first cross setpoint), overshoot, and settling time (time till PV lines out at setpoint) for various settings. The starting point was a near-integrating process with aggressive tuning that minimizes the peak and IE for unmeasured load disturbances. These tuning settings provide a minimum rise time but cause a large overshoot and settling time for a setpoint change. If integrating tuning rules with Lambda equal to dead time was used, the overshoot would be smaller but the rise time would be slightly longer.

Figure 2.5a shows the effect on the PV response to a setpoint change for different beta (setpoint factor for proportional mode) in a 2DOF structure for a gamma (setpoint factor derivative mode) that is zero (derivative on PV instead of error). For a beta equal to zero we have an “I on error and PD on PV” structure that has the least PV overshoot but slowest rise time. For a beta equal to one we have a PI on error and D on PV structure that has the most overshoot and fastest rise time the most commonly used structure. The overshoot for a beta of zero and 0.25 are about the same. The settling time is about the same for the range of beta values (zero to one).

Table 2.3. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	Effect
2.5a, b, c, d	Near-integrating	1 dimensionless	10	100	ΔSP = 5%	2DOF structure
2.6a, b, c, d	Near-integrating	1 dimensionless	10	100	ΔSP = 5%	Setpoint lead-lag
2.7a, b, c	Moderate self-reg.	1 dimensionless	20	20	ΔF_L = 10%	Deadband 4%
2.7d, e, f	Near-integrating	1 dimensionless	10	100	ΔF_L = 22%	Deadband 4%
2.7g, h, i	True integrating	0.01 1/sec	10	—	ΔF_L = 22%	Deadband 4%
2.8a, b, c	Moderate self-reg.	1 dimensionless	20	20	ΔF_L = 10%	Resolution 4%
2.8d, e, f	Near-integrating	1 dimensionless	10	100	ΔF_L = 22%	Resolution 4%
2.8g, h, i	True integrating	0.01 1/sec	10	—	ΔF_L = 22%	Resolution 4%

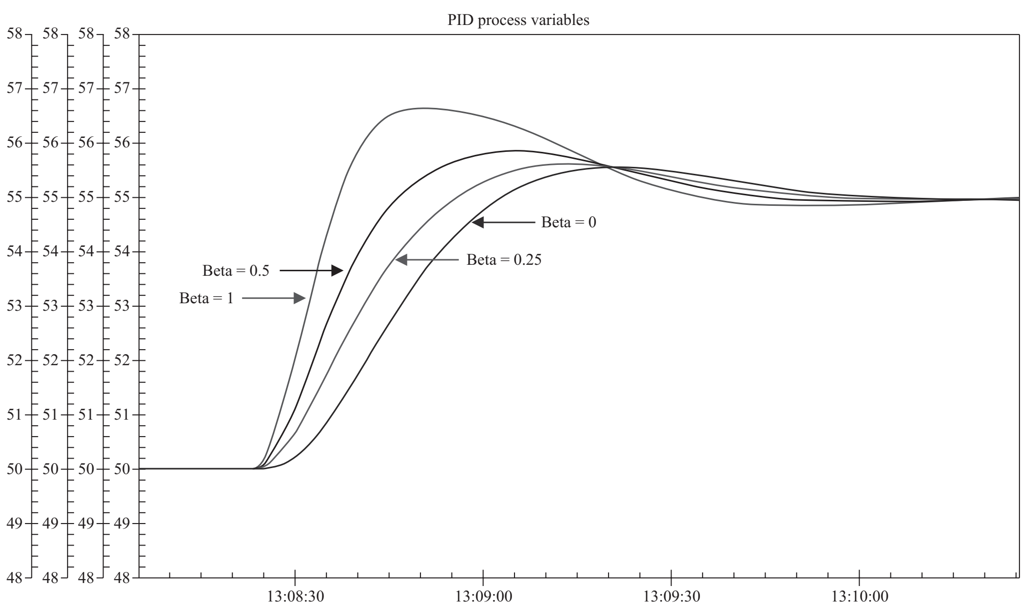


Figure 2.5a. Effect of *beta* setpoint factor with aggressive tuning on PV for 2DOF PID structure (*gamma* = 0).

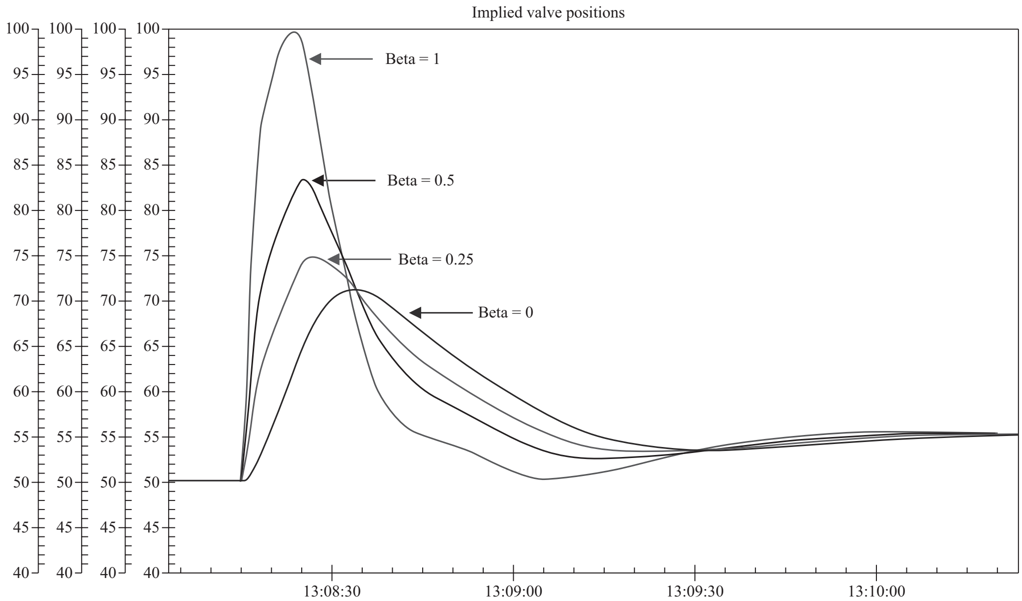


Figure 2.5b. Effect of *beta* setpoint factor with aggressive tuning on valve for 2DOF PID structure (*gamma* = 0).

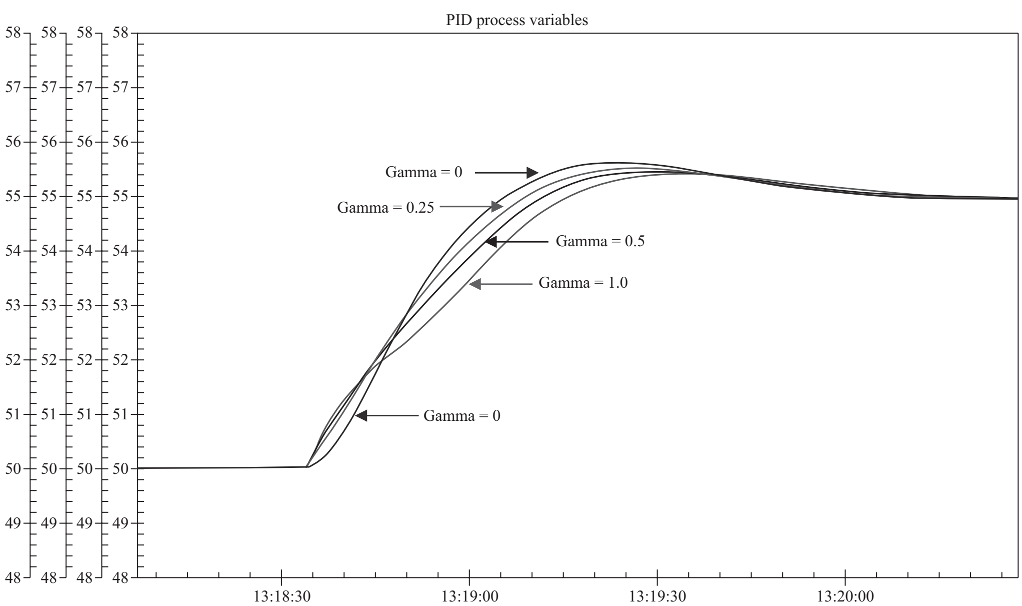


Figure 2.5c. Effect of *gamma* setpoint factor with aggressive tuning on PV for 2DOF PID structure (*Beta* = 0.25).

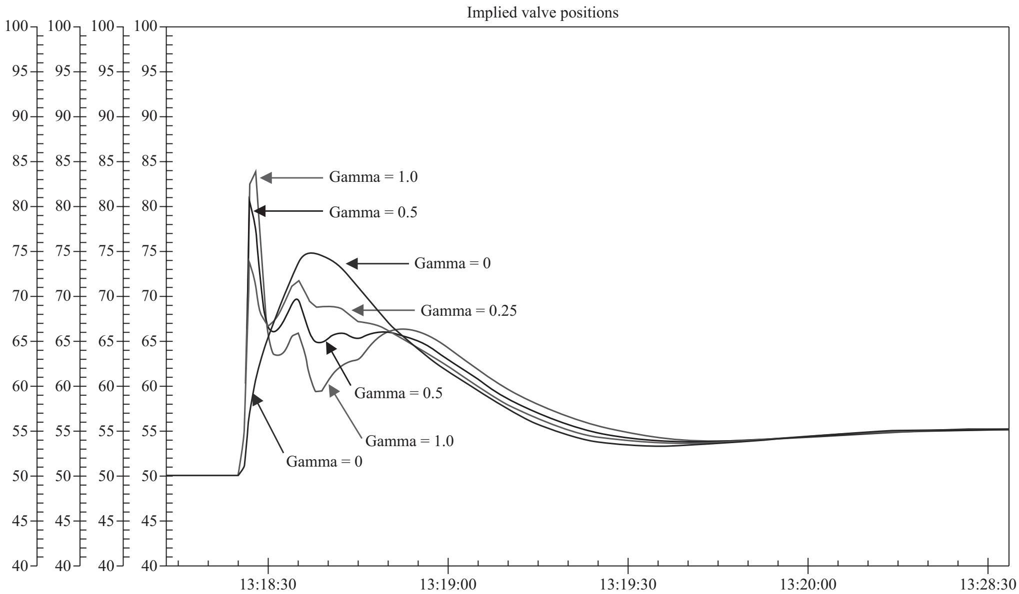


Figure 2.5d. Effect of *gamma* setpoint factor with aggressive tuning on valve for 2DOF PID structure (*beta* = 0.25).

Consequently, a beta of 0.25 gives the best compromise between rise time and overshoot for these tuning settings. Since less aggressive settings are normally used in industry, a beta of 0.5 is a better rule of thumb.

Figure 2.5b reveals the effect on the PID output response to a setpoint change for the Figure 2.5a test conditions. As expected, the PID output shows an immediate change almost to the output limit for the setpoint change. A slight setpoint filter of 2 seconds results in the output change not quite being a step change. The filter was used to smooth out setpoint noise from unsynchronized execution of modules used in the tests. The size of the peak decreases as the beta decreases. For a beta equal to zero, the increase in the PID output is solely from integral action. While there is some hesitation in the approach of the PID output to the final resting value, there is no real oscillation in the PID output.

Figure 2.5c illustrates the effect of changing gamma in a 2DOF structure for a beta equal to 0.25. The change in overshoot and rise time is not dramatic for the different gamma values. The settling time is about the same again for all cases. Since the difference in rise time is rather small when the gamma is increased from 0.25 to 0.5, a gamma of 0.25 provides the best compromise in minimizing overshoot and rise time considering the rate time is more conservatively set in industrial applications.

Figure 2.5d reveals the effect on the PID output response to a setpoint change for the Figure 2.5c test conditions. The plots show an oscillation in the PID output in the approach to the final resting value for a gamma greater than zero. This oscillation faster than the ultimate period is disconcerting to operations but for a near-integrating unit operation may pose no practical problem unless the oscillation upsets other users of a utility or raw material header or valve wear becomes a consideration from frequent setpoint changes.

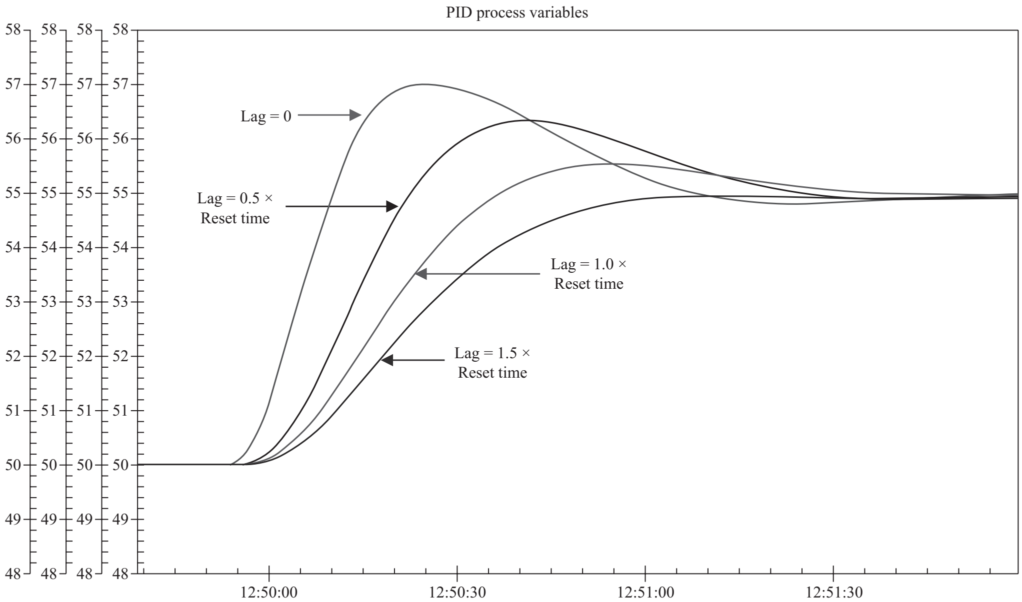


Figure 2.6a. Effect of setpoint *lag* with aggressive tuning on PV for PI on error and D on PV structure (*lead* = 0).

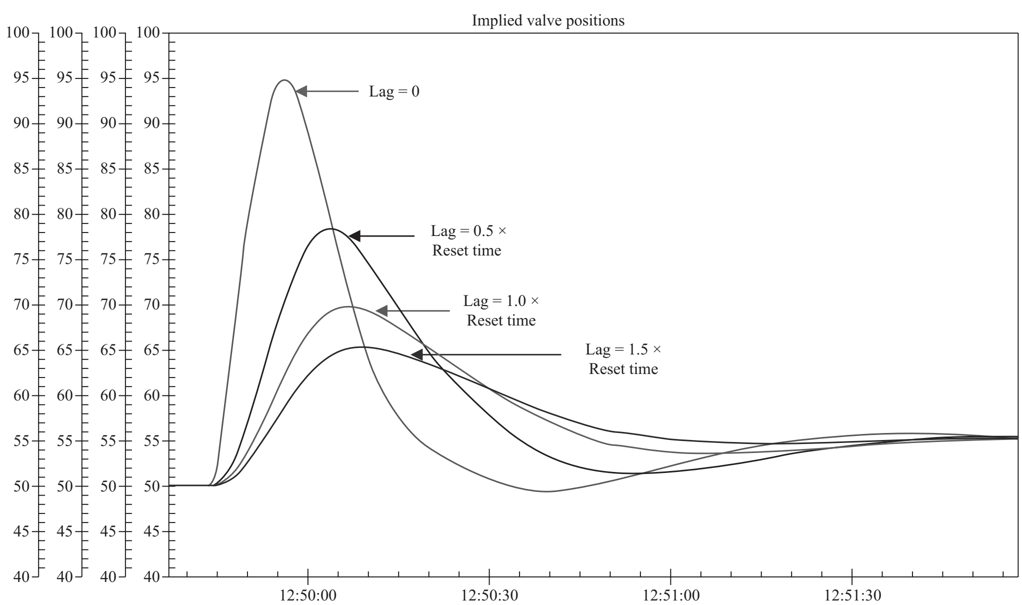


Figure 2.6b. Effect of setpoint *lag* with aggressive tuning on valve for PI on error and D on PV structure (*lead* = 0).

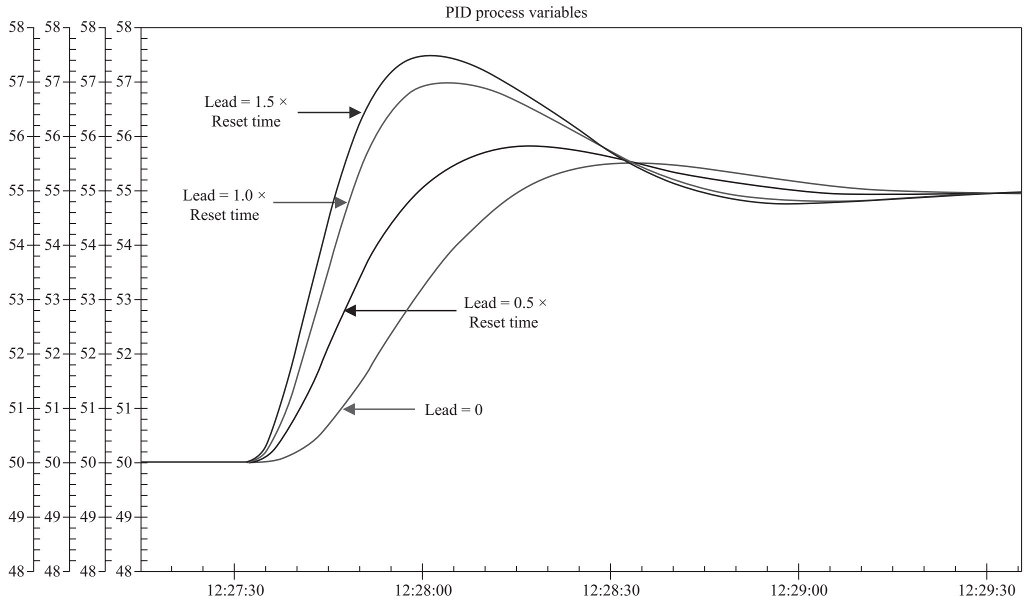


Figure 2.6c. Effect of setpoint *lead* with aggressive tuning on PV for PI on error and D on PV structure (*lag* = *reset time*).

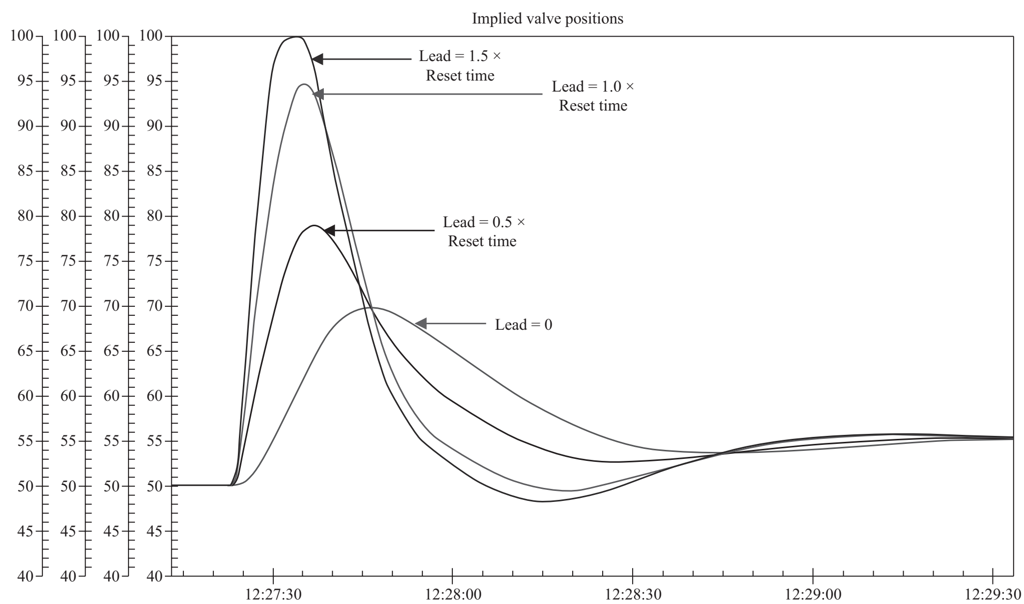


Figure 2.6d. Effect of setpoint *lead* with aggressive tuning on valve for PI on error and D on PV structure (*lag* = *reset time*).

Figure 2.6a shows the effect on the PV response to a setpoint change for different setpoint lag times in a “PI on Error and D on PV” structure for a lead time equal to zero. Due to the aggressive tuning, a setpoint filter of 1.5 times the reset time is needed to eliminate overshoot. This filter setting also provides the minimum settling time. Since industry normally uses less aggressive tuning, a setpoint filter time equal to the reset time is a better rule of thumb.

Figure 2.6b reveals the effect on the PID output response to a setpoint change for the Figure 2.6a test conditions. The hump is least as expected for the slowest setpoint lag time. For a zero lag time, a damped oscillation slower than the ultimate period is seen due to aggressive tuning.

Figure 2.6c shows the effect on the PV response to a setpoint change for different setpoint lead times in a PI on Error and D on PV structure for a lag time equal to the reset time. The lead time has quite a large effect on rise time and overshoot. The settling time is about the same for all cases. The decrease in rise time between a lead time equal to 0.5 times reset time and zero is quite large. The best compromise may be attained by a lead time equal to 0.25 times the reset time (case not shown).

Figure 2.6b reveals the effect on the PID output response to a setpoint change for the Figure 2.6c test conditions. As the lead time increases, the peak in the PID output dramatically increases.

For the aggressive tuning in the test cases a setpoint lead-lag with the lag time equal to 1.5 times the reset time and a lead time equal to 0.25 times the reset time is best. These settings offer slightly better PV metrics and a smoother change in PID output than 2DOF.

Figures 2.7a through 2.7i show the load response in moderate self-regulating, near-integrating, and true integrating processes of the PV, PID output, and actual valve position for a

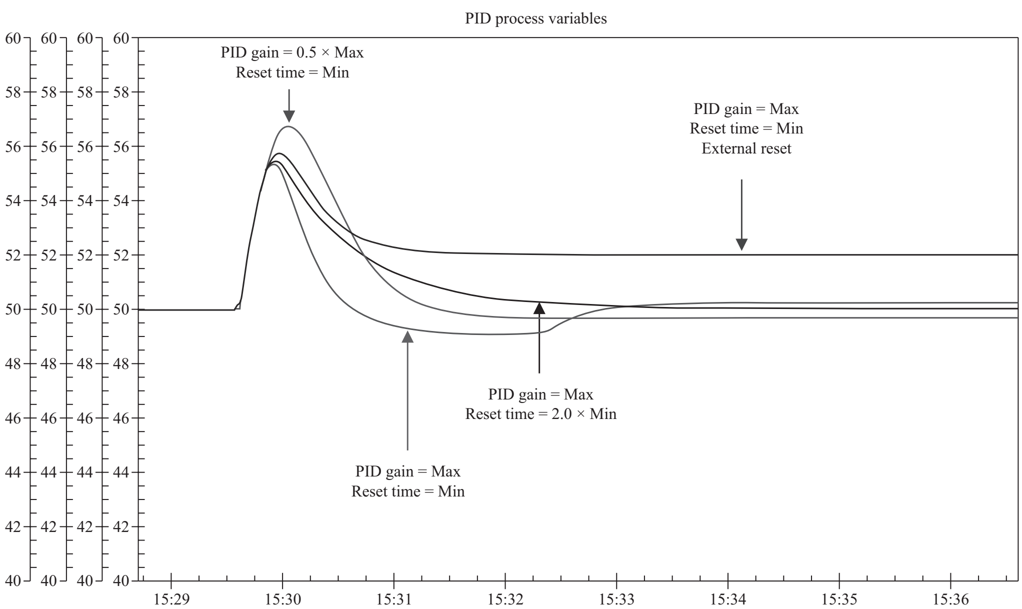


Figure 2.7a. Effect of 4% valve *deadband* on PID *PV* for 10% load upset in *moderate self-regulating* process.

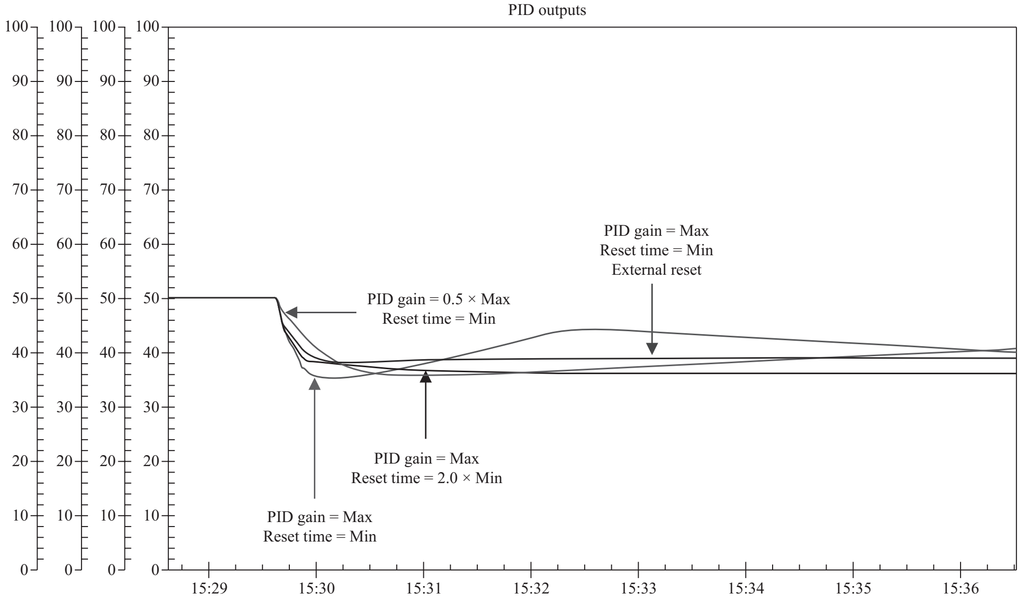


Figure 2.7b. Effect of 4% valve *deadband* on PID *output* for 10% load upset in *moderate self-regulating* process.

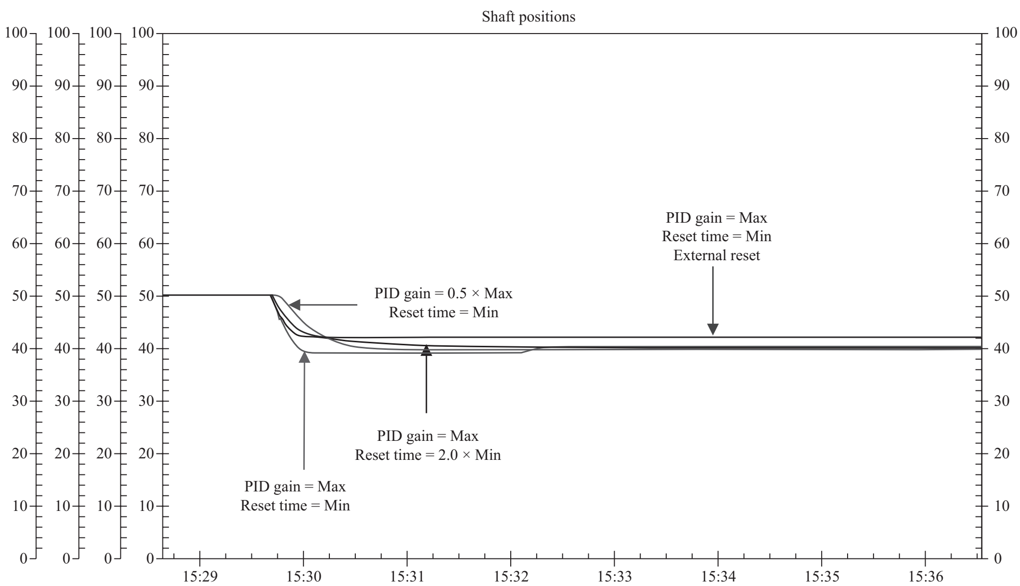


Figure 2.7c. Effect of 4% valve *deadband* on valve *stroke* for 10% load upset in *moderate self-regulating* process.

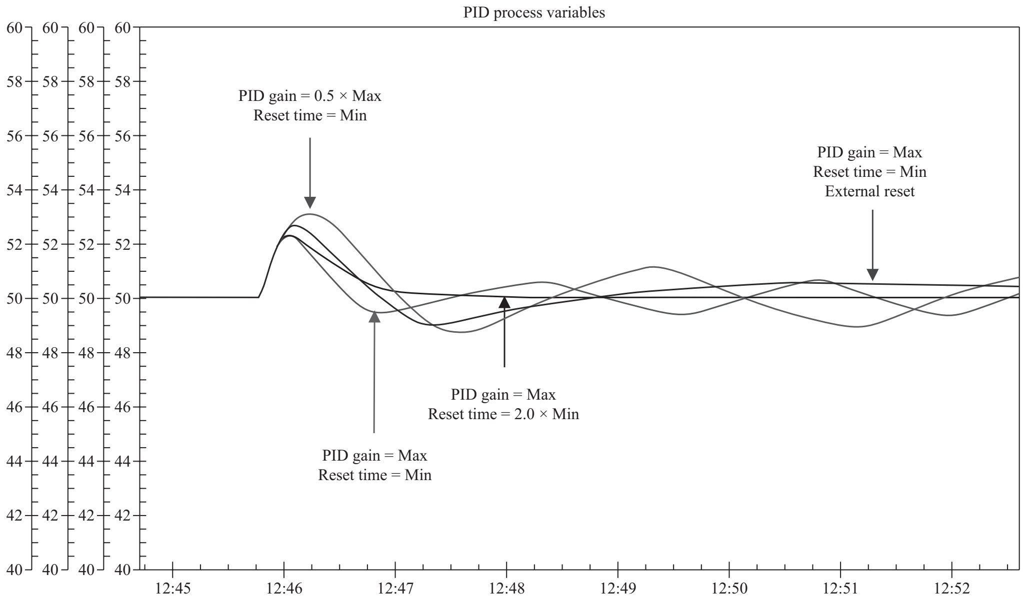


Figure 2.7d. Effect of 4% valve *deadband* on PID *PV for* 22% load upset in *near-integrating* process.

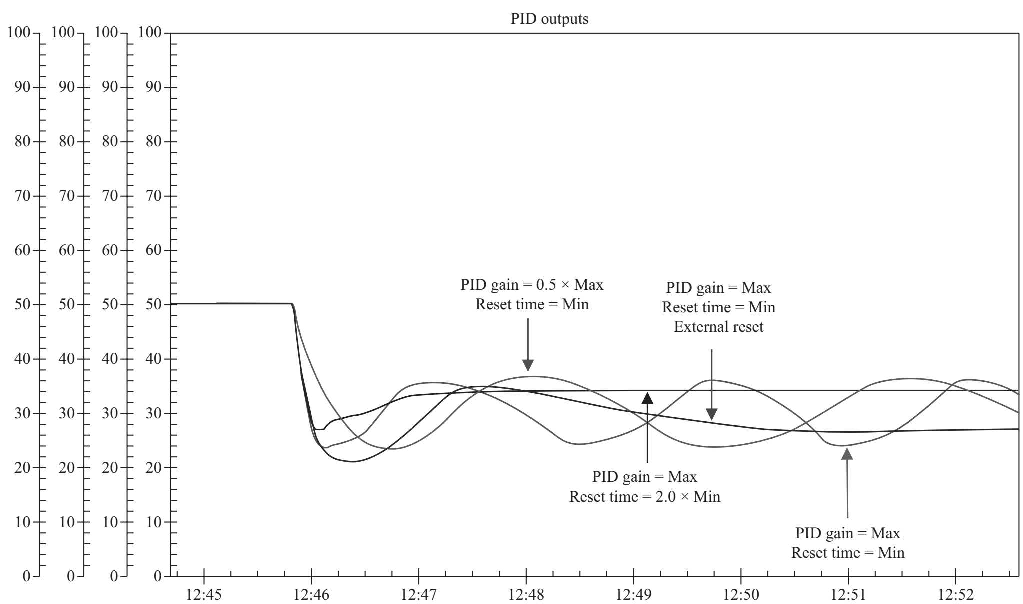


Figure 2.7e. Effect of 4% valve *deadband* on PID *output* for 22% load upset in *near-integrating* process.

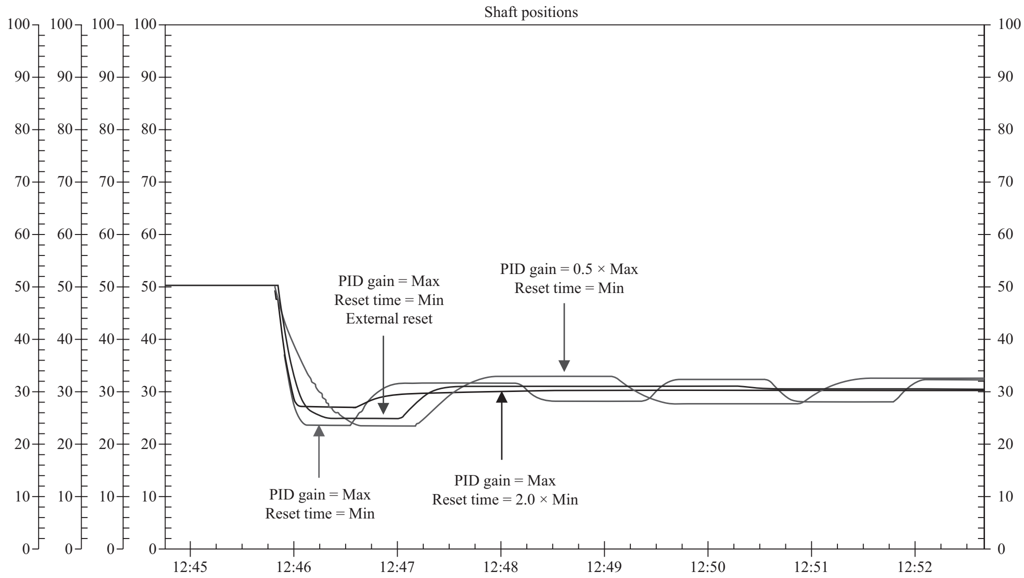


Figure 2.7f. Effect of 4% valve *deadband* on valve *stroke* for 22% load upset in *near-integrating* process.

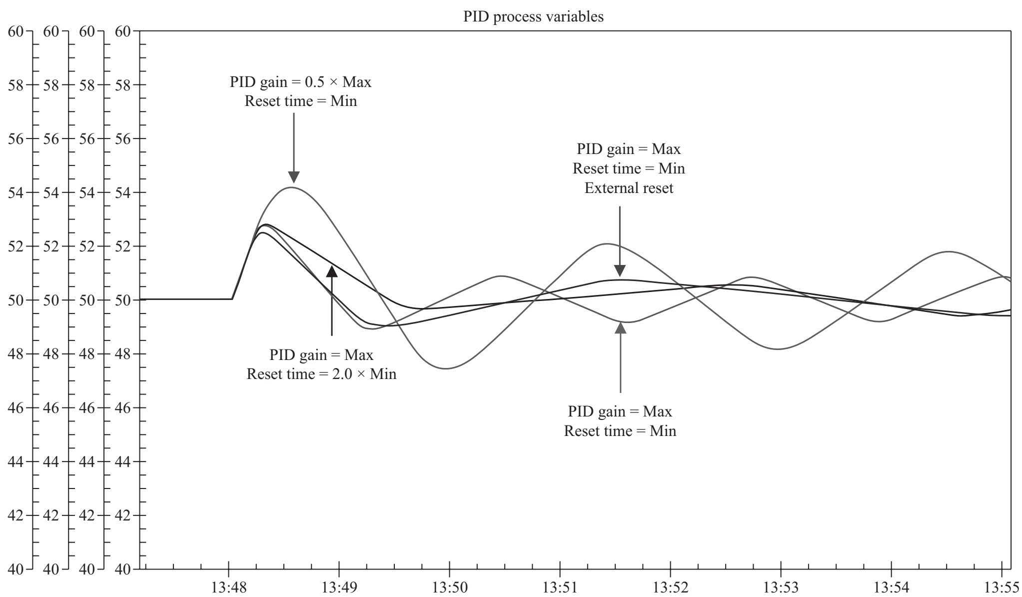


Figure 2.7g. Effect of 4% valve *deadband* on PID *PV* for 22% load upset in *true integrating* process.

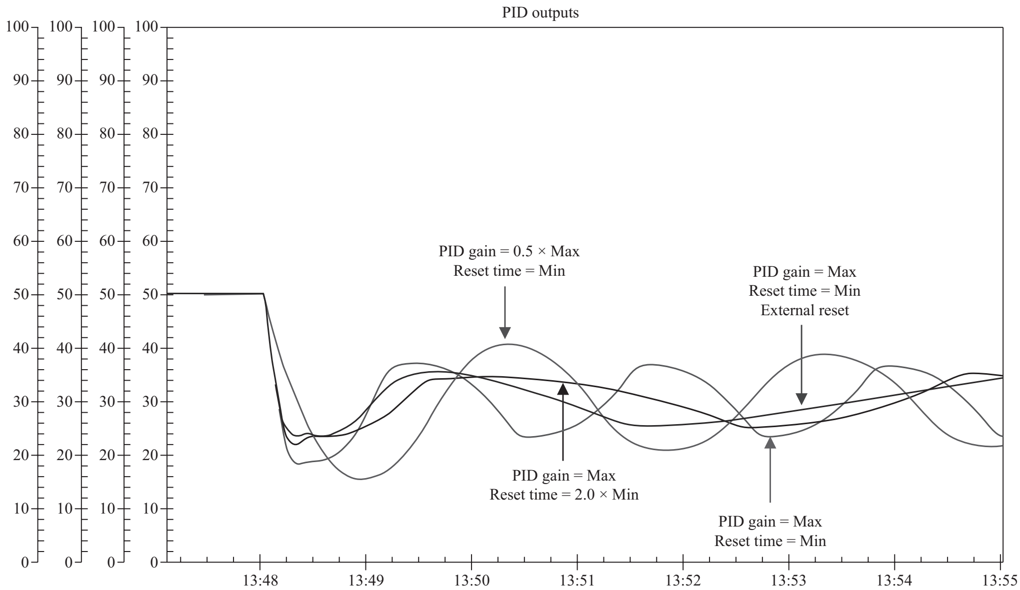


Figure 2.7h. Effect of 4% valve *deadband* on PID *output* for 22% load upset in *true integrating* process.

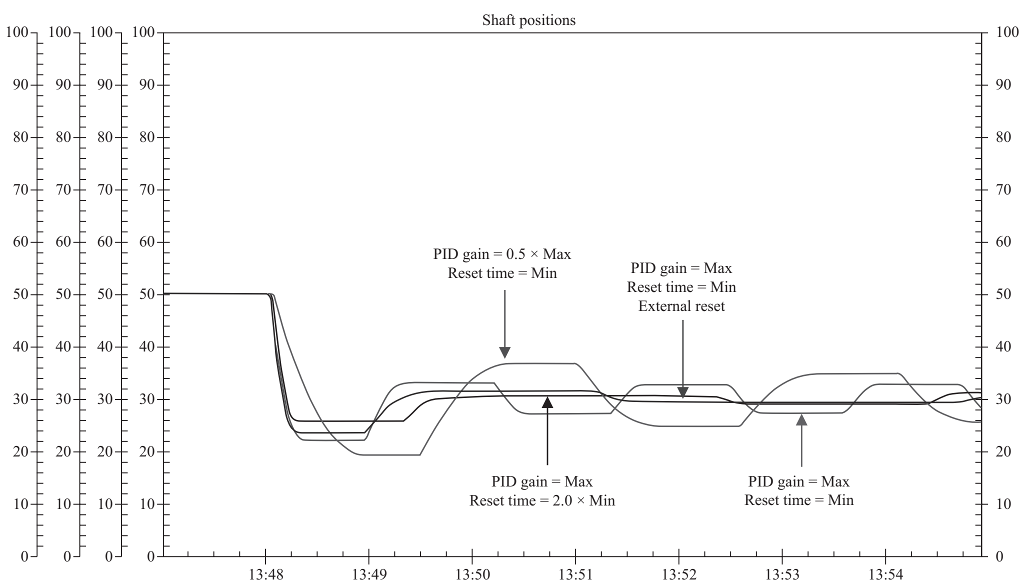


Figure 2.7i. Effect of 4% valve *deadband* on valve *stroke* for 22% load upset in *true integrating* process.

deadband of 4 percent for different gain and reset settings. The figures include a plot of external reset feedback enabled for the original aggressive tuning settings.

For the moderate self-regulating process, the original aggressive tuning settings create a damped oscillation with a very slow final return to setpoint. The incredible slowness on the last portion of the response is best seen in Figure 2.7b as the PID output slowly ramps through the deadband from the integral mode acting on a small error. A halving of the PID gain eliminates the oscillation in the PV but not in the PID output and significantly increases the peak error. A doubling of the PID reset seems to be the best tuning option since the effect on peak error is negligible and both the PV and PID output responses are smooth. Turning on external reset feedback freezes the output but the PV offset is large.

For the near-integrating process, a sawtooth develops in the PV response, an oscillation in the PID output, and a square wave in the actual valve position unless the reset time is doubled or external reset is turned on. The benefit of external reset is greater than for the moderate self-regulating process.

For the true integrating process, a true limit cycle develops. The size of the peak error, sawtooth in the PV response, oscillation in the PID output, and square wave in the actual valve position are larger than for the near-integrating process. The problem is exasperated by halving of the PID gain. A doubling of the reset time does not stop the sawtooth, oscillation, or square wave of the limit cycle but makes the period longer. External reset cannot stop the limit cycle because momentarily stopping PID integral action does not stop the ramping of the PV from the integrating process action.

Figures 2.8a through 2.8i show the load response in a moderate self-regulating, near-integrating, and true integrating processes of the PV, PID output, and actual valve position

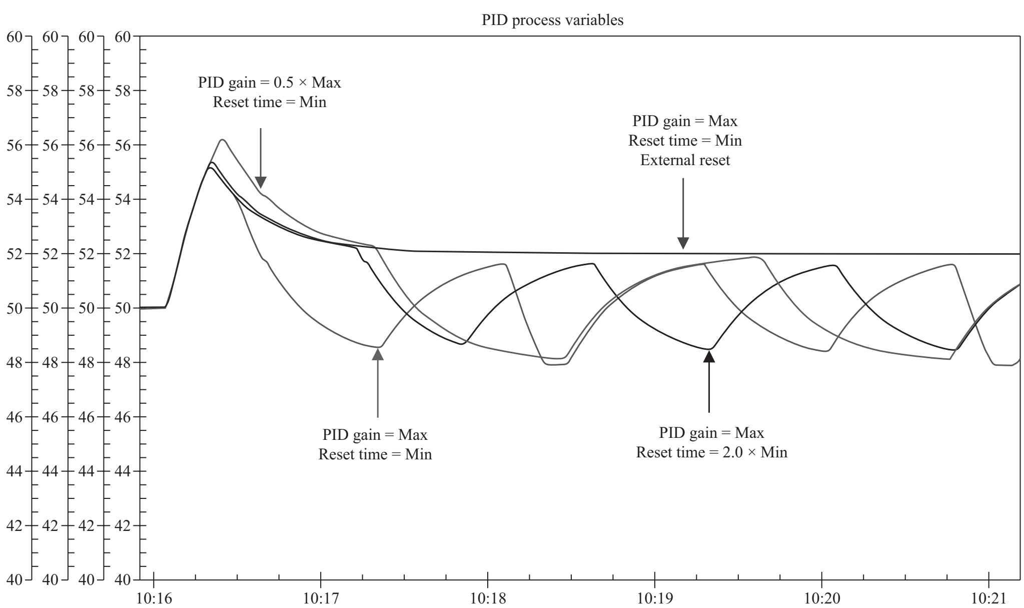


Figure 2.8a. Effect of 4% valve *resolution* on PID *PV* for 10% load upset in *moderate self-regulating* process.

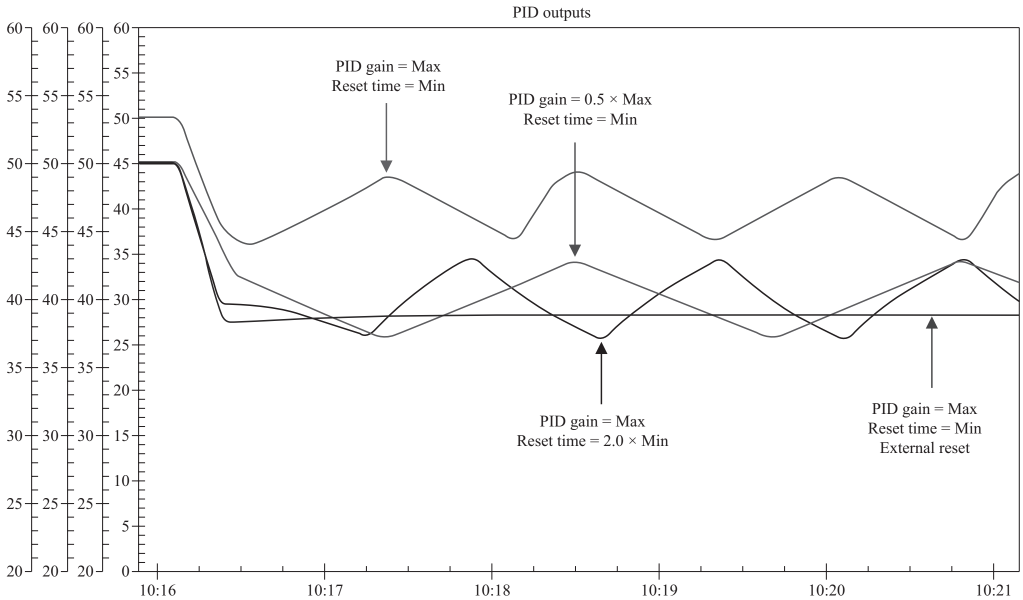


Figure 2.8b. Effect of 4% valve *resolution* on PID *output* for 10% load upset in *moderate self-regulating* process.

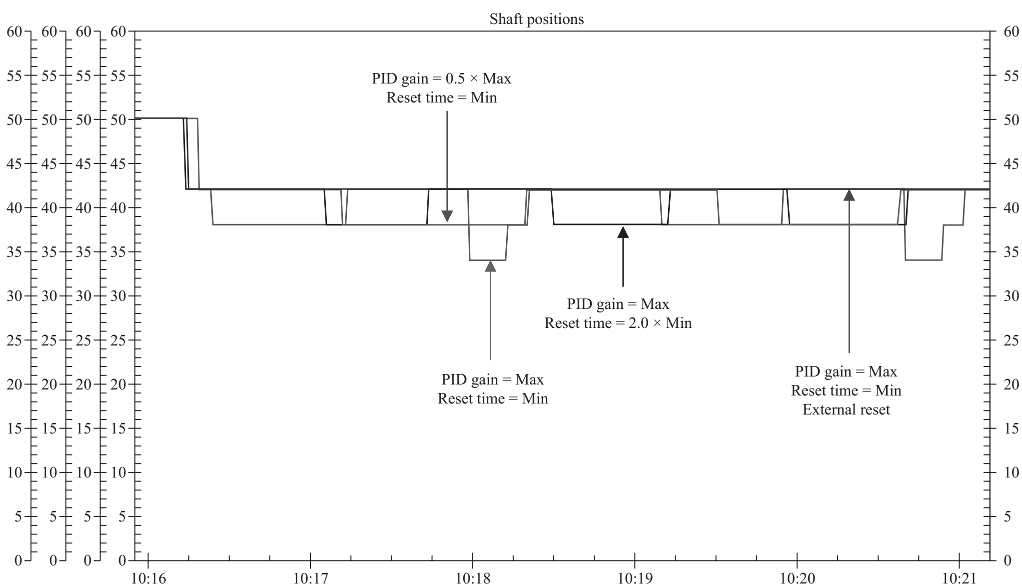


Figure 2.8c. Effect of 4% valve *resolution* on valve *stroke* for 10% load upset in *moderate self-regulating* process.

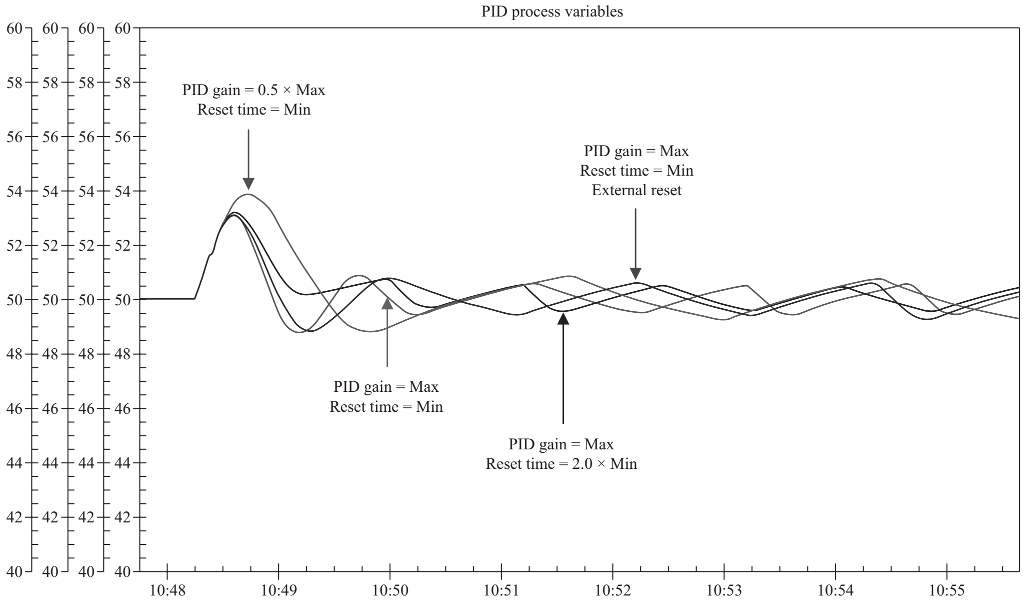


Figure 2.8d. Effect of 4% valve *resolution* on PID *PV* for 22% load upset in *near-integrating* process.

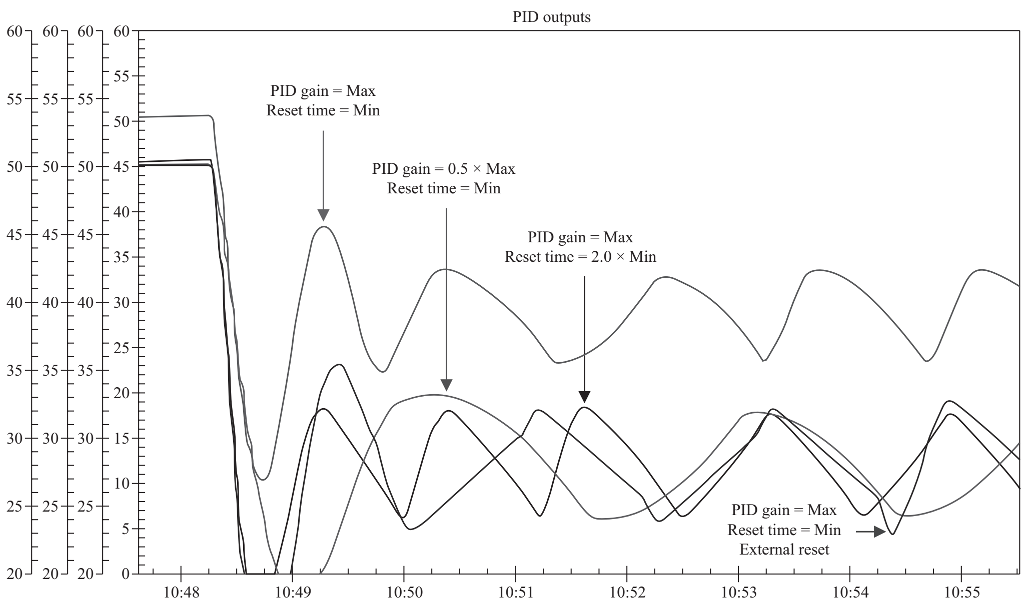


Figure 2.8e. Effect of 4% valve *resolution* on PID *output* for 22% load upset in *near-integrating* process.

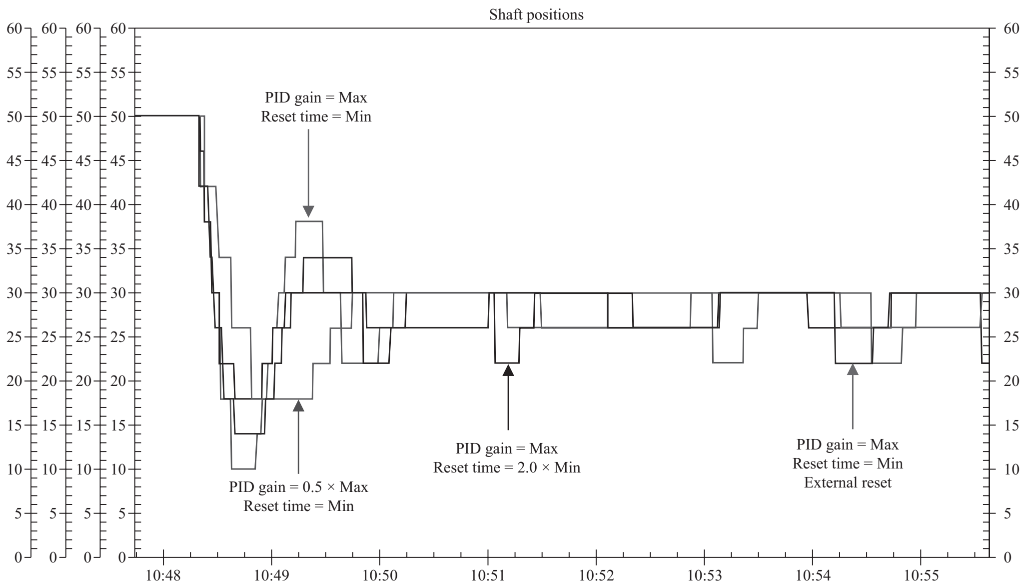


Figure 2.8f. Effect of 4% valve *resolution* on valve *stroke* for 22% load upset in *near-integrating* process.

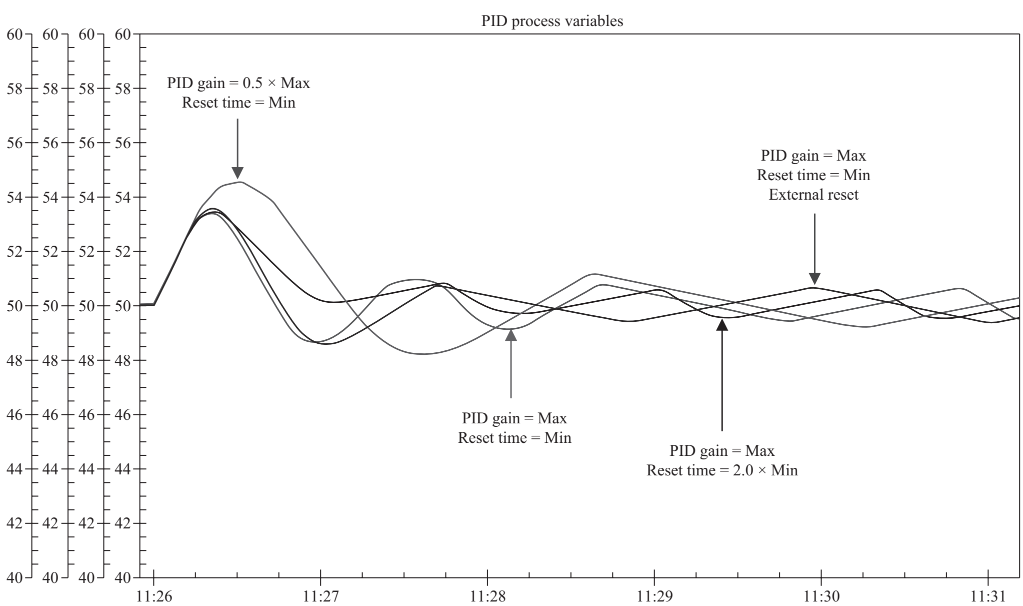


Figure 2.8g. Effect of 4% valve *resolution* on PID *PV* for 22% load upset in *true integrating* process.

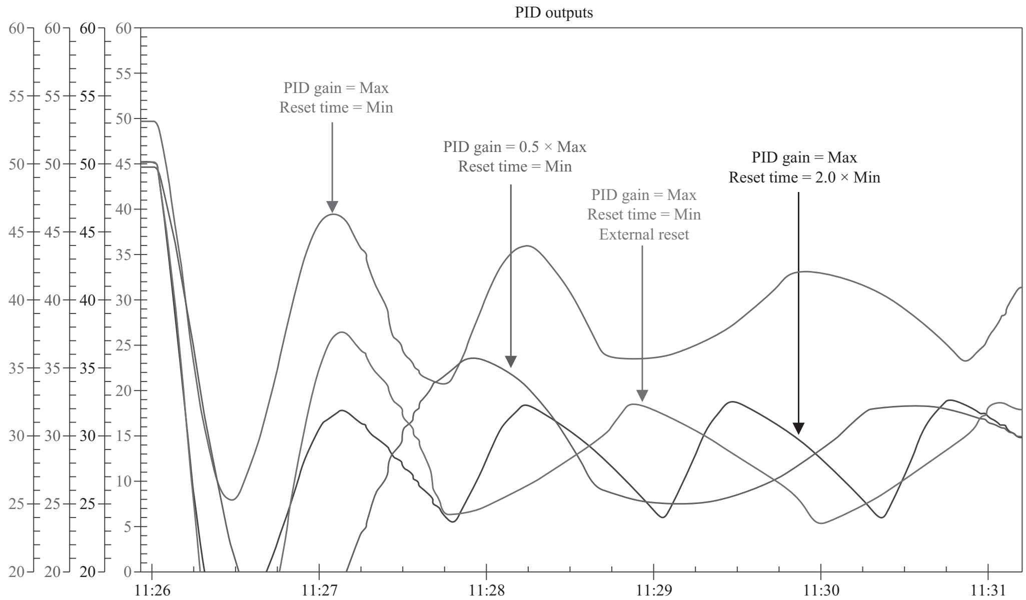


Figure 2.8h. Effect of 4% valve *resolution* on PID *output* for 22% load upset in *true integrating* process.

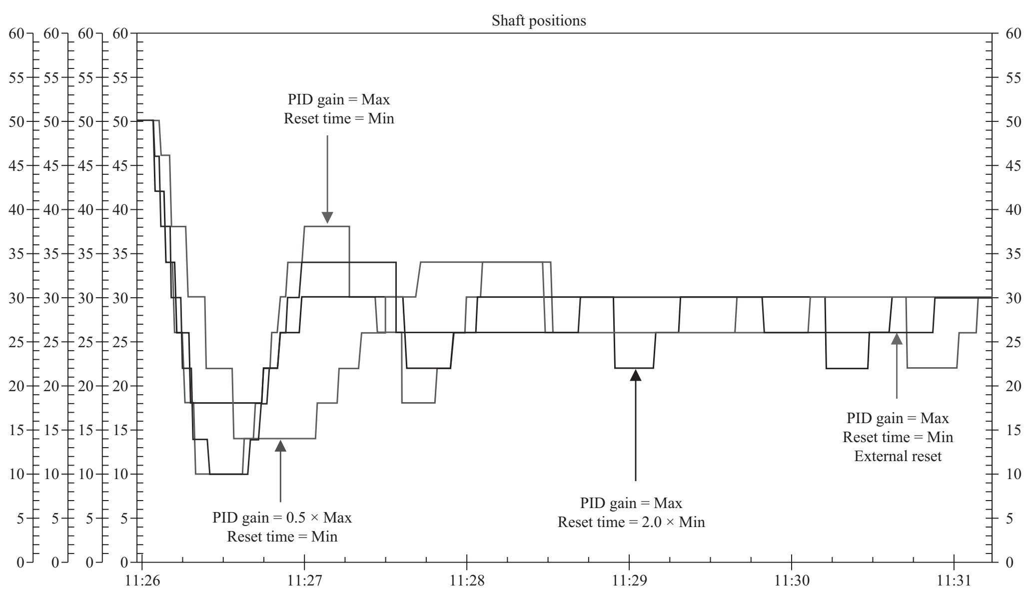


Figure 2.8i. Effect of 4% valve *resolution* on valve *stroke* for 22% load upset in *true integrating* process.

for a resolution of 4 percent for different gain and reset settings. The figures include a plot of external reset feedback enabled for the original aggressive tuning settings. External reset can stop a limit cycle in the moderate self-regulating process but the PV offset from setpoint may be unacceptable. The plots indicate nothing much else can be done to stop the limit cycle whose amplitude is basically fixed by the open loop gain. The limit cycle period can be increased by a halving the PID gain but the consequence in terms of a greater peak error is quite severe for integrating processes.

KEY POINTS

1. The capability of the measurements, valves, and VSDs and your imagination set the limits as to what PID can do.
2. Select measurements with the greatest precision (best resolution, threshold sensitivity, and repeatability) least drift, and required rangeability.
3. Design the measurement installation to provide the most representative PV with minimum noise, secondary lag, and dead time.
4. Select control valves, actuators, and positioners with the greatest precision (least backlash and stiction) and the best threshold sensitivity and resolution.
5. Use volume boosters with an adjustable bypass on each valve positioner output to make slewing rate faster and pre-stroke dead time smaller.
6. Use on-off valves for isolation and control valves for throttling and not vice versa.
7. Design for a valve pressure drop to system pressure drop greater than 0.2 and a static head to total head developed less than 0.1 for variable speed pumps.
8. Use VSD speed command input card with the best resolution (at least 12 bit).
9. Minimize deadband setting and maximize setpoint rate limit in VSD setup.
10. Minimize transmitter damping setting and PV signal filter.
11. Set the proper PID process action, valve action, output limits, and setpoint limits.
12. Use the full power of the PID by a better understanding of what, when, how, and why of PID features.
13. Select the best PID structure for the application.
14. Use signal characterization for known gain nonlinearity.
15. Use VPC to eliminate split range.
16. Add a feedforward signal of unmeasured load disturbances at the process input.
17. Use plantwide feedforward control to enable flexible manufacturing.
18. Use output tracking and remote output mode to provide and continuously improve repeatable intelligent actions practiced by operators and devised by process engineers to deal with startups, batch sequences, and abnormal conditions.
19. Use an enhanced PID with a threshold sensitivity limit to improve control using wireless devices and at-line or off-line analyzers and to provide directional move suppression for split range and VPC.
20. Use external reset feedback to prevent oscillations from a slow valve or VSD, slow secondary loop, and poor precision in a valve, VSD, or measurement.

21. Increase the Lambda from one dead time to a greater number of dead times (e.g., three) to provide more robustness and meet other objectives such as maximizing coordination (e.g., blending) and disturbance absorption (e.g., surge tank level).
22. Use integrating process tuning rules for near-integrating and runaway processes.

CHAPTER 3

Performance Criteria

3.1 INTRODUCTION

How well a loop performs depends upon the tuning. A slowly tuned loop will do as poorly as a loop with bad dynamics. Thus, money spent to improve loop dynamics by making valves and measurement faster or the delay in the process response smaller is wasted unless the controller is tuned for a reasonably fast response.

In this chapter, simple relationships are developed to show how tuning settings affect the ability of the loop to correct for unmeasured load responses to maximize process efficiency and capacity and reach setpoints faster to minimize batch cycle time and continuous system transition time. Guidance is also offered on how to eliminate setpoint overshoot while still achieving these other objectives.

This chapter focuses on how tuning settings determine the practical limits to control loop performance. The ultimate limits are determined by relative amount of loop dead time compared to the rate of change of the process variable. The practical limits become nearly as small as the ultimate limits for aggressive proportional-integral-derivative (PID) tuning settings that would give a critically damped response (fastest return to an existing setpoint or fastest approach to a new setpoint with no oscillation). However, these tuning settings make the PID excessively sensitive to nonlinearities and nonidealities. Consequently, the ultimate limits are more a benchmark than achievable goals. In practice, controllers are tuned sluggishly.

Appendix C will show the derivation of the ultimate limits determined by loop dynamics and the relationships to the practical limits determined by PID tuning. This appendix and Chapters 5 and 6 will show that an implied dead time can be simply computed from the tuning settings. Making the automation system faster or reducing process dead time has no effect if the total dead time is less than the implied dead time from the PID tuning. Thus, users can judge whether going with faster tuning is more important than capital investment in a faster automation system or a smaller process dead time, or vice versa.

3.1.1 PERSPECTIVE

The performance criteria primarily used in control theory text books is integrated absolute error (IAE) for a step disturbance. While IAE serves as a common benchmark, more focused metrics are needed along with other performance criteria. The accumulated amount of off-spec material within the unit operation is simply the integral of the error (integrated algebraic error) where plus and minus errors cancel out if there are no nonreversible reactions. The total amount of off-spec material sent downstream is the integral of the error multiplied by the unit operation discharge flow. During times of low feed rate, the effect on the error downstream is correspondingly diminished.

For plug flow unit operations, such as extruders, fluid bed reactor or cracking furnaces, the error from upstream operations is nearly fully propagated downstream. A unit operation with no back mixing is a sheet line or web line. Any variability in the feed is completely seen in sheets and webs. Short-term excursions are completely seen as changes in product gage and quality.

For unit operations involving process liquids, there is often considerable back mixing and attenuation of variability. The error amplitude in intermediate volumes and product storage tanks is filtered by the residence time that is the volume divided by the total feed flow. For parallel reaction or crystallization trains, concentration in the separation and purification area is a blend of multiple sources. Equations 6.6a and 6.6b in Chapter 6 can be used to estimate the filtering effect of well mixed volumes where the filter time is the residence time (e.g., liquid volume/feed rate). The filtering and blending effect must be included in most liquid chemical processes to provide a realistic estimate of the cost of the paths of variability. As inventories are reduced and production rates are increased, the cost of variability is increased.

The peak error (maximum error for a disturbance) is important when exceeding a limit can trigger a Safety Instrumented System (SIS), relief device, an environmental violation, or a hazardous reaction. Again a step disturbance is used for comparison of tuning methods and algorithms.

For a setpoint change, the metrics are rise time, overshoot, undershoot, and settling time. The time till the process variable (PV) reaches the new setpoint is the rise time. Overshoot is the maximum deviation of the PV observed as the maximum peak after the PV crosses setpoint. Undershoot is the subsequent maximum error of the opposite sign. The settling time is the time for the error to remain within a specified band around the setpoint.

The minimization of IAE, peak error, and rise time is often at the expense of overshoot, undershoot, and settling time. Consequently, unit operations where PV overshoot is of primary concern (e.g., chemical reactor temperature), the PID tuning is oriented towards a gradual approach to a new setpoint. For continuous operations, the IAE and peak error for disturbances are more of a concern since setpoint changes are relatively infrequent. The need for automated startup and flexible manufacturing has created more setpoint changes due to production rate and grade changes. Setpoint response is becoming increasingly important. Methods have been developed to eliminate overshoot without detuning.

For plug flow volumes particularly with recycle or heat integration, overshoot of the manipulated variable (MV) may also be a concern because of interactions and secondary effects. MV overshoot may be more important than rise time in refineries and some petrochemical plants.

The fact that the IAE is predicted and discussed almost exclusively, for a step disturbance being the worst case, opens an opportunity to track down and eliminate or slow down disturbances. These opportunities will be explored extensively in Chapter 8.

There is an ultimate limit to the IAE, peak error, and rise time based on dead time and the excursion rate in the PID time frame. However, the actual performance criteria observed depends upon the PID tuning. Thus, a slowly tuned PID will behave as badly in terms of these performance criteria as a poorly designed process or automation system with excessive dead time. The implied dead time from tuning is a useful parameter.

3.1.2 OVERVIEW

The equations developed in Appendix C for the ultimate limit based on dynamics and the practical limit based tuning are presented and discussed for IAE, peak error, and rise time. The use of a setpoint lead-lag or PID structure with adjustable weight factors on error for the proportional and derivative modes (e.g., two degrees of freedom [2DOF] PID) is recommended to minimize overshoot and undershoot.

The practical limit imposed by tuning is used to estimate an implied dead time. Efforts to reduce the dead time below this value have no observable benefit unless the PID is retuned. The implied dead time is an eye opener and helps the user realize whether money and time should be spent on tuning or process or automation system improvement.

3.1.3 RECOMMENDATIONS

1. To minimize the peak error for large and fast disturbances, maximize the controller gain.
2. To minimize the integrated algebraic error for large and fast disturbances in back mixed liquid volumes, maximize the PID gain time to give MV overshoot particularly if the process has an integrating or runaway response. Be careful not to have too small of an integral time particularly if the maximum PID gain is not used. The allowable integral time is inversely proportional to the product of the integrating process gain and PID gain.
3. To minimize the IAE for large and fast disturbances in plug flow liquid and gas volumes, minimize the integral time and decrease the controller gain as necessary to prevent MV overshoot.
4. To minimize the rise time, maximize the controller gain.
5. To prevent overshoot while minimizing rise time, use a setpoint lead-lag or a PID structure with adjustable setpoint weight factors (see Chapter 2). If rise time is not important, simply use a setpoint filter with the filter time set equal to reset time or a PD on PV and I on error structure (setpoint weight factors are zero).
6. For the fastest possible rise time and settling time use an output tracking strategy for a full throttle response (see Chapter 12).

3.2 DISTURBANCE RESPONSE METRICS

If we did not have unmeasured disturbances, we would not need feedback control beyond flow loops. We could simply set the controller output to a fixed value based on a process flow diagram (PFD) and improved by experimentation. In fact, this is the typical mindset of a process design where sequences of flows are optimized in fed-batch operations. In some cases, the process design engineer does not want to relinquish control of the MV (e.g., flow setpoint or valve or variable speed drive [VSD] signal).

A feedback loop is designed to transfer the variability from the controlled variable to the MV. In most cases, this transfer of variability is maximized to reduce peak and integrated errors for unmeasured disturbances. There are cases such as surge tank level control where the transfer of variability is minimized to just keep the PV within bounds. This objective is called maximization of the absorption of variability.

The fact that a feedback controller output changes for a given setpoint means there are disturbances either as load upsets or measurement signal drift. The advent of smart instrumentation has reduced and slowed down measurement drift to a negligible value except for pH electrodes and thermocouples. Mostly what we see as changes in the PID output today are the effect of process input disturbances (load upsets). For PID gains greater than one or reset times less than one minute, a trend chart of the PID output rather than the PID PV input is more indicative of what is going on in the process. In fact, for incredibly tight control, the PID PV chart is rather boring except for large sudden upsets.

Engineers tend to emphasize qualitative criteria such as loop importance and ease of tuning rather than quantitative criteria such as error size and duration. This qualitative emphasis is due partly to the complexity and diversity of the quantitative criteria and the associated analysis techniques. For example, a level loop that has a nonself-regulating response may be judged easy to control even though the errors may be large in size and duration because these errors are unimportant as long as the tank does not run dry or overflow. A temperature loop with a large time constant may be judged difficult to control even though the errors are small because the slowness of the loop response makes tuning extremely tedious. To judge objectively whether a loop is easy to control, quantitative criteria that are generally applicable and readily understood should be used. All quantitative criteria can be broadly classified as, and simplified to, either an accumulated error, or peak error criteria, or a combination thereof.

Figure 3.1 graphically depicts the accumulated error and peak error metrics. Figure 3.1 is the PID PV response to a large step disturbance at the process input for a near-integrating, true

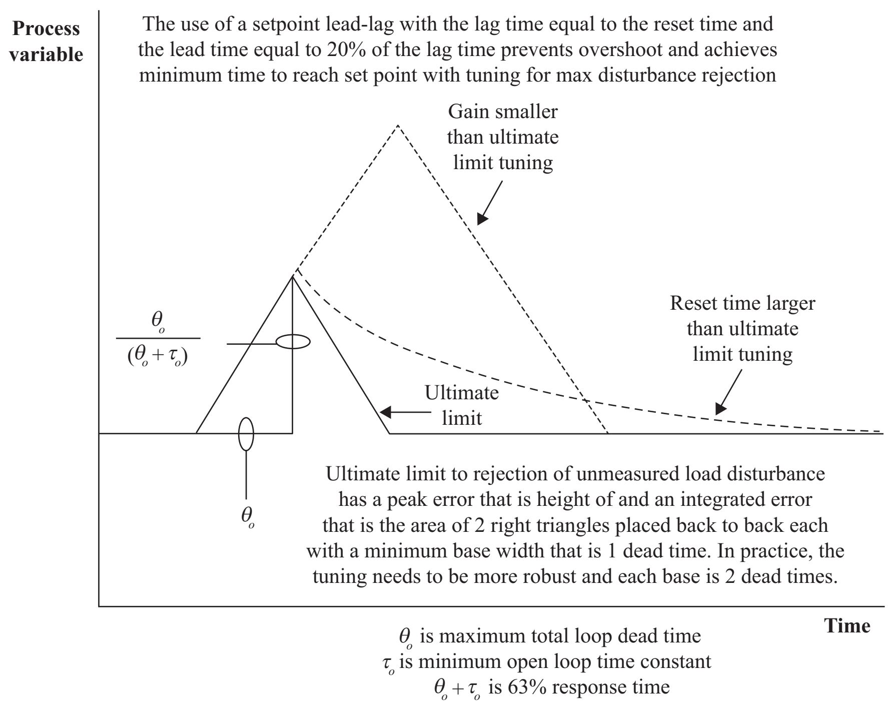


Figure 3.1. Peak error and integrated error for load response.

integrating, or runaway process. A moderate self-regulating process would show some deceleration of the PV response and dead time dominant processes would reach their final value. The straight line takeoff (lack of an initial bend) of the PV response is the result of a negligible secondary time constant.

The primary metrics for load disturbances (disturbances at the process input) are peak error and integrated error (accumulated error). Figure 3.1 shows that the minimum peak error for the most aggressive tuning is reached in one loop dead time from feedback control. The minimum integrated error is the area of two right triangles placed back to back. A smaller PID gain increases the peak error. A slower integral action (larger reset time) most notably slows the return to setpoint. Internal Model Control and Lambda tuning methods have been criticized for too low a PID gain and too large a reset time for loops with a large primary process time constant that resulted from an original emphasis on setpoint response. The use of an arrest time equal to the dead time and a switch from self-regulating to integrating process tuning for near-integrating processes with a low limit for reset time of four dead times for PI control and rate time of half the dead time for PID control results in tuning settings that is close to the optimum, eliminating this criticism. Appendix C extensively discusses the relationships between these metrics and loop dynamics.

3.2.1 ACCUMULATED ERROR

Accumulated error is the totalized deviation of the controlled variable from set point. For a composition control loop (e.g., a distillation column temperature or pH control loop), the accumulated error multiplied by the average product flow provides a measure of the total amount of the product that deviates from the optimum product purity specification. For a flow control loop (e.g., a reactor feed or a furnace fuel loop), the accumulated error provides a measure of the total amount of feed that deviates from the stoichiometric ratio specification that may be important not only for product purity, but also for safety (if the excess feed is flammable, explosive, exothermically reactive, or toxic) or emission control (if the excess feed or associated by-products are environmentally restricted). If the controlled variable is a utility flow (e.g., steam flow or cooling water flow), the accumulated error is representative of energy usage in excess of the set point. The accumulated error is the integrated error where the positive and negative errors are canceled by the volume of the system to provide a net accumulated positive or negative error. IAE is equal to the accumulated error for an over damped or critically damped (non-oscillatory) response. Integrated squared error (ISE) can be approximated by a combination of the accumulated error and peak error. The additional value of calculating the IAE and ISE errors is usually not worth the additional effort by the design or maintenance engineer for loop performance analysis. The exception occurs for split ranged flows where oscillations across the split range point can cause wasted energy from alternating between cooling and heating and wasted reagent from alternating between acid and base addition. However, a small measured accumulated error does not necessarily mean a well-tuned stable loop. A small measured accumulated error can result from a loop that is marginally stable since the positive and negative errors will cancel for sustained oscillations. For a limit cycle, the accumulated error will generally steadily increase with time because the oscillation is rarely centered on the setpoint. The accumulated error can be accurately calculated by use of a relatively simple algebraic equation if the oscillations are decaying in amplitude.

The equations presented here for a self-regulating process can be used for integrating and runaway processes. The open loop error is the ramp rate if the controller was in manual. The time units in the open loop error are cancelled out when the open loop integrating process gain is substituted for the open loop self-regulating process gain

Equation 3.1 shows that the accumulated (integrated) error for a closed loop can be calculated for a given size load disturbance from the PID gain, reset setting, execution time, and filter time (Shinskey 1996). The measurement gain (100 percent divided by measurement span is included to get the error in percent of PID scale. The disturbance size is expressed as the percent change in PID output required ($\Delta\%CO$). For dead time compensators and the enhanced PID for wireless, the reset time can be decreased to the point where these equations will predict too low of an integrated error. The result of these equations cannot be taken to be less than the ultimate limit as exemplified by Equation C.3 in Appendix C that is determined by the amount of actual dead time in the loop.

$$E_i = \frac{T_i + \Delta t_x + \tau_f}{K_m * K_c} * \Delta\%CO \tag{3.1}$$

A more useful formulation of the equation uses the size of the open loop error (error if PID was in manual) from the disturbance. The change in valve or VSD flow which is the change in controller output ($\Delta\%CO$) multiplied by the valve or VSD gain (K_v) must balance out the change in the disturbance variable (ΔDV) multiplied by the disturbance gain (K_d). The disturbance gain converts the load upset to the required change in manipulated flow.

$$\Delta\%CO * K_v = K_d * \Delta DV \tag{3.2}$$

The load multiplied by the gains gives the open loop error in engineering units.

$$E_o = K_d * K_r * K_p * \Delta DV \tag{3.3a}$$

Solving Equation 3.2 for ΔDV and substituting for ΔDV in Equation 3.3a, cancels out K_d.

$$E_o = K_v * K_r * K_p * \Delta\%CO \tag{3.3b}$$

Solving Equation 3.3b for $\Delta\%CO$ and substituting the result into Equation 3.1 gives the integrated error in terms of the open loop error.

$$E_i = \frac{T_i + \Delta t_x + \tau_f}{K_c * K_v * K_r * K_p * K_m} * E_o \tag{3.3c}$$

$$K_o = K_v * K_r * K_p * K_m \tag{3.3d}$$

$$E_i = \frac{T_i + \Delta t_x + \tau_f}{K_c * K_o} * E_o \tag{3.4a}$$

For a negligible filter and execution time, the Equation 3.4a simplifies to 3.4b

$$E_i = \frac{T_i}{K_c * K_o} * E_o \tag{3.4b}$$

The definition of the open loop gain in Equation 3.3d yields the final equation realizing that many loops are designed for a full scale valve stroke and gives a full scale change in the PV ($K_o = 1$).

$$K_o = \frac{\Delta F_v}{\Delta \% CO} * \frac{\Delta R}{\Delta F_v} * \frac{\Delta PV}{\Delta R} * \frac{\Delta \% PV}{\Delta PV} = 1 \tag{3.5}$$

$$K_o = K_v * K_r * K_p * K_m = 1 \tag{3.6}$$

For the open loop gain in Equation 3.6 and enabling a focus on the fundamental relationship of integral time and controller gain on the integrated error.

$$E_i = \frac{T_i}{K_c} * E_o \tag{3.7}$$

Nomenclature for Accumulated Error Metric:

E_i = integrated error from load disturbance (PV e.u. * sec)
E_o = open loop error (load disturbance error with controller in manual) (PV e.u.)
$\Delta\%CO$ = change in controller output converted to percent of PID output scale (%)
ΔDV = change in disturbance variable (load) (load e.u.)
ΔF_v = valve or VSD flow change (flow e.u.)
K_c = controller gain (dimensionless)
K_d = disturbance gain (flow e.u./load e.u.)
K_m = measurement gain (%/PV e.u.)
K_o = open loop self-regulating process gain (%/%) (dimensionless)
K_p = process gain (PV e.u.)
K_r = flow ratio gain (often imbedded in process gain [1/flow e.u.])
K_v = valve or VSD gain (flow e.u./%)
$\Delta\%PV$ = change in PV converted to % of controller input scale (%)
ΔPV = change in PV (PV e.u.)
ΔR = change in ratio of manipulated flow to load flow (dimensionless)
T_i = integral time (reset time) (sec)
τ_f = signal filter time constant (sec)
Δt_x = controller execution time (sec)

The equation for accumulated error leads to several conclusions important enough to emphasize. First, if the disturbances are nearly zero in magnitude (E_o = 0.0), even the most difficult loop will perform excellently. Thus, before one can decide whether a difficult loop justifies the additional expense of special equipment, special instruments, or advanced control algorithms, knowledge of the size of the disturbance is necessary. Second, if the controller was tuned with too small a gain (too large a proportional band) or too large a reset time (too slow a reset action), an easy loop will perform as poorly as a more difficult loop that required the slower mode settings. Any special efforts or expense during design to improve loop performance will be wasted if overly conservative controller tuning is used in the field. Third, if the PID execution time or filter time is large compared to the reset time, the accumulated error could be too large for even the best tuning. Fourth, if the process gain is increased, the open loop error, and hence the accumulated error, increases as per Equation 3.3b for a given disturbance. To recognize this relationship, you need to realize the K_p in the denominator is canceled by the K_p in the expression for PID gain K_c since the maximum K_c is inversely proportional to K_p. It is therefore important that the process engineer and control engineer review the effect of equipment design and operating conditions on the process gain at an early stage of the project (preferably before the project appropriation request is made final). In applications where the measurement is an inference of a process condition (e.g., temperature is an inferential measurement of composition), a high process gain (high process sensitivity) provides better recognition of the disturbance ensuring a small change in the process condition is above the threshold sensitivity limit of the sensor and resolution limit of the PID. The resolution limit of wide range thermocouple input cards in the 1980s vintage Distributed Control Systems (DCS) was about 0.25°C causing steps in the PID measurement that prevented the use of high PID gain and rate time settings. An increase in PID execution time and an increase in PID filter time settings helped to smooth the steps and increase the signal to noise ratio but the accumulated error as per Equation 3.4a was adversely affected.

An increase in valve, VSD, or measurement gain will increase the error and dead time from resolution and sensitivity limits since these are in percent of valve or VSD capacity and in percent of measurement scale span. Oversized control valves can make the problem even worse from the steeper slope and greater stiction from operation near the closed position. Undersized valves can cause operation on the nearly flat upper portion of the installed flow characteristic causing a loss of control from a control valve gain (slope of characteristic) that is nearly zero.

For pH measurement and control, the strategy is to reduce the process gain as much as possible by operating on the flatter portion of the titration curve. The curve gets flatter at points on the curve where an acid or base effect changes dramatically expressed logarithmically via an acid dissociation constant (pKa) (McMillan 2004). The incredible resolution and sensitivity in terms of hydrogen ion concentration from the logarithmic relationship between hydrogen ion concentration and pH enables a focus on reducing the process gain. However, the curve must not be too flat causing a loss of control from a process gain (titration curve slope) that is nearly zero. A titration curve can get too flat at high pH (e.g., pH > 12) from the proximity of the water dissociation constant (pKw).

For distillation column measurement and control, the strategy is to increase the process gain as much as possible by selecting the column tray that shows the highest sensitivity to changes in composition and hence reflux to feed flow or steam to feed flow ratio.

3.2.2 PEAK ERROR

Peak error is the maximum deviation of the controlled variable from set point. For a reactor temperature or reactor pH loop, the maximum deviation might have to be limited to prevent the start of an undesirable secondary reaction. For a pressure control loop, the maximum deviation might have to be limited to prevent actuation of relief valves or vacuum breakers. For a Claus furnace combustion control loop, the maximum emission level of sulfur dioxide must meet environmental regulations (a Claus furnace produces sulfur from sulfur dioxide waste gas). For a waste treatment neutralizer, the pH must always be between low and high environmental limits to prevent a recordable incident. The peak error is equal to approximately 1.5 times the steady-state offset for a proportional-only controller (Harriot 1964).

For proportional-only (P-only) controller:

$$E_x = \frac{1.5}{K_c * K_o + 1.0} * E_o \tag{3.8}$$

The peak error is reached in about one dead time. The integral mode contribution is the number of repeats of the proportional mode contribution until the peak error is reached giving Equation 3.9. For processes with a small dead time to time constant ratio, the integral time is much larger than the dead time and the peak error is mostly affected by the proportional mode. For processes with a large dead time to time constant ratio (dead time dominant), the integral time for most tuning methods is about half the dead time, and the integral mode contribution approaches the proportional mode contribution. For Lambda tuning of dead time dominant self-regulating processes, the integral time is set equal to the time constant which makes the integral contribution much larger than the proportional mode contribution. Note that the controller gain affects the contribution from both modes. For dead time dominant processes the controller gain is less than the inverse of the open loop gain resulting in the peak error approaching the open loop error.

For a proportional-integral (PI) controller:

$$E_x = \frac{1.5}{K_c * K_o * \left(1.0 + \dfrac{0.5 * \theta_o}{(T_i + \Delta t_x + \tau_f)}\right) + 1.0} * E_o \tag{3.9}$$

For processes with a small time constant to dead time ratio (dead time dominant), the product of the PID gain and open loop steady state gain is about 0.2 and the reset time is less than half the dead time. The denominator ends up being about equal to the numerator. The result is the peak error is about equal to the open loop error which is consistent with the observation that there is little if any reduction in open loop error by feedback control for an unmeasured step load disturbance. There is more integral action than proportional action and the process has reached close to its final value in one dead time, the minimum time needed for seeing and correcting for an unmeasured disturbance.

For processes with a large time constant to dead time ratio, a feedback controller is potentially capable of making the peak error a small fraction of the open loop error if the controller is aggressively tuned to make the most of the proportional and derivative mode to provide an immediate reaction. The numerator in Equation 3.9 reduces from 1.5 to 1.25 from the derivative mode contribution giving Equation 3.10a.

For a PID controller:

$$E_x = \frac{1.25}{K_c * K_o * \left(1.0 + \dfrac{0.5 * \theta_o}{(T_i + \Delta t_x + \tau_f)}\right) + 1.0} * E_o \tag{3.10a}$$

For PID control of these processes dominated by a large primary time constant tuned with an aggressive gain as a near-integrator to minimize peak error, the equation simplifies to:

$$E_x = \frac{1.1}{K_c * K_o} * E_o \tag{3.10b}$$

Nomenclature for Peak Error Metric:

E_x = peak error from load disturbance (e.u. * sec)
E_o = open loop error (load disturbance error with controller in manual) (e.u.)
K_c = controller gain (dimensionless)
K_o = open loop self-regulating process gain (%/%) (dimensionless)
T_i = integral time (reset time) (sec)
τ_f = signal filter time constant (sec)
Δt_x = controller execution time (sec)

3.2.3 DISTURBANCE LAG

A disturbance lag (time constant) does not have as much an effect on integrated error as the peak error. A disturbance lag that is much greater than the dead time offers a considerably smaller peak error. However, the accumulated error is about the same because of the long protracted return to setpoint as seen in the Chapter 8 test cases. A disturbance dead time does not affect the peak or integrated errors but does affect feedforward timing.

If we consider a load disturbance that is not a step change but has a time constant, the peak error is reduced to the fraction of the load change in one dead time. We can approximate the effect of a load time constant (τ_L) as the exponential response of the open loop error in one dead time giving Equation 3.11.

$$E_x = \frac{1.25}{K_c * K_o * \left(1.0 + \frac{0.5 * \theta_o}{(T_i + \Delta t_x + \tau_f)}\right) + 1.0} * (1 - e^{-\theta_o / \tau_L}) * E_o \qquad (3.11a)$$

For PID control of processes dominated by a large time constant (the case of more prominent interest), the equation simplifies to:

$$E_x = \frac{1.1}{K_c * K_o} * (1 - e^{-\theta_o / \tau_L}) * E_o \qquad (3.11b)$$

Nomenclature for the Effect of Load Disturbance Lag and Lambda on Peak Error:

E_o = open loop error (load disturbance error with controller in manual) (e.u.)
E_x = peak error from load disturbance (e.u.)
K_c = controller gain (dimensionless)
K_o = open loop gain (%/%) (dimensionless)
T_i = integral time (reset time) (sec)
θ_o = total loop dead time (sec)
τ_f = signal filter time constant (sec)
τ_L = load disturbance time constant (sec)
Δt_x = controller execution time (sec)

See Chapter 8 for a more extensive analysis and set of equations on the effect of disturbance time constants in the source (e.g., flow loop) and intervening volumes. Equations are also presented to show the effect of disturbance rate limits and oscillations on primary loops used for composition, pH, and temperature control.

3.3 SETPOINT RESPONSE METRICS

In the normal operation of continuous processes, the setpoint response is of minimal concern because the setpoint is rarely changed. During startup and product type or grade transitions, the setpoint response can become as important as for batch operations.

The time to reach a new setpoint (rise time) is minimized to reduce startup, transition, and batch times. The maximum deviation after the first crossing of the setpoint (overshoot) must be minimized for many of the same reasons for minimizing peak error. Note that overshoot is the maximum deviation below setpoint for a decrease in setpoint. Bioreactors are particularly sensitive to overshoot in batch temperature or pH. An overshoot can result in cell death. In highly exothermic chemical reactors, an overshoot can trigger a runaway reaction. The maximum deviation after a second crossing of the setpoint (undershoot) generally lowers reaction yield from reduced growth and product formation rates for bioreactors and low reaction rates for chemical reactors. Figure 3.2 shows the rise time, overshoot, and undershoot for step increase in setpoint.

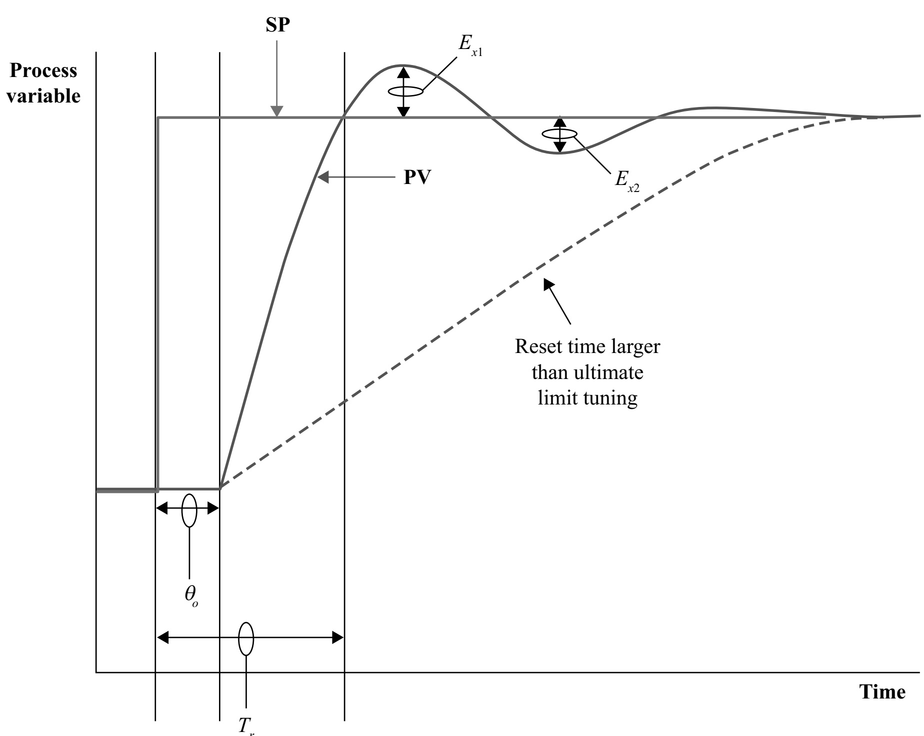


Figure 3.2. Rise time (T_r), overshoot (E_{x1}), and undershoot (E_{x2}) for setpoint response.

3.3.1 RISE TIME

The minimum rise time (T_r) can be approximated as the change in percent setpoint ($\Delta\%SP$) divided by the maximum rate of change of the PV plus the total loop dead time. For an integrating or *near-integrating* process, the maximum percent PV (%PV) ramp rate is the integrating process (K_i) gain multiplied by the change in controller output. The change in controller output is the minimum of the maximum allowable change per output limit ($\Delta\%CO_{\max}$) and the change as the result of the controller gain (K_c) and feedforward gain (K_{ff}) as detailed in the denominator of Equation 3.12a. If the step change in controller output from the proportional mode for a structure of proportional action on error is less than the maximum available output change ($\Delta\%CO_{\max}$) that is the difference between current output and output limit, Equation 3.12a simplifies to Equation 3.12b for feedback control. The output change must be corrected for methods used to make the setpoint response faster. For setpoint feedforward, the step change in output is a combination of the feedforward and feedback action. For smart bang-bang logic, the step output change is the maximum available output change. Setpoint rate limits and a small setpoint filter with external reset feedback can be used to smoothen out step changes in flow from utility and raw material systems.

$$T_r = \frac{\Delta\%SP}{K_i * \min(|\Delta\%CO_{\max}|, (K_c + K_{ff}) * \Delta\%SP)} + \theta_o \tag{3.12a}$$

For a maximum available output change larger than the step from the proportional mode ($\Delta\%CO_{max} > K_c * \Delta\%SP$) the change in setpoint in the numerator and denominator cancel out, yielding a simpler equation (no setpoint feedforward):

$$T_r = \frac{1}{(K_i * K_c)} + \theta_o \tag{3.12b}$$

For the *near-integrating* process response seen in vessel and column temperature loops where the process time constant is significantly larger than the total loop dead time, the integrating gain is the open loop gain (K_o) divided by the open loop time constant (τ_o) yielding Equation 3.12c.

$$T_r = \frac{\tau_o}{(K_o * K_c)} + \theta_o \tag{3.12c}$$

A structure with derivative action on error can help get through large a dead band and resolution limit for small setpoint changes and small gains. The result can be an increase in the initial rate of approach to the setpoint. Otherwise, the impact on rise time from derivative action on error is minimal. Normally, a structure of derivative action on PV (e.g., PI on error and D on PV) or a small 2DOF gamma is chosen to minimize kicking in the PID output for setpoint changes.

Step changes and oscillations in the PID output can be disruptive to other loops that are relying upon a constant pressure in the same volume (e.g., utility header or inline system) that is source or destination of the flow manipulated by the PID. Some tuning techniques are designed to make the changes in the PID output as smooth as possible. Some go so far as to choose to minimize overshoot of the PID output beyond its final resting value (FRV). While this objective may work for volumes with no back mixing and short residence times (e.g., catalyst bed gas reactors and plug flow liquid reactors), preventing step changes in the PID output and overshoot of the FRV is not feasible for integrating processes (e.g., level or batch temperature) and downright dangerous for runaway processes because an excursion in the PV will not reverse direction until the PID output correction goes beyond the FRV. While prevention of PID output overshoot is possible for continuous processes, it is not advisable for well mixed volumes with large residence times that consequently have a large primary process time constant. The rise time can become orders of magnitude larger for these near-integrating processes. A more prevalent goal is to focus on the minimization of PV overshoot (peak error after first crossing of the setpoint) and undershoot (peak error after second crossing of setpoint). Undershoot is in the opposite direction of overshoot.

Nomenclature for Rise Time Metrics:

$\Delta\%CO_{\max}$ = maximum change in controller output converted to % of PID output scale (%)
$\Delta\%SP$ = step change in setpoint converted to percent of PID input scale (%)
K_c = controller gain (dimensionless)
K_{ff} = feedforward gain (%/%) (dimensionless)
K_i = open loop integrating process gain %/sec per % delta PID output (1/sec)
K_o = open loop self-regulating process gain (%/%) (dimensionless)
T_r = rise time (time to reach a new setpoint) (sec)
τ_o = open loop time constant (primary time constant) (sec)
θ_o = total loop dead time (sec)

3.3.2 OVERSHOOT AND UNDERSHOOT

In general, a decrease in rise time causes an increase in overshoot for feedback control. Overshoot and undershoot of the PV can be reduced by making the contribution from the proportional and derivative modes optimally larger than the contribution from the integral mode. The integral mode is the principal cause of overshoot because this mode does not reverse direction of the change in PID output until the PV crosses setpoint. The integral mode has no sense of direction as noted in the Appendix B example of temperature control.

Overshoot and undershoot eliminated by a PD on PV and I on error structure or a setpoint filter time equal to the reset time at the expense of a considerable increase in rise time. The tradeoff between rise time and overshoot can be optimized by the use of 2DOF structure and a setpoint lead-lag. An output tracking (remote output) bang-bang control strategy can eliminate the tradeoff for near-integrating, true integrating, and runaway processes. Test results in Chapters 2 and 12 show the value and the adjustment of parameters for these approaches.

For batch temperature control where there is only heating or only cooling (no split range), and there is no significant heat generation from reaction or heat loss from evaporation in the process, the response is in one direction only (unidirectional). For heating the temperature can only rise and for cooling the temperature can only decrease. A one sided response also occurs in batch pH control where there is only an acid or only a base reagent and there is no significant reagent consumption from reaction. For these one sided responses, integral action will cause overshoot. The structure must have proportional action on error, derivative on PV and no integral action. When integral action is not used, an output bias must be specified. Generally for these one sided batch applications, the output bias is the PID output when the PV is at setpoint.

KEY POINTS

1. The accumulated (integrated) error provides a good measure of the amount of material that is off-spec and energy or raw material that is wasted.
2. The peak error is important to meet instantaneous constraints such as pressure ratings and environmental limits.
3. Even normally difficult loops such as pH can appear to be easy if the disturbance is very small or very slow.
4. Sluggish tuning settings will make a well-designed or easy loop perform as badly as a poorly designed or difficult loop.
5. A higher process gain determined by the process and equipment design causes a larger open loop error but decreases the effect of measurement sensitivity limits.
6. For pH measurement and control, the incredibly high sensitivity of the electrode to hydrogen ion concentration enables a focus on reducing the process gain by seeking to operate on the flat portion of the titration curve.
7. For distillation column measurement and control, the limited sensitivity of the sensor (particularly a thermocouple) means the tray with the greatest change in temperature for a change in feed concentration of manipulated flow ratio is best.
8. An increase in the control valve or VSD capacity will provide more muscle to compensate for errors but the resolution limit or dead band in terms of flow is undesirable.
9. A reduction in the transmitter span provides better recognition of the disturbance and improves measurement accuracy and resolution but increases the measurement gain and consequently the open loop gain. The peak and integrated errors in engineering units will be the same if the PID gain is decreased so that the product of the open loop gain and PID gain remains the same. However, the errors in terms of percent of scale will be larger due to the smaller measurement span. If the gain is not decreased and the response is nonoscillatory, the peak and integrated errors in engineering units will be smaller than before the recalibration.
10. The accumulated and peak errors decrease as the PID gain increases (proportional band decreases). The accumulated error also decreases as the reset time decreases (repeats per minute increases). Note that the accumulated error for a proportional-only loop is infinite due to offset.

CHAPTER 4

Effect of Process Dynamics

4.1 INTRODUCTION

Most engineers are not taught how equipment and piping design affect loop dynamics and performance. Many of the more demanding control applications are the result of poor process dynamics. Automation engineers can help bridge the gap and be able to intelligently discuss how plant design is affecting plant performance. The perspective, overview, recommendations, concepts, and relationships offered can provide the understanding and starting point for making the most of process knowledge. The result can be a better focus on process control and mechanical design improvements with the greatest impact on the bottom line.

4.1.1 PERSPECTIVE

In general, an increase in back mixed volumes is beneficial in terms of increasing the process time constant or decreasing the integrating process gain slowing down and attenuating load disturbances. In contrast, an increase in a plug flow volume (e.g., pipe and dip tube volume) or decrease in mixing (e.g., increase in turnover time) is detrimental in terms of increasing process dead time. A decrease in heat transfer coefficient will increase the secondary time constant presenting a problem in terms of slowing down corrective actions that is particularly problematic for integrating and runaway processes. Such insights are seen in the typical values in Table 4.1 and the relationships in Table 4.2 in Section 4.2 on the Effect of Mechanical Design.

An essential realization for proper loop analysis is that time constants in series create dead time. These time constants can occur anywhere in the control loop. Multiple process time constants primarily arise from volumes in series and heat transfer surfaces. Section 4.3 will show how to estimate the equivalent dead time from small time constants anywhere in the loop and large equal process time constants.

In the first order plus dead time (FOPDT) approximation, the largest time constant is singled out and a fraction of all other time constants is taken as dead time. For equal time constants, a fraction of the totalized time constants is used as dead time. The equations are for noninteracting time constants. Pairs of interacting time constants can be converted to noninteracting time constants by means of equations in Appendix I.

Table 4.1. Typical values of process dynamics

Control loop description	Process dead time	Process time constant	Open-loop self-regulating or integrating process gain
Linear liquid flow	0.05–0.5 sec	0.1–1.0 sec	0.5–2
Liquid pressure	0.05–0.5 sec	0.05–0.5 sec	0.5–2
Linear gas flow	0.05–0.5 sec	0.05–5.0 min	0.5–2
Gas pressure (psig)	0.05–0.5 sec	0.05–5.0 min	0.5–2
Gas pressure (inches w.c.)	0.05–0.5 sec	0.05–5.0 min	10–100
Liquid level (vertical vessel or sump)	0.05–0.5 sec	0.05–1.0 min (secondary)	0.000001–0.01 per sec ramp
Batch vessel temperature (coolant to jacket/coils)	1–5 min	2–10 min (secondary)	0.000001–0.01 per sec ramp
Continuous vessel temperature (coolant to jacket/coils)	2–10 min	20–200 min	0.5–5
Exchanger temperature (coolant to shell/tubes)	1–2 min	2–10 min	0.5–5
In-line temperature (direct mix hot and cold)	2–10 sec	2–10 sec	0.5–5
Column temperature (material balance)	0.2–2 hrs	2–20 hrs	0.1–1.0
Vertical well-mixed tank pH (circulation line injection)	0.1–1.0 min	5–20 min	0.1–10,000
Horizontal poor-mixed tank pH (circulation line injection)	10–50 min	10–50 min	0.1–10,000
Vertical well-mixed tank pH (dip tube injection)	1.0–60 min	5–20 min	0.1–10,000
Pond or lagoon pH	5–50 hrs	1–5 hrs	0.1–10,000
In-line pH (injector tip check valve)	2–10 sec	2–10 sec	0.1–10,000
In-line pH (back filled injector)	1–20 min	2–10 sec	0.1–10,000

The equivalent dead time from time constants in series is added to the pure dead time. If we are just looking at the process, the total is a process dead time. However, we need to include the dead time from the automation system to arrive at a total loop dead time that determines tuning and loop performance.

The extremely high pH gain of 10,000 shown in the table is for a strong acid and strong base with no buffering and negligible carbon dioxide absorption.

The process sources of pure dead time typically involve transportation or mixing delays. Examples of transportation delays are composition or temperature changes propagating through a dip tube, extruder, pipeline, sheet line, static mixer, or any plug flow volume.

If the largest time constant is in the process downstream of load disturbances, the process time constant will act to slow down the excursion rate for these disturbances giving the proportional-integral-derivative (PID) controller time to catch up and correct for the load change. If the largest time constant is in the measurement or PID input signal, the disturbance seen by the PID is still slowed down but the actual process variable is not. The PID is seeing a filtered version of the real disturbance. The equations for estimating PID tuning and loop performance do not distinguish where the largest time constant is located. Consequently, the loop performance predicted must be corrected if the largest time constant is in the measurement or PID input signal as per Equations 6.7a or 6.7c in Chapter 6.

The fraction of time constants not converted to dead time should be added to the largest time constant to give an open loop time constant. Open loop denotes the fact that the response being measured is for the PID controller output in manual, remote output, or output tracking. The term *open loop* denotes no feedback action.

The process gain for composition and temperature control is a function of a flow ratio (e.g., ratio of manipulated flow to feed flow). The process gain for level and gas pressure is a function of the difference in flow going into and out of the volume. Plots of the process variable gain should be versus a flow ratio for composition and temperature and versus a flow difference for level and gas pressure.

The controller tuning depends upon the product of the valve or variable speed drive (VSD) gain, process gain, and measurement gain, termed the open loop gain in this book. The term denotes the gain does not include the effect of closed loop control and is identified when the loop is open (e.g., PID is in manual or remote output). The open loop gain is the percent change in the process variable divided by the percent change in PID output. The open loop gain for self-regulating and runaway processes is dimensionless (%/%). The open loop gain for integrating processes has units of 1/sec (%/sec per %) because the process variable is ramping.

Note that the literature commonly calls the open loop gain a process gain; the open loop time constant a process time constant; and the total loop dead time a process dead time. The more descriptive terms used in this book serves as a reminder that the dynamics of the automation system, besides the process, affect tuning and loop performance.

A key concept important for a unified methodology as described in Chapter 2 is the approximation of a process with a large process time constant as a near-integrator so the tuning for integrating processes can be used. The response can be considered to be a near-integrator when the process time constant is larger than four times the total loop dead time. This transition is consistent with the view that most of the response of PID tuned for disturbance rejection occurs in the four dead times. The transition also prevents the reset time in Lambda tuning from becoming more than four times the dead time.

The estimation of a secondary time constant can be important for tuning and loop analysis, particularly for integrating and runaway processes. For self-regulating processes, the secondary time constant is the second largest time constant in the loop. For an integrating process, it is also the second largest if you consider the integrating process gain can be equated to a large time constant in the near-integrator approximation. For a runaway process, the secondary time constant is also the second largest time constant because the largest time constant must be a positive feedback time constant for stability.

4.1.2 OVERVIEW

Chapter 4 starts with a methodology for estimating the dead time from process time constants in series. The equations presented work for time constants anywhere in the loop and are consequently useful for estimating the effect of sensor lags, transmitter damping, and signal filters. Since the focus in this chapter is on the process, a table provides a summary of the first principle relationships for the most common process dead times and process time constants. The remaining sections in this chapter compute the ultimate gain and ultimate period in terms of process dynamics.

The dynamic response of processes can be categorized based on internal feedback within the process. If the process has negative feedback, the process is self-regulating and will eventually reach a steady state if there are no more disturbances or subsequent changes in PID output. If the process has zero feedback, the process is integrating and will ramp for any unbalance in flows into and out of the volume. If the process has positive feedback, the process is runaway and can accelerate if left in manual long enough. In all cases, equipment limits, Safety Instrumented Systems, and relief devices may prevent the process from reaching a steady state or continuing to ramp or accelerate.

Simple equations enable the user to estimate the ultimate gain and period. For self-regulating processes, the equations use a primary process time constant (negative feedback), open loop gain, and total loop dead time. For integrating processes, the equations use the integrating process gain, secondary process time constant, and total loop dead time. For runaway processes, the equations use a primary process time constant (positive feedback), a secondary time constant, an open loop gain, and total loop dead time. Runaway processes are termed *open loop unstable*. Equation 4.17 can be used to convert a self-regulating or runaway process with a large primary process time constant to an equivalent near-integrating process. This capability enables all processes with a gradual response to be tuned by rules developed for integrating processes as discussed in Chapters 1 and 2.

While there are three major types of process responses, the self-regulating process is sub classified into three types based on the ratio of the total loop dead time to the largest time constant creating a total of five types of dynamic responses. Self-regulating processes with a time constant to dead time ratio greater than four are treated as near-integrating processes and can use the tuning rules for integrating processes. Self-regulating processes with a time constant to dead time ratio between one-fourth and four are termed in this book as moderate self-regulating processes that can use the tuning rules commonly seen in the literature. Self-regulating processes with a time constant to dead time ratio less than one-fourth are called dead time dominant processes that can take advantage of a smaller reset time.

The decrease in the PID gain required to eliminate excessive reaction to an inverse response (initial response is in the opposite direction of final response) will be covered in Chapter 9 (Effect of Nonlinearities). The special algorithm needed in feedforward control when inverse response is excessive will be addressed in Chapter 12 (Advanced Control).

4.1.3 RECOMMENDATIONS

1. To see the effect of process design on process dynamics, use table and equations in the chapter to estimate the process gain, process time constants, and process dead time.

2. To see the effect of dynamics on tuning and loop performance, use the equations to estimate the ultimate gain and period depending upon whether the process is self-regulating, integrating, or runaway.
3. Convert lag dominant self-regulating and runaway processes to integrating processes and use Equation 1.5d in Chapter 1 to prevent the start of oscillations from a reset time not being increased when a PID gain is decreased.
4. For near-integrating, integrating, and runaway processes, identify any significant secondary time constant (e.g., thermal lag from a heat transfer surface and thermowell for temperature control).
5. For runaway processes, make sure the PID gain is well above the limit for stability that is the inverse of the open loop gain as per Equation 4.22f. The open loop gain is the product of the manipulated variable gain, process gain detailed in Appendix F, and the measurement gain.
6. For runaway processes, make sure the secondary time constant and total loop dead time are each significantly less than the primary process time constant to prevent closing of the window of allowable PID gains causing the process to be unstable for all tuning settings. To safely stay far enough away from a closed PID gain window, the largest possible secondary time constant and total loop dead time should each be less than one-tenth of the smallest possible primary time constant.

4.2 EFFECT OF MECHANICAL DESIGN

Equipment and piping design determine the inherent performance of the process. The mechanical design often is based on specifications developed by process engineers. In general, changes after a project is completed are undesirable from both capital and installation cost and process interruption standpoint. A dialog early in a project between the automation, mechanical, and process design engineers based an understanding of the key relationships and common mistakes made can enable better process control leading to better process efficiency and capacity. Most advanced control systems are addressing limitations in the mechanical and process design.

4.2.1 EQUIPMENT AND PIPING DYNAMICS

The peak and integrated errors for unmeasured disturbances at the process input (load upsets) are minimized by minimizing the total loop dead time and maximizing the primary time constant in the process and then tuning the PID to take full advantage of these favorable dynamics (see Chapter 3 and Appendix C). A large process time constant slows down the disturbance giving more time for the PID to do its job. The term *primary time constant* denotes the time constant is the largest in the loop. In general, the larger process time constants are associated with liquid volumes. The second largest time constant (secondary time constant) is often the result of thermal lags and volumes in series. The process dead times are typically due to mixing and transportation delays. Table 4.2 provides the equations and range of values for the more common types of equipment and piping in the process industry. The process dead time is essentially zero for gas pressure control of single volume. If there were multiple volumes in series, there would be significant dead time as noted in the example given in Section 4.6.4 and Appendix G.

The dead time and time constants for concentration and temperature control of gas volumes commonly occurring in refineries and petrochemical processes are not included in the table. In general, most of the residence time (volume/feed rate) for these gas volumes becomes a transportation delay rather than a time constant, which is the opposite of agitated liquid volumes. The transportation delay is due to plug flow (lack of back mixing). The largest time constant in these unit operations originates in the heat transfer surface or the sensor (e.g., thermowell lag) which slows down the correction or measurement of the load upset.

Note that there are other significant sources of dead times and time constants as seen in Figure 2.4, the block diagram of a control loop with valve or VSD, measurement, and controller dynamics denoted besides process dynamics. Improvements in automation system dynamics are generally less expensive, faster, and less disruptive. However, sources of dead time that are small compared to those in the process do not have much of an impact on the bottom line.

For inline systems and gas volumes, most of the dead time is in the automation system. This is also true for vessels with high agitation rates, negligible injection delays, and minimal heat transfer surface lags. Here the automation engineer has the keys to the ultimate limit to loop performance detailed in Appendix C.

A lot can go wrong in mechanical design that can make the job difficult or impossible. The next section offers a list of the more common equipment and piping mistakes with the consequences noted.

The parameters in the Table 4.2 equations are often a function of operating conditions and are not generally precisely known. The value of Table 4.2 is more in showing the relative effect of these parameters on process dynamics. For example, the user can estimate the benefit of a higher agitation rate in terms of decreasing dead time and a larger well mixed volume in terms of increasing the primary time constant. The user can also employ these equations in virtual plants to prototype improvements and provide operator training. The actual tuning settings in the plant should be based on field tests.

Process Nomenclature:

A_0 = cross sectional area of liquid level (m^2)
A = overall heat transfer surface area (m^2)
A_s = shell heat transfer surface area (m^2)
A_t = tube heat transfer surface area (m^2)
A_w = wall heat transfer surface area (m^2)
C_p = process heat capacity (kJ/kg*°C)
C_w = wall heat capacity of heat transfer surface (kJ/kg*°C)
C_s = shell heat capacity of heat transfer surface (kJ/kg*°C)
C_t = tube heat capacity of heat transfer surface (kJ/kg*°C)
f_p = friction factor for pipeline (0.1 to 1.0)
F_a = agitator pumping rate (kg/sec)
F_f = total feed flow (kg/sec)
F_g = gas flow (kg/sec)
F_i = feed stream i flow (kg/sec)
F_o = vessel outlet flow (kg/sec)
F_p = pump flow (kg/sec)
F_r = recirculation flow (kg/sec)
F_v = vaporization rate (kg/sec)
H_v = heat of vaporization (kJ/kg)
H_x = heat of reaction (kJ/kg)

L_p = pipeline length (m)
M_g = gas mass (kg)
M_o = liquid mass (kg)
M_w = wall mass of heat transfer surface (kg)
M_t = tube mass of heat transfer surface (kg)
MW_g = molecular weight of gas (kg per mole)
N_t = number of tubes in heat exchanger
P_g = gas pressure (kPa)
T_c = coolant temperature (°C)
T_f = total feed temperature (°C)
T_g = gas temperature (°K)
T_j = jacket temperature (°C)
T_i = feed stream i temperature (°C)
T_o = outlet temperature (°C)
t = time (sec)
Q_r = heat from reaction (kJ)
R = ratio of valve pressure drop to system pressure drop (0.1 to 1.0)
R_x = reaction rate (kg/sec)
R_g = universal constant for gas law (8.3145 (kPa * m^3)/(°K * kg per mole))
ρ_g = gas density (kg/m^3)
ρ_i = stream i density (kg/m^3)
ρ_o = overall density (kg/m^3)
ρ_p = pump flow density (kg/m^3)
ρ_v = density of vapor (kg/m^3)
U = overall heat transfer coefficient (kJ/m^2 * °C)
v_f = fluid velocity in pipeline (m/sec)
v_s = velocity of sound in fluid (m/sec)
V_g = gas volume (m^3)
V_i = injection (e.g., dip tube or sparger ring) volume (m^3)
V_p = piping volume (m^3)
x = fraction of volume that is plug flow
Z_g = compressibility factor for nonideal gas (0.1 to 1.0)

Dynamics Nomenclature:

K_i = open loop integrating process gain (1/sec)
K_p = process gain (e.u./e.u.) (includes flow ratio gain)
θ_p = process dead time (sec)
θ_{pi} = injection process dead time (sec)
θ_{pm} = mixing process dead time (sec)
θ_{px} = transport process dead time (sec)
τ_p = primary process time constant (sec)
τ_{p2} = secondary process time constant (sec)
τ'_p = runaway process time constant (positive feedback) (sec)
τ_t = tube fluid time constant (heat exchanger) (sec)
τ_w = tube wall constant (heat exchanger) (sec)
τ_s = shell fluid constant (heat exchanger) (sec)

Table 4.2. Equations for equipment and piping dynamics

Equipment or piping	Controlled variable	Manipulated variable	Process dynamics from first principle equations (typical range in manufacturing plants)
Agitated vessel (well mixed)	Concentration (% mass)	Flow (kg/sec)	$\theta_{pi} = V_i / (F_i / \rho_i)$ (1–10 sec)(note 1) $\theta_{pm} = 0.5 * (M_o / \rho_o) / \left[(F_f + F_a + F_r) / \rho_o + F_v / \rho_v \right]$ (6–60 sec) $\tau_p = M_o / (R_x + F_f) - \theta_p$ (600–6,000 sec) Self-regulating: $K_p = 100\% * X_{Af} / (R_x + F_f)$ (0.0001–10.0 %/kg/sec)(note 2) Integrating: $K_p = 100\% * X_{Af} / M_o$ (0.00000001–0.01 %/kg)
Static mixer	Concentration (% mass)	Flow (kg/sec)	$\theta_{pi} = V_i / (F_i / \rho_i)$ (2–20 sec) $\theta_{pm} = x * V_p / \Sigma(F_i / \rho_i)$ (1–2 sec) $\theta_{px} = L_p * V_p / \Sigma(F_i / \rho_i)$ (2–12 sec) $\tau_p = (1 - x) * V_p / \Sigma(F_i / \rho_i)$ (0.2–0.5 sec)
Liquid pipeline	Flow (kg/sec)	Pump flow (kg/sec)	$\theta_p = L_p / v_s$ (0.001–1 sec)(note 3) $\tau_p = \dfrac{L_p * A_p * f_p^2}{F_p / \rho_p}$ (0.1–1 sec) $K_p = 1$
Liquid pipeline	Flow (kg/sec)	Valve flow (kg/sec)	$\theta_p = L_p / v_s$ (0.001–1 sec)(note 3) $\tau_p = \dfrac{D_p}{4 * f_p * v_f * (1 + R)}$ (0.1–1 sec) $K_p = 1$

(*Continued*)

Table 4.2. (*Continued*)

Equipment or piping	**Controlled variable**	**Manipulated variable**	**Process dynamics from first principle equations (typical range in manufacturing plants)**
Liquid heat exchanger	Tube temperature (°C)	Shell temperature (°C)	$\tau_t = \dfrac{M_t * \left(C_t / N_t\right)}{U_t * \left(A_t / N_t\right) + F_t * C_t}$ (6–60 sec) $\tau_w = \dfrac{M_w * C_w}{U_t * A_t + U_s * A_s}$ (1–6 sec) $\tau_s = \dfrac{M_s * \left(C_s / N_t\right)}{U_s * \left(A_s / N_t\right) + F_s * C_s}$ (2–20 sec) $K_P = (F_s * C_s)/(F_t * C_t)$ (1–10)
Liquid level	Level (%)	Flow (kg/sec)	$\theta_p = L_p/v_s$ (0.001–1 sec)(note 3) secondary time constant for manipulating pump flow $\tau_{p2} = \dfrac{L_p * A_p * f_p^2}{F_p / \rho_p}$ (0.1–1 sec) secondary time constant for manipulating valve flow $\tau_{p2} = \dfrac{D_p}{4 * f_p * v_f * (1+R)}$ (0.1–1 sec) $K_p = 1/(\rho_o * A_o)$ (0.00000001–0.01 m per kg)
Gas pressure	Pressure kPa	Flow (kg/sec)	$K_p = \left[Z_g * (R_g * T_g) / \left(MW_g * V_g\right)\right]$ (1–10,000 kPa/kg)
Agitated vessel (well mixed)	Vessel temperature (°C)	Jacket temperature (°C)	$\tau_p = (C_p * M_o) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A\right]$ (600–6,000 sec)

(*Continued*)

Table 4.2. (*Continued*)

Equipment or piping	Controlled variable	Manipulated variable	Process dynamics from first principle equations (typical range in manufacturing plants)
			secondary time constant from heat transfer surface $\tau_{p2} = (C_w * M_w) / [U * A]$ (6–60 sec) $K_p = (U * A) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right]$ (0.1–10) Integrating: $K_p = (U * A) / (C_p * M_o)$ (0.00000001–0.01 per sec)
Vessel jacket	Jacket temperature (°C)	Coolant flow (kg/sec)	$\theta_p = V_j / (F_j / \rho_j)$ (1–5 sec) $K_p = \left(T_j - T_c\right) / F_j$ (0.1–10 °C/kg/sec)
Agitated vessel (well mixed)	Vessel temperature (°C)	Feed flow (kg/sec)	$\theta_{pi} = V_i / (F_i / \rho_i)$ (1–10 sec)(note 1) $\theta_{pm} = 0.5 * (M_o / \rho_o) / \left[(F_f + F_a + F_r) / \rho_o + F_v / \rho_v \right]$ (6–60 sec) $\tau_p = M_o / F_f - \theta_p$ (600–6,000 sec) $K_p = \left(T_o - T_f\right) / F_f$ (0.1–10 °C/kg/sec)

Notes:
[1]For pH control the injection delay (θ_{p1}) in dip tubes can be as large as 6,000 seconds for low reagent flow.
[2]For pH control the *overall* process gain (K_p) is extremely low in terms of percent H^+ ion concentration (10^{-5} % H^+/kg/sec) but the transformation to pH creates extremely high *localized* process gains as seen on titration curve near the isopotential point (e.g., pH 7).
[3]For long pipelines (miles rather than feet) the dead time can be orders of magnitude larger (pipeline length divided by speed of sound in fluid).

4.2.2 COMMON EQUIPMENT AND PIPING DESIGN MISTAKES

Here is a compilation with the help of Solutia retiree and Fellow Hall of Famer Terry Tolliver and ISA Mentor Program participants Hector Torres and Hunter Vegas.

1. *Boiler drum that is too small for the steam production rate*. Shrink and swell can be so severe for changes in steam demand that the boiler may trip on low or high level. The normal three-element control that is enforcing the material balance by feedforward

where the feedwater flow is ratioed to steam flow can make the initial level swing worse. A kicker algorithm creates a transient feedforward of the opposite sign as the normal three-element drum feedforward. For an increase in firing rate for an increase steam demand there is an immediate temporary step decrease in feedwater flow to help mitigate the swell. The increase decays out as rapidly as the swell. The conventional feedforward can step in once the swell is under control. The swell develops from an increase in bubbles in the downcomers from the increase in firing rate that pushes level up into the drum. Note that differential pressure level transmitters see the increases in head from the liquid but not the bubbles pushed up into the drum.

2. *Gas pressure control volume that is too small for gas flow rate.* Rapid increases in gas flow can activate pressure relief devices. An extreme example was an incinerator where increases in waste gases could cause the pressure to ramp off-scale within a tenth of a second. In this case, the 0.1 second scan time of a Distributed Control Systems (DCS) PID was too slow and we needed to revert back to the analog electronic controller. The fan had a VSD with no dead time or dead band. The transmitter damping was minimized.
3. *Temperature or pH control in a pipeline with a long distance between the sensor and control valve*. The transportation delay is excessive. Feedforward and dead time dominant tuning can help but the real solution is to move the sensor closer.
4. *A control valve that is too big or a butterfly is used to control low flow.* The valve ends up riding the seat. The equal percent characteristic is steepest and the sealing friction is greatest near the closed position. Plus the actuator torque must build up to be greater than the breakaway torque to open the valve. The result is a limit cycle near the closed position and on-off control that wears out the valve and increases system variability. This is a common problem for pH control because of the extreme rangeability requirement. The solution is a single correctly sized sliding stem valve with a smart positioner. If rangeability is needed, a combination of split range and valve position control (VPC) can be used. When demand is high, the splitter block output to increase large valve position is added as a feedforward signal to the VPC. Conventional flow feedforward helps the VPC deal with upsets.
5. *Large oversized blender discharge valves.* They affect upstream and downstream feed systems. If a PID is controlling the blender's level it will react abruptly to every discharge and to the sudden large decrease in level. The resulting sudden step changes to the upstream feeding system and subsequent decay in rate demand could affect raw material ratios and concentrations, especially if the associated feeders control loops are not properly tuned. Downstream the large amount of material being discharged will create stagnation zones and blocking when the discharge chute and vibrating conveyor are not large enough. Tuning of the level controller in the receiving hopper will be a hard task as stagnation and blocking will introduce constant load upsets varying in period and amplitude. Aggressive tuning leads to overfill of the hopper and high torque of the associated agitator; when the tuning goes to the other extreme, the material will travel slowly thru the conveyor leading to the mentioned blocking in the discharge chute or conveyor. While scheduled and adaptive tuning may help, the solution if the process material permits is the replacement of the slide-gate with a properly sized rotary airlock valve (rotary airlock feeder) to provide a modulated continuous flow.
6. *Control valve upstream of flow sensor.* The control valve distorts the flow profile and can cause bubble formation from the flashing in the vena contracta. A Coriolis meter

can deal with the irregular profile and some degree of bubbles in special models but the real solution is to have the flow sensor upstream of the control valve.

7. *Steam desuperheater temperature sensor too close to condensate injection.* The steam has water droplets and an inconsistent temperature. The result is noisy measurement. A signal filter can help but the real solution is to move the sensor downstream enough where all of the liquid is vaporized and the cross sectional temperature profile is more constant. Just do not get carried away and have it located so far that the transportation delay is greater than 10 seconds.
8. *pH sensor too close to injection of reagent for inline pH control.* Inline system has no back mixing and even the radial mixing is adversely affected at the discharge of the mixer. pH is an extremely sensitive measurement, thus small fluctuations in concentration will show up as high amplitude noise. Flashing reagents such as ammonia will require more time for the bubbles to dissolve. A signal filter can help but the real solution is to move the sensor downstream enough where the fluid is better mixed from turbulent flow and the cross sectional pH profile is more constant. Just do not get carried away and have it located so far that the transportation delay is greater than 10 seconds.
9. *Improper location of control tray in distillation column.* Temperature is an inference of composition in a column. The proper tray is the one that shows the largest temperature change in both directions for an increase and decrease in the ratio of the manipulated flow to feed flow (e.g., reflux flow to feed flow). Process simulators can point to the best candidate. Temperature sensors should be installed in the recommended and adjacent trays and online tests done to confirm best tray location.
10. *Radial rather than axial mixing*. The concentration, pH, and temperature response is erratic with a variable process dead time due to poor mixing. There is a dead zone and a nonuniform mixture in upper half of the vessel from agitation pattern that is swirling liquid near the bottom, rather pulling liquid from and pushing liquid to the surface. If the surface is calm, there is insufficient axial agitation.
11. *Horizontal neutralization vessel or sump.* The concentration, pH, and temperature response is erratic with an extremely large and variable process dead time due to poor mixing. There are large dead zones and nonuniform mixtures outside the cylindrical agitation zone established by agitator blade diameter.
12. *Feed entry point near vessel discharge nozzle*. Reagents, reactants, and additives bypass to a variable degree the main volume and exit a continuous vessel. This short circuiting is inconsistent and leads to poor product quality and yield.
13. *Horizontal distillate receiver*. The cross sectional area is so large that the level change for a change in flow is less than sensor noise or sensitivity limit. For the column control scheme where tray temperature is controlled by manipulating distillate flow, the tray temperature control has no effect on the column until the receiver level controller sees a change in level and causes a change in reflux flow. Also, the level controller gain needed is way beyond the comfort zone and will amplify noise from a poor signal to noise ratio due to true change in level being small relative to sensor noise and sensitivity.
14. *Low flow causing low heat transfer coefficient and fouling.* The heat transfer coefficient is proportional to liquid velocity to the 0.8 power. Fouling greatly increases for velocities less than five feet per second. Fouling causes a dramatic decrease in the heat transfer coefficient.

15. *Low flow causing low pH sensor response and fouling.* The sensor lag can increase by an order of magnitude for velocities less than 0.5 feet per second. Fouling greatly increases for velocities less than five feet per second. Fouling causes an even more dramatic increase in the sensor lag (two more orders of magnitude).
16. *Variable jacket or coil flow.* The process gain, transportation delay, and valve stiction increase as the jacket or coil flow decreases. Operation at low flow can lead to oscillations that you cannot get rid of by just tuning. Solution is a recirculation loop.
17. *Steam to water transition versus steam injector.* When split range control makes a transition from cooling to heating and vice versa, water droplets or pockets exist in the steam and steam bubbles or pockets exist in cooling water leading to erratic heat transfer. A steam injector that changes cool water to hot water eliminates the multiple phases in the jacket or coil enabling a smooth transition between heating and cooling.
18. *Insufficient surge tank volume.* If the change in flow into a surge tank requires an almost equal change in outlet flow to prevent activating high or low level alarms, most of flow variability that enters the tank is transferred as flow variability out of the tank. Often this flow variability become feed upsets to important downstream unit operations for separation, purification, and final product processing.
19. *Variable speed pump head not high enough at high static head.* As the pump curve approaches the static head downstream (e.g., vessel pressure) as the pump flow decreases, the flow takes a dive and responds more to downstream pressure fluctuations than speed. The slope of the installed flow characteristic becomes steep, resembling that of a quick opening control valve. The result is a loss of controllability and rangeability.
20. *Variable speed pump head is not high enough at high flow.* If the pump curve takes a nose dive at high flow, the change in flow with speed is small. The flat installed flow characteristic is similar to what is seen for butterfly valves operating more than 40 degrees open or in systems where the valve to system pressure drop ratio is too low.
21. *Control valve to system pressure drop ratio is too low.* For valves with linear trim, the flow characteristic becomes quick opening. For equal percentage trims, the minimum flow becomes much larger and the curve becomes flatter much sooner. The loss of rangeability is great and aggravated by the increase in stiction and flow near the closed position.
22. *Sparger ring is near electrode.* Bubbles from the sparger momentarily touch and even attach to the electrode surface causing an erroneous measurement. The frequency of resulting noise may not be consistent making the judicious use of a filter more difficult.
23. *Axial compressor shaft inertia is too low.* The unloading of the compressor upon surge causes the speed to accelerate. The higher speed causes more unloading which causes more acceleration resulting in a runaway response and high speed shutdown.
24. *Exothermic reactor cooling system capability is too low.* An increase in reactor temperature causes the reaction rate to increase. The increase in reaction rate causes an increase in heat of reaction causing an acceleration resulting in a runaway response and high temperature or high pressure shutdown.
25. *Biological reactor shift to product formation rate is too late***.** If the substrate and nutrients are being used just for cell reproduction, the increase in cell concentration causes an increase in cell production. The increase in growth rate accelerates until the cells start to die due to overcrowding and the buildup of sodium and lactate.

26. *Heat exchanger in recirculation line with low heat transfer area or coefficient***.** If the heat exchanger does not have enough muscle to quickly change the temperature, the recycle effect of vessel contents can become noticeable in the heat exchanger temperature controller response, drastically changing the dynamics and tuning rules from self-regulating to integrating with a significant secondary time constant from thermal lag.
27. *Dip tubes too large for low reagent flow***.** The injection delay (volume/flow) becomes huge, approaching an hour or more to start or stop a reagent flow. Before the reagent flow is started, the dip tube is back filled with process fluid from the vessel. Reagent flow will not get into the vessel until the process fluid in the dip tube is flushed out. When the reagent valve closes the reagent flow entering the vessel does not stop until the dip tube is completely empty of reagent.
28. *Gravity feed of reagents.* The pressure available to the control valve varies with the head in the equipment that is the source of reagent flow. A change in control valve position does not immediately result in a corresponding change of reagent flow coming out of the dip tube because the pipe and dip tube are not completely full or pressurized. A delay for a change in flow depends upon the time a wave traveling down the pipeline takes to reach the dip tube and then how long a change in falling film takes to reach surface level in the dip tube. The resulting dead time is large and variable.
29. *Solid or gaseous reagents.* The time it takes for solids or bubbles to dissolve varies with solid and bubble size and operating conditions. Often the time to completely dissolve is longer than the residence time, resulting in reagent solids or bubbles downstream of the neutralizer and a continuing shift in pH. The result is erratic pH control and excessive reagent use and in some cases additional maintenance to deal with the consequences of solids (fouling) and bubbles (erosion).
30. *Several volumes between valve and sensor.* The residence times in series become either pure dead time (plug flow volumes) or equivalent dead time from time constants in series (back mixed volumes). This dead time increases as the feed rate decreases.

4.3 ESTIMATION OF TOTAL DEAD TIME

The total loop dead time is the sum of all pure dead times plus the equivalent dead time from time constants in series. For a FOPDT approximation for a self-regulating process, a fraction of all time constants smaller than the largest time constant (primary time constant) is computed as an equivalent dead time. The part of the time constant not converted to dead time is added to the primary time constant. For a second order plus dead time (SOPDT) approximation for a self-regulating process or for an integrating or runaway process, the conversion to equivalent dead time is done for all time constants smaller than the secondary time constant. The part of these time constants not converted to dead time is added to the secondary time constant.

The equivalent dead time from a small time constant in series with the primary or secondary time constant can be computed by Equations 4.1a through 4.1c (Ziegler and Nichols 1943). Figure 4.1 shows nearly all of a small time constant is converted to dead time if the respective time constant is very small compared to the large time constant.

$$X_i = \frac{\tau_i}{\tau_x} \tag{4.1a}$$

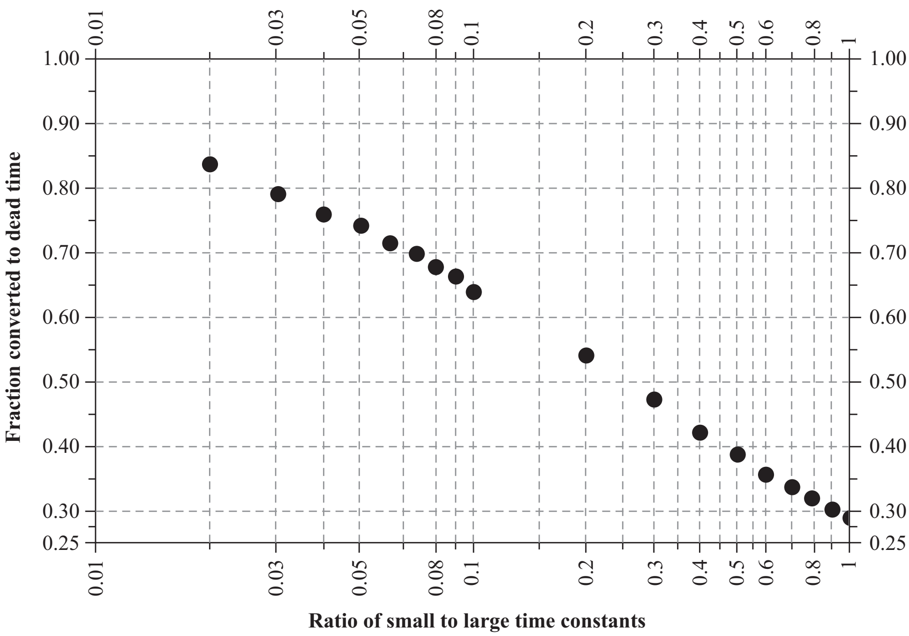


Figure 4.1. Fraction of small time constant converted to dead time.

$$Z_i = \frac{1}{1 - X_i} \tag{4.1b}$$

$$Y_i = 1 + \frac{1}{X_i} - \frac{1}{X_i^{Z_i}} - \frac{\ln(X_i)}{1 - X_i} \tag{4.1c}$$

$$\theta_i = \sum_i Y_i * \tau_i \tag{4.1d}$$

The equivalent dead time from a multiple equal large time constant in series with the primary or secondary time constant can be computed by Equations 4.2a and 4.2b (Ziegler and Nichols 1943). Figure 4.2 shows nearly all of the large equal time constants are converted to dead time if the number of the respective time constant is very large.

$$Y_j = A_o + A_1 * [\ln(N)]^1 + A_2 * [\ln(N)]^2 + A_3 * [\ln(N)]^3 + A_4 * [\ln(N)]^4 \tag{4.2a}$$

$$\theta_j = Y_j * N * \tau_j \tag{4.2b}$$

Equation 4.3a details that the total dead time is the sum of the equivalent dead time from small time constants and multiple equal time constant plus the pure dead times. Equation 4.3b shows the increase in the open loop time constant from the fraction of the smaller time constants not converted to dead time. If the equal time constants are the largest in the loop, then the open loop time constant is the leftover fraction. While the equations are shown in terms of an open loop time constant (primary time constant), the procedure works for a secondary time constant as well. Hopefully the largest time constant is in the process in terms of disturbance rejection but the procedure is still valid no matter where the largest time constant is located. If

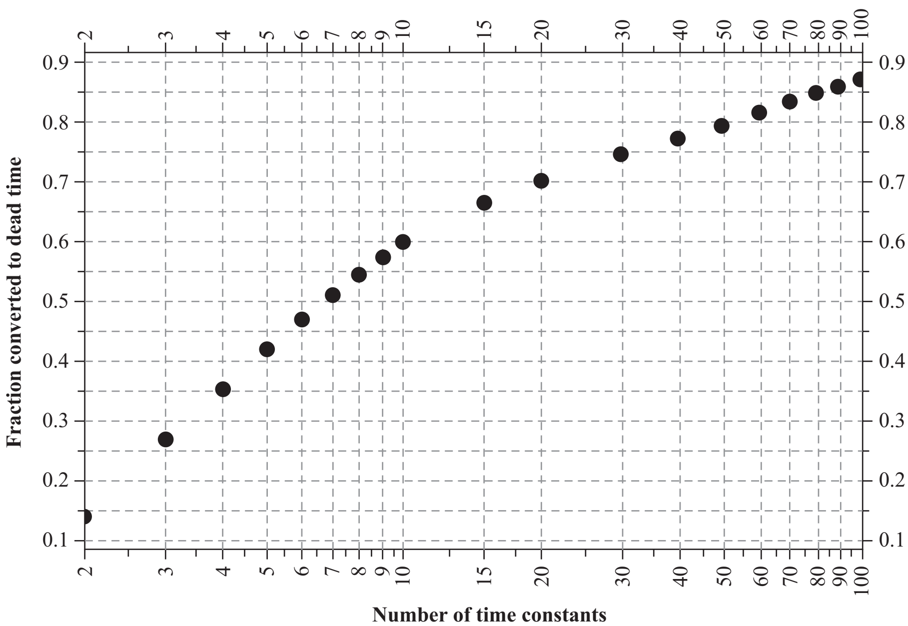


Figure 4.2. Fraction of multiple equal time constants converted to dead time.

the largest time constant is in the measurement, the tuning and observed errors in percent for the PID are the same but the actual error in the process variable is larger due to the filtering effect as discussed in Chapters 5 and 6.

$$\theta_o = \theta_i + \theta_j + \sum \theta_k \tag{4.3a}$$

$$\tau_o = \sum_i \left[(1 - Y_i) * \tau_i\right] + \sum_j \left[(1 - Y_j) * N * \tau_j\right] + \tau_x \tag{4.3b}$$

where

θ_i = dead time from i small time constants (sec)
θ_j = dead time from j equal large time constants (sec)
θ_k = pure time delays k (sec)
θ_o = total open loop dead time (sec)
τ_i = small time constants i (sec)
τ_j = large equal time constants j (sec)
τ_o = open loop time constant (primary time constant) (sec)
τ_x = maximum time constant (sec)
X_i = ratio of small time constant i to maximum time constant (sec)
Z_i = exponent in equation for conversion of small time constant i to dead time
Y_i = fraction of small time constant i converted to dead time (sec)
Y_j = fraction of summation of large time constants j converted to dead time (sec)

4.4 ESTIMATION OF OPEN LOOP GAIN

The percent change in the PID process variable for a given percent change in the PID output with the PID in manual (MAN), remote output (ROUT) or local override (LO) mode is the open loop gain. By not having feedback action we see the effect of automation system dynamics (gains, dead times, and time constants) without the effect of controller tuning. The open loop gain takes into account the valve or VSD installed flow characteristic and the measurement calibration span.

$$K_o = K_v * K_r * K_p * K_m \tag{4.4a}$$

$$K_v = \frac{\Delta F_v}{\Delta \% CO} \tag{4.4b}$$

For composition, temperature, and pH loops:

$$K_r = \frac{\Delta R}{\Delta F_v} = \frac{\Delta F_v / F_p}{\Delta F_v} = \frac{1}{F_p} \tag{4.4c}$$

$$K_p = \frac{\Delta PV}{\Delta R} \tag{4.4d}$$

For flow, level, and pressure loops:

$$K_r = 1 \tag{4.4e}$$

$$K_p = \frac{\Delta PV}{\Delta F_v} \tag{4.4f}$$

$$K_m = \frac{100\%}{(PV_{\max} - PV_{\min})} \tag{4.4g}$$

For integrating processes, the open loop gain (K_o) is replaced with an integrating process gain (K_i) with inverse time units (1/sec) that results from the process gain term units not cancelling out the flow time units as seen in Equation 4.18c for level.
where

K_m = measurement gain (%/PV e.u.)
K_i = integrating process open loop gain (%/%/sec) (1/sec)
K_o = self-regulating process open loop gain (%/%) (dimensionless)
K_p = process gain (PV e.u.)
K_r = flow ratio gain often imbedded in process gain (1/flow e.u.)
K_v = valve or VSD gain (manipulated flow per % output)
$\Delta\%CO$ = change in controller output converted to percent of controller output scale (%)
ΔF_v = change in valve or VSD flow (flow e.u.)
F_p = process flow at current production rate (flow e.u.)
$\%PV$ = change in process variable converted to percent of controller input scale (%)
$\%PV_{\max}$ = maximum process variable scale value (PV e.u.)
$\%PV_{\min}$ = minimum process variable scale value (PV e.u.)
ΔR = change in ratio of manipulated flow to process flow (dimensionless)

4.5 MAJOR TYPES OF PROCESS RESPONSES

The sign of internal feedback with the process has a major impact on the type of process response and the relative importance of the role of the PID in getting the process to line out.

Negative feedback is needed for a process to settle out at an operating point. More aggressive PID tuning is needed for tight control as you go from negative to zero to positive feedback within the process.

Self-regulating processes have negative feedback within the process and will reach a steady state for a given disturbance without PID control. Integrating processes have zero feedback within the process and will continue to ramp for a given disturbance without PID control. Integrating processes are sometimes called non-self-regulating processes. Runaway processes have positive feedback within the process and will accelerate away from the operating point for a given disturbance without PID control. Runaway processes are called open loop unstable.

The rate of change of the open loop response of a process with negative, zero, and positive internal feedback will eventually decelerate, stay the same, and accelerate as the deviation of the process variable increases, respectively. Here we look at the impact of process type on ultimate period and ultimate gain. Even though one is not using Ziegler Nichols tuning, these parameters are benchmarks as to closed loop control dynamics and control loop performance.

Self-regulating processes with large primary time constants have no or little deceleration ramp within the time frame of the PID response (four dead times). These processes are termed near-integrating and are best treated as integrating processes in terms of analysis and PID tuning.

4.5.1 SELF-REGULATING PROCESSES

The most common type of process is the self-regulating process. Liquid flow and pressure control, continuous unit operation concentration, pH, and temperature control have self-regulating response. For large liquid volumes with mixing (e.g., continuous stirred reactors and distillation columns), the primary time constant is so large that the response is near-integrating.

Figure 4.3a for the ultimate period indicates a curvature to lower ultimate periods as the primary time constant to dead time ratio decreases. The period goes from four times the dead time to two times the dead time as this ratio decreases. Figure 4.3b for the ultimate gain shows a nearly linear relationship with this ratio. These plots for the ultimate period and ultimate gain

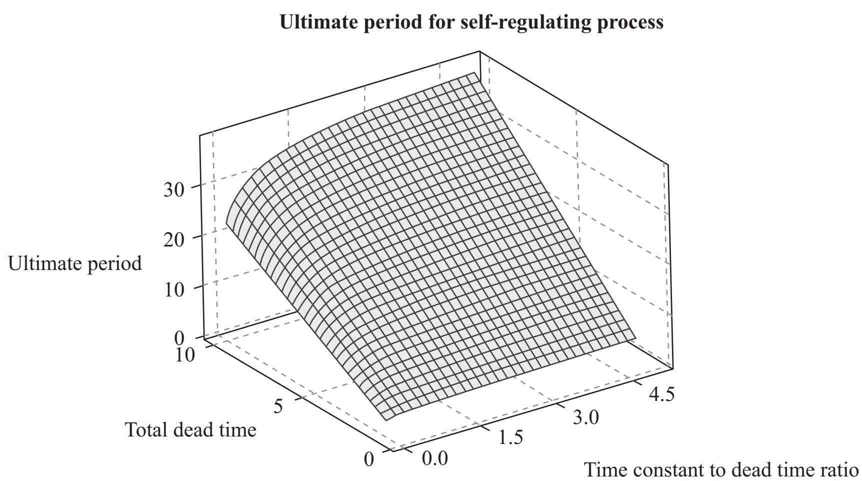


Figure 4.3a. Ultimate period for self-regulating process.

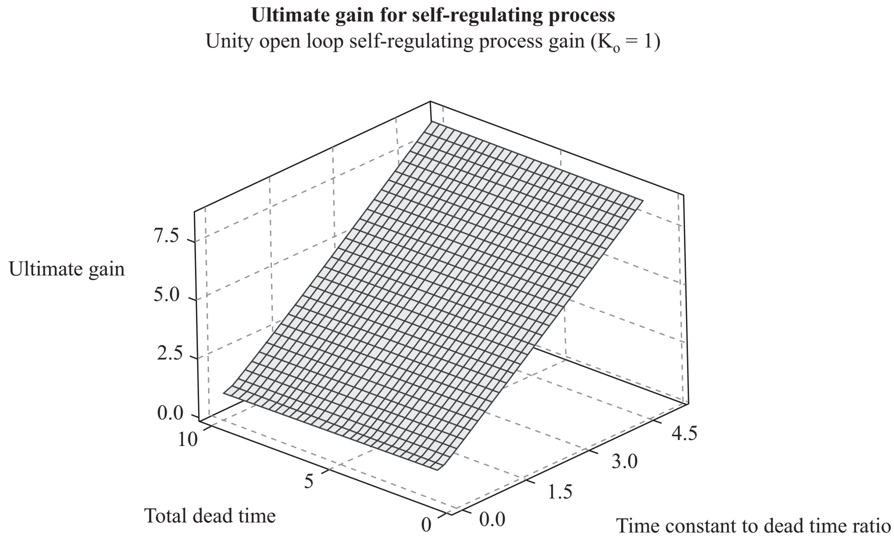


Figure 4.3b. Ultimate gain for self-regulating process (unity open loop gain).

were generated from the equations developed here for self-regulating processes. The accuracy of the curve fit for the ultimate period is about 20 percent. The value is more in showing relationships than computing actual values. An auto tuner should be used to find the actual ultimate period and ultimate gain.

For self-regulating (negative feedback) processes based on a curve fit for Bode plots:

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o} \right)^{0.65} \right] * \theta_o \tag{4.5a}$$

For loop dominated by a large time constant ($\tau_0 >> \theta_0$), the ultimate period equation simplifies to:

$$T_u = 4 * \theta_o \tag{4.5b}$$

For loop dominated by a large dead time ($\theta_0 >> \tau_0$), the ultimate period equation simplifies to:

$$T_u = 2 * \theta_o \tag{4.5c}$$

The effect on tuning can be most directly seen in the Ziegler Nichols ultimate oscillation method for an ISA Standard form proportional-integral (PI) controller:

$$K_c = 0.4 * K_u \tag{4.6}$$

$$T_i = 0.8 * \frac{T_u}{Min(4,\ 10 * \left(\frac{4 * \theta_o}{T_u} - 1 \right)^2 + 1)} \tag{4.7}$$

For an ISA Standard form PID controller:

$$K_c = 0.6 * K_u \tag{4.8}$$

$$T_i = 0.6 * \frac{T_u}{Min(4, 10 * \left(\frac{4 * \theta_o}{T_u} - 1\right)^2 + 1)} \tag{4.9}$$

$$T_d = Min\left[0.25 * T_i, 0.8 * Max\left[0.0, 0.25 * (T_i - 0.5 * \theta_o) + \tau_s\right]\right] \tag{4.10}$$

In the process industry, the gain setting is cut in half as noted in Chapter 1 to provide a smooth nonoscillatory response and to increase the robustness of the tuning.

Starting with fundamental relationship that ultimate gain is the inverse of the product of the open loop gain and amplitude ratio at −180 degrees phase shift

$$K_u = \frac{1}{K_0 * AR_{-180}} \tag{4.11a}$$

For self-regulating (negative feedback) processes:

$$AR_{-180} = \frac{1}{\sqrt{1 + (\tau_o * \omega_n)^2}} \tag{4.12a}$$

$$K_u = \frac{\sqrt{1 + (\tau_o * \omega_n)^2}}{K_o} \tag{4.12b}$$

Using natural frequency relationship to ultimate period $\omega_n = \frac{2 * \pi}{T_u}$

$$K_u = \frac{\sqrt{1 + (\tau_o * 2 * \pi / T_u)^2}}{K_o} \tag{4.12c}$$

For loop dominated by a large time constant ($\tau_o >> \theta_o \Leftrightarrow T_u << \tau_o$), the ultimate gain equation simplifies to:

$$K_u = \frac{2 * \pi * \tau_o}{K_o * T_u} \tag{4.12d}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$K_u = \frac{1.6 * \tau_o}{K_o * \theta_o} \tag{4.12e}$$

For loop dominated by a large dead time ($\theta_o >> \tau_o \Leftrightarrow T_u >> \tau_o$), the ultimate gain equation simplifies to:

$$K_u = \frac{1}{K_o} \tag{4.12f}$$

For PID control of processes dominated by a large time constant (the case of more prominent interest), we have using Equation 3.4b ($K_c = 0.6 * K_u$ and $T_i = 0.6 * T_u$):

$$E_i = \frac{0.6 * T_u}{0.6 * K_u * K_o} * E_o = \frac{T_u}{K_u * K_o} * E_o \tag{4.13a}$$

Knowledge of K_o to calculate the integrated error is not needed since K_u is inversely proportional to K_o.

For a loop dominated by a large time constant ($\tau_o >> \theta_o \Leftrightarrow T_u << \tau_o$), based on Equations 3.4b and 3.10b, and aggressive tuning as per Equations 4.8 and 4.9 the minimum integrated and peak errors determined by dynamics for PID control are:

$$E_i = \frac{T_u}{2 * \pi * \tau_o} * T_u * E_o \tag{4.13b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_i = 2.6 * \frac{\theta_o^{\,2}}{\tau_o} * E_o \tag{4.13c}$$

For PID control of processes dominated by a large time constant (the case of more prominent interest), we have using Equation 3.10b ($K_c = 0.6 * K_u$):

$$E_x = \frac{1.1}{0.6 * K_u * K_o} * E_o = \frac{1.8}{K_u * K_o} * E_o \tag{4.14a}$$

Knowledge of K_o to calculate the peak error is not needed since K_u is inversely proportional to K_o.

For the same conditions we stated for Equations 4.13b, the peak error simplifies to:

$$E_x = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o \tag{4.14b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_x = 1.2 * \frac{\theta_o}{\tau_o} * E_o \tag{4.14c}$$

where

AR_{-180} = amplitude ratio at −180 degrees phase shift (dimensionless)
E_i = integrated error from load disturbance (e.u.* sec)
E_o = open loop error (load disturbance error with controller in manual) (e.u.)
E_x = peak error from load disturbance (e.u.)
K_c = controller gain for maximum disturbance rejection (dimensionless)
K_o = self-regulating process open loop gain (dimensionless)
K_u = controller ultimate gain (dimensionless)
T_i = integral time (reset time) (sec)
T_d = derivative time (rate time) (sec)
T_u = ultimate oscillation period (sec)
θ_o = total loop dead time (sec)
τ_o = open loop time constant in self-regulating process (sec) (negative feedback)
ω_i = natural frequency (critical frequency) (radians/sec)

4.5.2 INTEGRATING PROCESSES

Processes where changes in the process variable have no effect on the rate of change of the process variable due to a disturbances or setpoint changes have zero internal feedback and have an integrating process response.

A change in level in a tank does not change the discharge flow unless there is gravity flow. A change in gas pressure in a vessel does not change the vent flow unless the vent valve pressure drop is very small. The discharge and vent flow depend only upon valve position and fluid density.

Unit operations where there is no discharge flow will have an integrating response for concentration, pH, and temperature. During batch operations and the startup of continuous unit operations, the discharge valve is closed giving an integrating response.

Figure 4.4a for the ultimate period indicates a large increase as the secondary time constant to dead time ratio increases. The period goes from four times the dead time to an unlimited multiple of the dead time as this ratio increases. Figure 4.4b for the ultimate gain shows a considerable upward curvature as the dead time approaches zero for a given secondary time constant to dead time ratio. These plots for the ultimate period and ultimate gain were generated from the equations developed here for integrating processes. The accuracy of the curve fit for the ultimate period is about 20 percent. The value is more in showing relationships than computing actual values. An auto tuner should be used to find the actual ultimate period and ultimate gain.

For integrating (ramping) processes based on a curve fit for Bode plots:

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o} \right)^{0.65} \right] * \theta_o \tag{4.15a}$$

For loop where the dead time is much greater than the secondary time constant ($\theta_o \gg \tau_s$), the ultimate period equation simplifies to:

$$T_u = 4 * \theta_o \tag{4.15b}$$

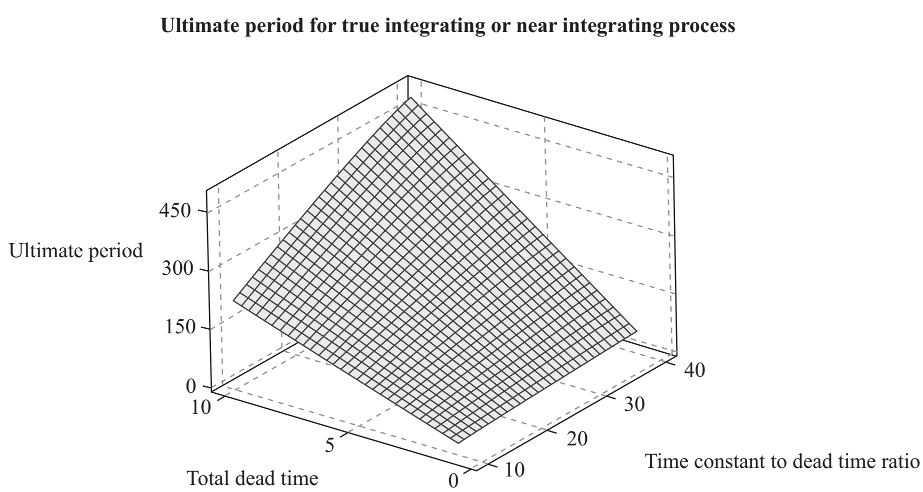


Figure 4.4a. Ultimate period for integrating processes.

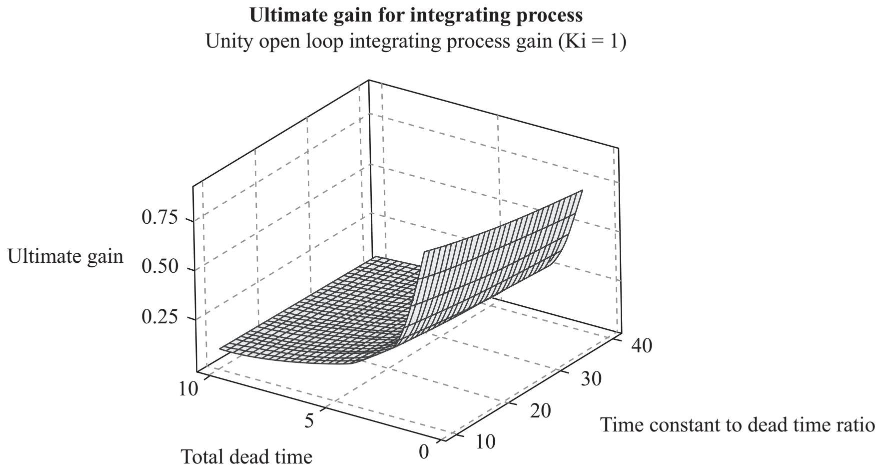


Figure 4.4b. Ultimate gain for integrating processes (unity open loop gain).

Starting with fundamental relationship that ultimate gain is the inverse of the product of the open loop gain and amplitude ratio at −180 degrees phase shift

$$K_u = \frac{1}{K_i * AR_{-180}} \tag{4.11b}$$

$$AR_{-180} = \frac{1}{\omega_n * \sqrt{1 + (\tau_s * \omega_n)^2}} \tag{4.16a}$$

$$K_u = \frac{\omega_n * \sqrt{1 + (\tau_s * \omega_n)^2}}{K_i} \tag{4.16b}$$

Using natural frequency relationship to ultimate period $\omega_n = \dfrac{2 * \pi}{T_u}$

$$K_u = \frac{2 * \pi / T_u * \sqrt{1 + (\tau_s * 2 * \pi / T_u)^2}}{K_i} \tag{4.16c}$$

For loop where the dead time is much larger than the secondary time constant ($\theta_o >> \tau_s \Leftrightarrow T_u >> \tau_s$), the ultimate gain equation simplifies to:

$$K_u = \frac{2 * \pi}{K_i * T_u} \tag{4.16d}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$K_u = \frac{1.6}{K_i * \theta_o} \tag{4.16e}$$

Assuming the largest time constant is in the process, we can approximate a near-integrating process gain as:

$$K_i = \frac{K_o}{\tau_o} \tag{4.17}$$

If we substitute the near-integrating process gain as per Equation 4.17 into 4.16e we end up with Equation 4.12e providing affirmation that a self-integrating process dominated by a large lag can be treated and tuned as integrating process. For slow continuous processes such as temperature, the tuning test time is greatly decreased and the tuning settings are greatly improved by the identification of the initial ramp rate rather than waiting for the process to reach a new steady state. In fact, Chapter 1 and Appendix C shows that the tuning settings for maximum disturbance rejection of most major tuning methods converge for near-integrating and integrating processes. Appendix D shows that the initial time response in the time frame of interest for disturbance rejection (e.g., four dead times) is a ramp for slow self-regulating, integrating, and runaway processes.

For level we can eliminate the ratio gain term in Equation 4.4a for the open loop gain (K_o) for a self-regulating process giving an equation for the integrating process gain (K_i) with the distinction of no longer being dimensionless but having units of 1/sec.

$$K_i = K_v * K_p * K_m \tag{4.18a}$$

$$K_v = \frac{\Delta F_v}{\Delta \% CO} \tag{4.18b}$$

$$K_p = \frac{1}{A * \rho} \tag{4.18c}$$

$$K_m = \frac{\Delta \% PV}{\Delta L} \tag{4.18d}$$

For PID control of processes with a slow integrating process gain (the case of more prominent interest), we have using Equation 3.4b ($K_o = 0.6*K_u$ and $T_i = 0.6*T_u$):

$$E_i = \frac{0.6 * T_u}{0.6 * K_u * K_i} * E_o = \frac{T_u}{K_u * K_i} * E_o \tag{4.19a}$$

Knowledge of K_i to calculate the integrated error is not needed since K_u is inversely proportional to K_i.

For a loop with a slow integrating process gain ($\tau_o >> \theta_o \Leftrightarrow T_u << \tau_o$), based on Equations 3.4b and 3.10b, and aggressive tuning as per Equations 4.8 and 4.9, the minimum integrated and peak errors determined by dynamics for PID control are:

$$E_i = \frac{T_u}{2 * \pi} * T_u * E_o \tag{4.19b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_i = 1.6 * \theta_o^2 * E_o \tag{4.19c}$$

For PID control of processes with a slow integrating process gain (the case of more prominent interest), we have using Equation 3.10b ($K_c = 0.6 * K_u$):

$$E_x = \frac{1.1}{0.6 * K_u * K_i} * E_o = \frac{1.8}{K_u * K_i} * E_o \tag{4.20a}$$

Knowledge of K_i to calculate the peak error is not needed since K_u is inversely proportional to K_i.

For the same conditions we stated for Equation 4.19b the peak error simplifies to:

$$E_x = \frac{1.8 * T_u}{2 * \pi} * E_o \tag{4.20b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_x = 1.2 * \theta_o * E_o \tag{4.20c}$$

Most integrating process gains are very slow resulting in a very high maximum that is well beyond the comfort range of the user. The controller gain is often set an order of magnitude lower. To prevent the start of slow rolling oscillations, the integral time must be increased by the same order of magnitude to obey the following relationship (Equation 1.5c in Chapter 1 solved for reset time):

$$T_i > \frac{2}{K_c * K_i} \tag{1.5c}$$

Consequently, a more useful set of relationships for integrated error and peak error are ones that use the actual tuning settings used. If we use a rate of change of open error with the controller in manual we end up with the same equations in Chapter 3 (Equations 3.4b and 3.10b) with the integrating process gain substituted for the open loop gain.

$$E_i = \frac{T_i}{K_c * K_i} * \Delta E_o / \Delta t \tag{4.20d}$$

$$E_x = \frac{1.1}{K_c * K_i} * \Delta E_o / \Delta t \tag{4.20e}$$

where

AR_{-180} = amplitude ratio at −180 degrees phase shift (dimensionless)
E_i = integrated error from load disturbance (e.u. * sec)
E_o = open loop error (load disturbance error with controller in manual) (e.u./sec)
E_x = peak error from load disturbance (e.u.)
A = cross sectional area (m^2)
ΔF_v = change in flow through valve or VSD (kg/sec)
K_c = controller gain for maximum disturbance rejection (dimensionless)
K_i = integrating process open loop gain (1/sec)
K_m = measurement gain (%/m for level)
K_o = open loop gain (%/%) (dimensionless)

K_p = process gain (m/kg for level)
K_v = valve or VSD gain (kg/sec/%)
K_u = controller ultimate gain (dimensionless)
T_i = integral time (reset time) (sec)
T_d = derivative time (rate time) (sec)
T_u = ultimate oscillation period (sec)
θ_o = total loop dead time (sec)
τ_o = open loop time constant (largest time constant in self-regulating process) (sec)
τ_s = secondary time constant (sec)
ω_n = natural frequency (critical frequency) (radians/sec)
$\Delta\%CO$ = change in controller output (%)
ΔF_v = change in flow through valve or VSD (kg/sec)
$\Delta\%PV$ = change in process variable (%)

4.5.3 RUNAWAY PROCESSES

The most common runaway process is a highly exothermic reaction. If the temperature gets high enough, even a well-designed coolant system may not be able to prevent an acceleration to a high temperature shutdown. Common equipment and piping design mistakes 23 to 26 in Section 4.2.2 can cause an unexpected runaway response.

Figure 4.5a for the ultimate period indicates a tremendous increase as the secondary time constant to dead time ratio increases. The period goes from four times the dead time to an unlimited multiple of the dead time as this ratio increases. Figure 4.5b for the ultimate gain shows a dramatic upward curvature as the dead time approaches zero for a given secondary time constant to dead time ratio. While the plots look similar to those for integrating processes, the scale ranges are an order of magnitude larger. These plots for the ultimate period and ultimate gain were generated from the equations developed here for runaway processes. The accuracy of the curve fit for the ultimate period is about 20 percent for typical dynamics in industrial

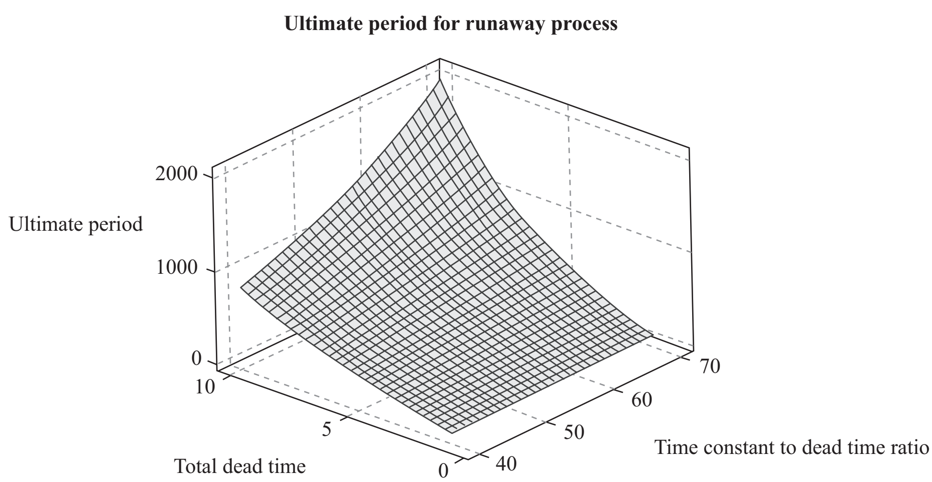


Figure 4.5a. Ultimate period for runaway process (1000 sec positive feedback time constant).

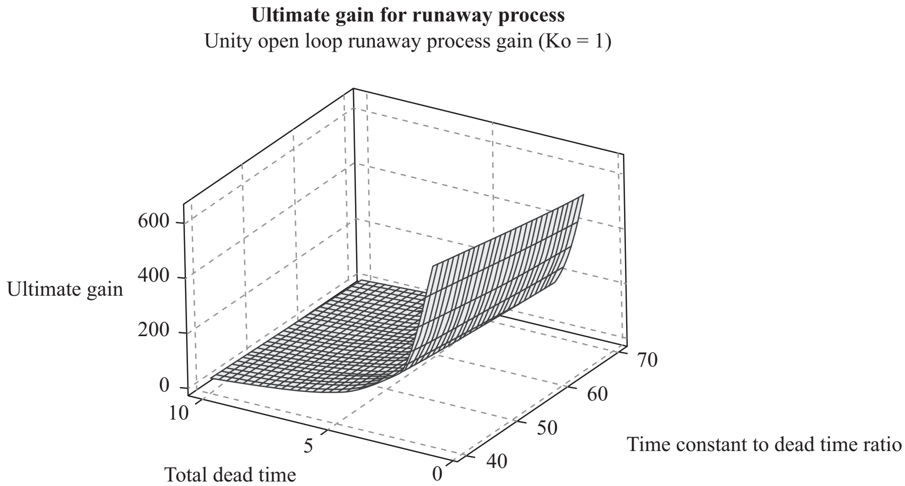


Figure 4.5b. Ultimate gain for runaway process (unity open loop gain and 1000 sec positive feedback time constant).

processes. The value is more in showing relationships than computing actual values. An auto tuner should be used to find the actual ultimate period and ultimate gain.

For runaway (positive feedback) processes based on a curve fit from Nyquist plots:

$$T_u = 4 * \left[1 + \left(\frac{N}{D} \right)^{0.65} \right] * \theta_o \tag{4.21a}$$

$$N = \left(\tau_p' + \tau_s \right) * \left(\tau_p' * \tau_s \right) \tag{4.21b}$$

$$D = \left(\tau_p' - \tau_s \right) * \left(\tau_p' - \theta_o \right) * \theta_o \tag{4.21c}$$

For loop where the dead time is much greater than the secondary time constant ($\theta_o >> \tau_s$), the ultimate period equation simplifies to:

$$T_u = 4 * \theta_o \tag{4.21d}$$

Starting with fundamental relationship that ultimate gain is the inverse of the product of the open loop gain and amplitude ratio at −180 degrees phase shift based on an analogous relationship to what was used for self-regulating and integrating processes.

$$K_u = \frac{1}{K'_o * AR_{-180}} \tag{4.11c}$$

$$AR_{-180} = \frac{1}{\sqrt{1 + (\tau_p' * \omega_n)^2} * \sqrt{1 + (\tau_s * \omega_n)^2}} \tag{4.22a}$$

$$K_u = \frac{\sqrt{1 + (\tau_p' * \omega_n)^2} * \sqrt{1 + (\tau_s * \omega_n)^2}}{K_o'} \tag{4.22b}$$

Using natural frequency relationship to ultimate period $\omega_n = \dfrac{2 * \pi}{T_u}$

$$K_u = \frac{\sqrt{1 + (\tau_p' * 2 * \pi / T_u)^2} * \sqrt{1 + (\tau_s * 2 * \pi / T_u)^2}}{K_o'} \tag{4.22c}$$

For loop dominated by a large time constant ($\tau_p' >> \tau_s \Leftrightarrow T_u << \tau_p'$), the ultimate gain equation simplifies to:

$$K_u = \frac{2 * \pi * \tau_p'}{K_o' * T_u} \tag{4.22d}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$K_u = \frac{1.6 * \tau_p'}{K_o' * \theta_o} \tag{4.22e}$$

For stability, the controller gain must be greater than the inverse of the open loop gain.

$$K_c > \frac{1}{K_o'} \tag{4.22f}$$

For PID control of processes dominated by a large time constant (the case of more prominent interest), we have using Equation 3.4b (K_c = 06 * K_u and T_i = 0.6 * T_u):

$$E_i = \frac{0.6 * T_u}{0.6 * K_u * K_o'} * E_o = \frac{T_u}{K_u * K_o'} * E_o \tag{4.23a}$$

Knowledge of K_o' to calculate the integrated error is not needed since K_u is inversely proportional to K_o'.

For a loop dominated by a large time constant ($\tau_p' >> \theta_o \Leftrightarrow T_u << \tau_p'$), based on Equations 3.4b and 3.10b, and aggressive tuning as per Equations 4.8 and 4.9, the minimum integrated and peak errors determined by dynamics for PID control are:

$$E_i = \frac{T_u}{2 * \pi * \tau_p'} * T_u * E_o \tag{4.23b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_i = 2.6 * \frac{\theta_o^{\,2}}{\tau_p'} * E_o \tag{4.23c}$$

For PID control of processes dominated by a large time constant (the case of more prominent interest), we have using Equation 3.4b (K_c = 0.6 * K_u):

$$E_x = \frac{1.1}{0.6 * K_u * K_o'} * E_o = \frac{1.8}{K_u * K_o'} * E_o \tag{4.24a}$$

Knowledge of K_o to calculate the peak error is not needed since K_u is inversely proportional to K_o.

For the same conditions we stated for Equations 4.13b the peak error simplifies to:

$$E_x = \frac{1.8 * T_u}{2 * \pi * \tau'_p} * E_o \tag{4.24b}$$

For an ultimate period being about four dead times ($T_u \cong 4 * \theta_o$):

$$E_x = 1.2 * \frac{\theta_o}{\tau'_p} * E_o \tag{4.24c}$$

where

AR_{-180} = amplitude ratio at −180 degrees phase shift (dimensionless)
E_i = integrated error from load disturbance (e.u. * sec)
E_o = open loop error (load disturbance error with controller in manual) (e.u.)
E_x = peak error from load disturbance (e.u.)
K_c = controller gain for maximum disturbance rejection (dimensionless)
K'_o = positive feedback process open loop gain (%/%) (dimensionless)
K_u = controller ultimate gain (dimensionless)
T_i = integral time (reset time) (sec)
T_d = derivative time (rate time) (sec)
T_u = ultimate oscillation period (sec)
θ_o = total loop dead time (sec)
τ'_p = positive feedback process time constant (sec)
τ_s = secondary time constant (sec)
ω_n = natural frequency (critical frequency) (radians/sec)

4.6 EXAMPLES

The following examples for common unit operations show how process dynamics affect the ultimate period and ultimate gain and consequentially the loop performance. More information on the source of the solutions can be obtained by going to the portion of the chapters and appendixes denoted by the equation numbers shown in the solutions.

4.6.1 WASTE TREATMENT pH LOOPS (SELF-REGULATING PROCESS)

Given:

a. Set point is pH 7.
b. Measurement range is pH 0 to 14.
c. Minimum influent flow is 10 gallons per minute (gpm) (F_{il}).
d. Normal influent flow is 22 gpm (F_{in}).
e. Maximum influent flow is 100 gpm (F_{ih}).
f. Influent concentration is 32 percent by weight HCl (10.17 normality) (C_i).
g. Influent disturbance is rapid increase of 2.2 normality HCl concentration (ΔC).

h. Reagent concentration is 20 percent by weight NaOH (7.93 normality) (C_r).
i. Vertical tank liquid volume is 1,000 gallons (V).
j. Axial blade agitator diameter is 1 foot (D).
k. Axial blade agitator speed is 120 rpm (N_s).
l. Axial blade agitator discharge coefficient is 1 (N_q).

Find: minimum possible peak and integrated pH errors for one, two, and three tanks in series with individual control loops and with just one overall control loop determined by process dynamics (automation response is instantaneous and controller tuning is ideal).

Solution:

a. Calculate the ultimate period of an individual pH loop on each vessel
Equivalent flow due to agitation via Equation L.3 from Appendix L on liquid mixing:

$$F_a = 7.48 * N_q * N_s * D_t^3 \tag{L.3}$$

$$F_a = 7.48 * 1 * 60 * 1 = 450\ gpm$$

Minimum reagent flow:

$$F_{rl} = \frac{C_i}{C_r} * F_{il} = \frac{10.17}{7.93} * 10 = 12.82\ gpm$$

Normal reagent flow:

$$F_{rn} = \frac{C_i}{C_r} * F_{in} = \frac{10.17}{7.93} * 22 = 28.21\ gpm$$

Maximum reagent flow:

$$F_{rh} = \frac{C_i}{C_r} * F_{ih} = \frac{10.17}{7.93} * 100 = 128.2\ gpm$$

Total feed flow from normal influent and reagent flows:

$$F_f = F_{rn} + F_{in} = 28.21 + 22 = 50.2\ gpm$$

Dead time is half the turnover time (volume divided by total of stream and agitation flows):

$$\theta_p = 0.5 * \frac{V}{F_f + F_a} = 0.5 * \frac{1000}{50 + 450} = 1 \min \tag{L.1}$$

Process time constant is residence time less mixing dead time (turnover time):

$$\tau_p = \frac{V}{F_{in}} - \theta_p = \frac{1000}{50} - 1 = 19 \text{ min} \tag{L.2}$$

Since we are only considering the effects of process dynamics (measurement, controller, and valve dynamics not included) we have $\theta_o = \theta_p$ and $\tau_o = \tau_p$.

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o}\right)^{0.65}\right] * \theta_o = 2 * \left[1 + \left(\frac{19}{19+1}\right)^{0.65}\right] * 1 = 4 \min \tag{4.5a}$$

b. Calculate the ultimate period for one overall pH loop three vessels:

$$\theta_j = Y_j * N * \tau_j = 0.28 * 3 * 19 = 16 \min \tag{4.2c}$$

$$\tau_o = (1 - Y_j) * N * \tau_j = (1 - 0.28) * 3 * 19 = 41 \min \tag{4.4}$$

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o}\right)^{0.65}\right] * \theta_o = 2 * \left[1 + \left(\frac{41}{41+16}\right)^{0.65}\right] * 16 = 58 \min \tag{4.5a}$$

c. Calculate the ultimate limit of concentration peak errors for individual loops on each vessel:
Calculate peak error of first vessel from E_o step disturbance:

$$E_{x1} = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o = \frac{1.8 * 4}{2 * \pi * 19} * 2.2 = 0.13 \textit{ normality} \tag{4.14b}$$

Calculate the peak error of second vessel from peak error of first vessel taking into account the disturbance is slowed down by the first tank volume and the open loop error for the loop on the second vessel can be approximated as the peak error from the first vessel:

$$E_{x2} = (1 - e^{-\theta_o/\tau_o}) * \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_{x1} = 0.2 * \frac{1.8 * 4}{2 * \pi * 19} * 0.13 = 0.0016 \textit{ normality}$$

Calculate peak error of third vessel from peak error of second vessel:

$$E_{x3} = (1 - e^{-\theta_o/\tau_o}) * \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_{x2} = 0.2 * \frac{1.8 * 4}{2 * \pi * 19} * 0.0016 = 0.00002 \textit{ normality}$$

d. Calculate the ultimate limit of pH peak errors for individual loops on each vessel using pH definition:

$$E_{xpH1} = 7 - \log(E_{x1}) = 7 - \log(0.13) = 6.1\,pH$$

$$E_{xpH2} = 7 - \log(E_{x2}) = 7 - \log(0.0016) = 4.2\,pH$$

$$E_{xpH3} = 7 - \log(E_{x3}) = 7 - \log(0.00002) = 2.3\,pH$$

e. Calculate the ultimate limit of concentration integrated errors for individual loops on each vessel:
The integrated error in terms of the peak error is approximately the ratio of Equation 4.13a to 4.14a:

$$R = \frac{E_i}{E_x} = \frac{\dfrac{T_u}{K_u * K_o}}{\dfrac{1.8}{K_u * K_o}} = \frac{4}{1.8} = 2.2$$

$$E_{i1} = R * E_{x1} = 2.2 * 0.13 = 0.286 \; normality * \min$$

$$E_{i2} = R * E_{x2} = 2.2 * 0.0016 = 0.0035 \; normality * \min$$

$$E_{i3} = R * E_{x3} = 2.2 * 0.00002 = 0.000044 \; normality * \min$$

f. Calculate the ultimate limit of pH integrated errors for individual loops on each vessel:

$$E_{ipH1} = R * E_{xpH1} = 2.2 * 6.1 = 13 \; pH * \min$$

$$E_{ipH2} = R * E_{xpH2} = 2.2 * 4.2 = 9 \; pH * \min$$

$$E_{ipH3} = R * E_{xpH3} = 2.2 * 2.3 = 5 \; pH * \min$$

g. Calculate the ultimate limit of the concentration peak error and the associated pH peak error and the consequential integrated errors for an overall loop:

$$E_x = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o = \frac{1.8 * 58}{2 * \pi * 41} * 2.2 = 0.9 \; normality \tag{4.14b}$$

$$E_{xpH} = 7 - \log\left(E_x\right) = 7 - \log\left(0.9\right) = 6.95 \, pH$$

The integrated error in terms of the peak error is approximately the ratio of Equation 4.13a to 4.14a:

$$R = \frac{E_i}{E_x} = \frac{\dfrac{T_u}{K_u * K_o}}{\dfrac{1.8}{K_u * K_o}} = \frac{58}{1.8} = 32$$

$$E_i = R * E_x = 32 * 0.9 = 28.8 \; normality * \min$$

$$E_{ipH} = R * E_{xpH} = 32 * 6.95 = 222 \; pH * \min$$

Conclusions: The concentration peak and integrated errors are several orders of magnitude larger for the overall loop compared to the individual loop on the last vessel. The increase in errors expressed in pH units does not seem as bad because of the exponential relationship between normality and pH. However, normality errors represent the reagent (process) errors, which are a better index of the loop's performance than the pH (measurement) errors. The use of pH or oxidation reduction potential (ORP) for composition control greatly increases the rangeability and sensitivity for set points near neutrality but decreases the awareness of the loop's performance. Also, the nonlinearity of the pH process gain necessitates decreasing the controller gain to match the highest process gain (at neutrality for pH loops), which increases the errors over the rest of the pH measurement range.

4.6.2 BOILER FEEDWATER FLOW LOOP (SELF-REGULATING PROCESS)

Given:

a. Set point is 100,000 pounds per hour (pph).
b. Measurement range is 0 to 200,000 pph.

c. Disturbance is rapid 20 percent increase in flow (ΔF).
d. Pipe diameter is 4 inches or 0.33 foot (D_p).
e. Fluid velocity is 5 feet per second (fps) (V_f).
f. Pipe friction factor is 0.01 (C_f).
g. Pipe wall modulus of elasticity is 500,000,000 lb/ft^2 (E_p).
h. Pipe wall thickness is 0.34 inch or 0.03 foot (H).
i. Fluid density is 62 lb/ft^3 (ρ).
j. Fluid modulus of elasticity is 5,000,000 lb/ft^2 (E_f).
k. Acceleration due to gravity is 32 ft/(sec * sec) (G).
l. Pipe length from valve to transmitter or discharge is 150 feet (S_p).
m. Ratio of valve pressure drop to pipe pressure drop is 0.7 (R).

Find: minimum possible peak and integrated flow errors determined by process dynamics (automation response is instantaneous and controller tuning is ideal).

Solution:

a. Calculate the ultimate period using equations from Table 4.1

The process dead time is the distance divided by the velocity of a pressure wave:

$$V_w = \sqrt{\frac{G}{\rho * \left(\frac{D_p}{E_p * H} + \frac{1}{E_f} \right)}} = \sqrt{\frac{32}{62 * \left(\frac{0.33}{500,000,000 * 0.03} + \frac{1}{5,000,000} \right)}} = 1500\, fps$$

V_w is the velocity of a pressure wave as per an equation in Bergeron (1961). Fluid does not start to accelerate at the transmitter until the pressure change due to change in stroke of the valve propagates from the valve to the discharge.

$$\theta_p = \frac{S_p}{V_w} = \frac{150}{1500} = 0.1\,\text{sec} \qquad \text{(Table 4.1)}$$

This process dead time is much less than the automation system dead time from sensor lag, transmitter damping, and PID module execution time.

$$\tau_p = \frac{D_p}{4 * C_f * V_f * (1+R)} = \frac{0.33}{4 * 0.01 * 5 * (1+0.7)} = 1\,\text{sec} \qquad \text{(Table 4.1)}$$

Since we are only considering the effects of process dynamics (measurement, controller, and valve dynamics not included) we have $\theta_o = \theta_p$ and $\tau_o = \tau_p$.

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o} \right)^{0.65} \right] * \theta_o = 2 * \left[1 + \left(\frac{1}{1+0.1} \right)^{0.65} \right] * 0.1 = 0.4\,\text{sec} \qquad (4.5a)$$

b. Calculate the ultimate limit of peak and integrated errors

$$E_o = \Delta F = 40000\ pph = 11\, lb/\text{sec}$$

$$E_x = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o = \frac{1.8 * 0.4}{2 * \pi * 1} * 11 = 1.2\ lb/\sec \qquad (4.14b)$$

$$E_i = \frac{T_u}{2 * \pi * \tau_o} * T_u * E_o = \frac{0.4}{2 * \pi * 1} * 0.4 * 11 = 0.28\ lb \qquad (4.13b)$$

4.6.3 BOILER DRUM LEVEL LOOP (INTEGRATING PROCESS)

Given:

a. Set point is 2 feet.
b. Measurement range is 0 to 4 feet.
c. Boiler feedwater flow dynamics are as described in Section 4.6.2.
d. Level controller output goes directly to feedwater valve.
e. Disturbance is rapid 20 percent decrease in steam flow (*DF*).
f. Boiler drum diameter is 4 feet (*Dd*).
g. Boiler drum length is 10 feet (*Sd*).

Find: peak and integrated level errors determined by process dynamics where shrink and swell is negligible (consistent gradual response).

Solution:
If we ignore automation system dynamics, the level response dead time and secondary time constant is the Example 4.6.2 boiler feedwater flow process time constant and dead time, respectively.

a. Calculate the ultimate period for the level loop from flow process dynamics

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o}\right)^{0.65}\right] * \theta_o = 4 * \left[1 + \left(\frac{1}{0.1}\right)^{0.65}\right] * 0.1 = 2.2 \sec \qquad (4.15a)$$

b. Calculate the open loop integrating process gain
Using Equations 4.18a, 4.18b, 4.18c, and 4.18d, we can estimate valve gain, process gain, measurement gain, and the consequential open loop integrating process gain.

$$K_v = \frac{\Delta F_v}{\Delta \% CO} = \frac{2000000\ pph}{100\%} = \frac{56\ lb/\sec}{100\%} = \frac{0.56\ lb/\sec}{\%} \qquad (4.18b)$$

$$K_p = \frac{1}{A * \rho} = \frac{1}{D_d * S_d * \rho} = \frac{1}{4 * 10 * 42} = \frac{0.0006\, ft}{lb} \qquad (4.18c)$$

$$K_m = \frac{\Delta \% PV}{\Delta L} = \frac{100\%}{ft} = \frac{25\%}{ft} \qquad (4.18d)$$

$$K_i = K_v * K_p * K_m = \frac{0.56\ lb/\sec}{\%} * \frac{0.0006\, ft}{lb} * \frac{25\%}{ft} = \frac{0.008\ \%/\sec}{\%} \qquad (4.18a)$$

c. Calculate the open loop error

$$E_o = K_p * K_m * \Delta F = \frac{0.0006\, ft}{lb} * \frac{25\%}{ft} * 11\ lb/\sec = 0.165\ \%/\sec$$

d. Calculate the PID controller gain and corresponding integral time settings using Equation 4.16d for the ultimate gain and Equation 4.6 for a PI controller gain:

$$K_u = \frac{2 * \pi}{K_i * T_u} = \frac{6.28}{0.008 * 2.2} = 357 \tag{4.16d}$$

$$K_c = 0.4 * K_u = 143 \tag{4.6}$$

The controller gain used depends more on measurement noise than the ultimate gain. A more practical gain would be 10 provided that shrink and swell is not a factor ($K_c = 10$).

To prevent slow rolling oscillations, the integral time must be increased by about the same factor that the gain was decreased.

$$T_i > \frac{2}{K_c * K_i} = \frac{2}{10 * 0.008} = 25 \text{ sec} \tag{1.5d}$$

e. Calculate the practical limit of the peak and integrated errors for a PID controller gain of 10 instead of 214 maximum gain using Equations 3.10b and 3.4b from Chapter 3.

$$E_x = \frac{1.1}{K_c * K_i} * E_o = \frac{1.1}{10 * 0.008} * 0.165 = 2.3\,\% \tag{3.10b}$$

$$E_i = \frac{T_i}{K_c * K_i} * E_o = \frac{25}{10 * 0.008} * 0.165 = 52\,\% * \text{sec} \tag{3.4b}$$

4.6.4 FURNACE PRESSURE LOOP (NEAR-INTEGRATING PROCESS)

Given:

a. Set point is 5 inches water column (w.c.) gage.
b. Measurement range is 0 to 10 inches w.c. gage.
c. Furnace volume is 10,000 ft^3 (V_f).
d. Quench volume is 1,000 ft^3 (V_q).
e. Scrubber volume is 1,000 ft^3 (V_s).
f. Furnace inlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_f).
g. Quench inlet flow resistance pressure drop is 5 inches w.c. (ΔP_q).
h. Scrubber inlet flow resistance pressure drop is 10 inches w.c. (ΔP_s).
i. System outlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_o).
j. Flue gas flow is 1,000 scfm (F_f).
k. Atmospheric pressure is 408 inches w.c. (P_a).
l. Disturbance is a rapid 20 percent increase in inlet pressure (ΔP).
m. Inlet pressure (discharge of forced draft fan) is 15 inches w.c. (P_i).

Find: minimum possible peak and integrated pressure errors determined by process dynamics (automation response is instantaneous and controller tuning is ideal).

Solution:

a. Calculate the ultimate period from the system resistances and capacitances using Equations G.3 and G.4 for each volume's resistance and capacitance in Appendix G on gas dynamics and Equations I.1–I.5 in Appendix I on interacting time constants.

$$R_f = \frac{2*\Delta P_f}{F_f} = \frac{2*2.5\ \textit{inches w.c.}}{1000\ \textit{scfm}} = 0.005\ \textit{inches w.c. per scfm} \tag{G.3}$$

$$R_q = \frac{2*\Delta P_q}{F_f} = \frac{2*5\ \textit{inches w.c.}}{1000\ \textit{scfm}} = 0.01\ \textit{inches w.c. per scfm} \tag{G.3}$$

$$R_s = \frac{2*\Delta P_s}{F_f} = \frac{2*10\ \textit{inches w.c.}}{1000\ \textit{scfm}} = 0.02\ \textit{inches w.c. per scfm} \tag{G.3}$$

$$R_o = \frac{2*\Delta P_o}{F_f} = \frac{2*2.5\ \textit{inches w.c.}}{1000\ \textit{scfm}} = 0.005\ \textit{inches w.c. per scfm} \tag{G.3}$$

$$C_f = \frac{V_f}{P_a} = \frac{10000\ \textit{scf}}{408\ \textit{inches w.c.}} = 24.5\ \textit{scf per inch w.c.} \tag{G.4}$$

$$C_q = \frac{V_q}{P_a} = \frac{1000\ \textit{scf}}{408\ \textit{inches w.c.}} = 2.45\ \textit{scf per inch w.c.} \tag{G.4}$$

$$C_s = \frac{V_s}{P_a} = \frac{1000\ \textit{scf}}{408\ \textit{inches w.c.}} = 2.45\ \textit{scf per inch w.c.} \tag{G.4}$$

$$\tau_{p1} = \frac{2*A}{B+\sqrt{B^2-4*A*C}} \tag{I.1}$$

$$\tau_{p2} = \frac{2*A}{B-\sqrt{B^2-4*A*C}} \tag{I.2}$$

$$A = R_q * C_q * R_s * C_q * R_o \tag{I.3}$$

$$A = 0.01*2.45*0.02*0.005 = 0.000006$$

$$B = R_q * C_q * R_o + R_q * C_q * R_s + R_s * C_s * R_o + R_q * C_s * R_o \tag{I.4}$$

$$B = R_q * C_q * (R_o + R_s) + (R_s + R_q) * C_s * R_o$$

$$B = 0.01*2.45*(0.005+0.02)+(0.02+0.01)*2.45*0.005 = 0.001$$

$$C = R_q + R_s + R_o = 0.01+0.02+0.005 = 0.035 \tag{I.5}$$

$$\tau_{p1} = \frac{2*0.000006}{0.001+\sqrt{0.001^2-4*0.000006*0.035}} = \frac{0.000012}{0.001+0.004} = 0.0086\ \text{min} = 0.5\,\text{sec}$$

$$\tau_{p2} = \frac{2*0.000006}{0.001-\sqrt{0.001^2-4*0.000006*0.035}} = \frac{0.000012}{0.001-0.004} = 0.02\ \text{min} = 1.2\,\text{sec}$$

Find the fraction of the smallest time constant converted to equivalent dead time:

$$\theta_o = \sum Y_i * \tau_i = Y_1 * \tau_{p1} = 0.4 * 0.5 = 0.2 \text{ sec} \tag{4.2a}$$

The secondary time constant for an integrating process is the fraction of the smaller time constant not converted to dead time plus the larger time constant:

$$\tau_s = \sum (1 - Y_1) * \tau_{p1} + \tau_{p2} = (1 - 0.4) * 0.5 + 1.2 = 1.5 \text{sec} \tag{4.4}$$

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o} \right)^{0.65} \right] * \theta_o = 4 * \left[1 + \left(\frac{1.5}{0.2} \right)^{0.65} \right] * 0.2 = 3.8 \text{ sec} \tag{4.15a}$$

b. Calculate the integrating process gain using the near integrator gain approximation:

$$K_i = \frac{K_o}{\tau_o} = \frac{R_q}{R_f + R_q} * \frac{R_f + R_q}{R_f * R_q * C_f} = \frac{1}{R_f * C_f} = \frac{1}{0.005 * 24.5} = 8.2 \text{ per min} \tag{4.17}$$

$$K_i = 8.2 / 60 = 0.14 \text{ per sec}$$

c. Calculate the open loop error

$$E_o = K_i * \Delta P = 0.14 * 3 = 0.4 \text{ inches per sec.}$$

d. Calculate the ultimate limit for peak and integrated errors:

$$E_x = \frac{1.8 * T_u}{2 * \pi} * E_o = \frac{1.8 * 3.8}{2 * \pi} * 0.4 = 0.4 \textit{ inches w.c.} \tag{4.20b}$$

$$E_i = \frac{T_u}{2 * \pi} * T_u * E_o = \frac{3.8}{2 * \pi} * 3.8 * 0.4 = 0.9 \textit{ inches w.c.} * \text{sec} \tag{4.19b}$$

4.6.5 EXOTHERMIC REACTOR CASCADE TEMPERATURE LOOP (RUNAWAY PROCESS)

Given:

a. Set point is 150 °F.
b. Measurement range is 100 to 200 °F.
c. Reactant feed flow is 2,000 pph (W_r).
d. Reactant mass is 3,000 pounds (M_r).
e. Reactant heat capacity is 0.5 Btu/(lb * °F) (C_r).
f. Heat transfer coefficient*area is 8,000 Btu/(hr * °F) (UA).
g. Change in heat generation with temperature is 12,000 Btu/(hr * °F)($\Delta Q/\Delta T$).
h. Coolant mass is 400 pounds (M_c).
i. Coolant heat capacity is 1 Btu/(lb * °F) (C_c).
j. Coolant flow is 80,000 pph (W_c).
k. Disturbance is a 20 percent increase in coolant temperature (ΔT).

l. Coolant temperature is 100 °F.
m. Continuous reactor.

Find: minimum possible peak and accumulated temperature errors determined by process dynamics (automation response is instantaneous and controller tuning is ideal).

Solution:

a. Calculate the ultimate period using equations from Table 4.1 and ultimate gain

$$\tau'_p = \frac{M_r * C_r}{UA + W_r * C_r - \Delta Q / \Delta T} = \frac{3000 * 0.5}{8000 + 2000 * 0.5 - 12000} = 0.5\ hour = 30\ \text{min}$$

$$\tau_s = \frac{M_c * C_c}{UA} = \frac{400 * 1}{8000} = 0.05\ hour = 3\,\text{min}$$

$$\theta_p = \frac{M_c}{W_c} = \frac{400}{80000} = 0.005\ hour = 0.3\,\text{min}$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D}\right)^{0.65}\right] * \theta_o \tag{4.21a}$$

$$N = \left(\tau'_p + \tau_s\right) * \left(\tau'_p * \tau_s\right) = (30 + 3) * 30 * 3 = 2970 \tag{4.21b}$$

$$D = \left(\tau'_p - \tau_s\right) * \left(\tau'_p - \theta_o\right) * \theta_o = (30 - 3) * (30 - 0.3) * 0.3 = 241 \tag{4.21c}$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D}\right)^{0.65}\right] * \theta_o = 4 * \left[1 + \left(\frac{2970}{241}\right)^{0.65}\right] * 0.3 = 7.4\ \text{min}$$

b. Compute the ultimate gain from the ultimate period and the primary and secondary time constants.

$$K_u = \frac{\sqrt{1 + (\tau'_p * 2 * \pi / T_u)^2} * \sqrt{1 + (\tau_s * 2 * \pi / T_u)^2}}{K'_o} \tag{4.22c}$$

$$K_u = \frac{\sqrt{1 + (30 * 2 * \pi / 7.4)^2} * \sqrt{1 + (3 * 2 * \pi / 7.4)^2}}{K'_o} = \frac{25.5 * 2.7}{K_o} = \frac{68.8}{K_o}$$

c. Calculate the open loop positive feedback process gain assuming other gain factors are one:

$$K'_o = \frac{UA}{UA + W_r * C_r - \Delta Q / \Delta T} = \frac{8000}{8000 + 2000 * 0.5 - 12000} = 2.7$$

d. Calculate the open loop error:

$$E_o = K'_o * \Delta T = 2.7 * 20\ \text{deg}\ F = 53\ {}^oF$$

e. Calculate the ultimate limit for peak and integrated errors:

$$E_x = \frac{1.1}{0.6 * K_u * K'_o} * E_o = \frac{1.8}{\frac{68.8}{K_o} * K'_o} * 53 = 1.4 \ {}^oF \tag{4.24a}$$

$$E_i = \frac{0.6 * T_u}{0.6 * K_u * K'_o} * E_o = \frac{T_u}{K_u * K'_o} * E_o \tag{4.23a}$$

$$E_i = \frac{T_u}{K_u * K'_o} * E_o = \frac{7.4}{\frac{68.8}{K_o} * K'_o} * 53 = 5.7 \ {}^oF * \min$$

4.6.6 BIOLOGICAL REACTOR BIOMASS CONCENTRATION LOOP (RUNAWAY PROCESS)

Given:

a. Average nutrient concentration is 4 nanograms per milliliter (ng/ml) (C_i).
b. Nutrient concentration where growth rate is half its maximum is 1 ng/ml (K_i).
c. Maximum cell growth rate is 0.01 generation per minute (U_x).
d. Cell death rate is 0.001 generation per minute (K_d).
e. Reactor working volume is 600 gallons (V).
f. Average throughput flow is 10 gpm (F).
g. Agitator pumping rate is 200 gpm (F_a).
h. Open loop positive feedback process gain is 1 (K'_o).

Find: the window of allowable controller gains based on process dynamics (automation system response is instantaneous)

Solution:

a. Calculate the ultimate period using equations from Table 4.1 and ultimate gain:

$$\theta_p = 0.5 * \frac{V}{F + F_a} = \frac{600}{200 + 10} = 3 \text{ min}$$

$$\tau_s = \frac{V}{F} - \theta_p = \frac{600}{10} - 3 = 57 \text{ min}$$

$$U = U_x * \frac{C_i}{K_i + C_i} - K_d = 0.01 * \frac{4}{1+4} - 0.001 = 0.007$$

$$\tau'_p = \frac{1}{U} = \frac{1}{0.007} = 142 \text{ min}$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D}\right)^{0.65}\right] * \theta_o \tag{4.21a}$$

$$N = \left(\tau'_p + \tau_s\right) * \left(\tau'_p * \tau_s\right) = \left(142 + 57\right) * \left(142 * 57\right) = 1610706 \tag{4.21b}$$

$$D = \left(\tau_p' - \tau_s\right) * \left(\tau_p' - \theta_o\right) * \theta_o = \left(142 - 57\right) * \left(142 - 3\right) * 3 = 35445 \tag{4.21c}$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D}\right)^{0.65}\right] * \theta_o = 4 * \left[1 + \left(\frac{1610706}{35445}\right)^{0.65}\right] * 3 = 155 \text{ min} \tag{4.21a}$$

$$K_u = \frac{\sqrt{1 + (\tau_p' * 2 * \pi / T_u)^2} * \sqrt{1 + (\tau_s * 2 * \pi / T_u)^2}}{K'_o} \tag{4.22c}$$

$$K_u = \frac{\sqrt{1 + (142 * 2 * \pi / 155)^2} * \sqrt{1 + (57 * 2 * \pi / 155)^2}}{1}$$

$$K_u = \sqrt{1 + (5.75)^2} * \sqrt{1 + (2.3)^2} = 5.8 * 2.5 = 20.8$$

b. Calculate the minimum and maximum PID controller gains.
The minimum PID gain to provide enough feedback action of open loop unstable process:

$$K_c > \frac{1}{K_o} = 1 \tag{4.22f}$$

The maximum PID gain based on the ultimate gain:

$$K_c = 0.3 * K_u = 0.3 * 20.8 = 6 \tag{4.8}$$

KEY POINTS

1. Pure process dead time usually originates from transportation delays in the piping, static mixer, dip tube, jacket, and coil design and mixing delays in the agitator- equipment design.
2. Equivalent process dead time is predominately associated with thermal lags and volumes in series.
3. The fraction of small time constants converted to dead time increases as the size of the small time constant relative to the primary time constant decreases.
4. The fraction of large equal time constants converted to dead time increases as the number of the equal time constants increases.
5. In a continuous liquid *well-mixed* vessel (vessel with good agitation rate and pattern), most of the residence time (liquid volume/feed rate) is a beneficial primary process time constant.
6. In a poorly mixed volume, the dead time is large and variable and the response is erratic and noisy.
7. Mixing is best provided by an axial agitator but adequate mixing may be attained by a large and directed recirculation.
8. Excessive dead time and extraneous responses in mechanical design is best prevented early in a project design by the automation, mechanical, and process design engineers

getting together with a basic understanding of process control. Chemical engineering and mechanical design are taught from a steady state rather than a dynamic viewpoint with little to no understanding of the consequences in terms of tuning and control loop performance.

9. Missing process negative feedback action (true integrating or runaway processes) or slow negative feedback action (near-integrating processes) requires more immediate aggressive PID action (higher PID gain and rate settings) to provide the negative feedback needed for stability and to reach setpoint quickly.
10. While Ziegler–Nichols ultimate oscillation method is rarely used, the estimation of the ultimate period and gain provides a benchmark for diagnosing problems.
11. The ultimate period for a loop with good process and automation system dynamics (consistent gradual response) is about four times the dead time.
12. The ultimate period for a loop where the dead time is much greater than the primary time constant is about two times the dead time. While these loops are not as common, there are important examples in extruder and sheet line control.
13. True integrating and runaway processes with a large secondary time constant (often due to large thermal lag in heat transfer surfaces or a thermowell) can have an ultimate period much larger than four times the dead time.
14. The maximum PID gain should be less than one-fourth the ultimate gain for a smooth closed loop response (exceptions are an extremely accurate dead time compensator or an enhanced PID is used on a loop where nearly all of the dead time is from a discontinuous update such as that provided by an at-line analyzer).
15. The ultimate gain is about equal to the inverse of the maximum open loop gain for dead time dominant processes.
16. The ultimate gain is much greater than the inverse of the maximum open loop gain for near-integrating, true integrating, and runaway processes.
17. The additional pure delays and the small lags that become equivalent dead time from the controller (Chapter 5), measurement (Chapter 6), and valve or variable frequency drive (Chapter 7), should be added to the process dead time to get the total dead time in the loop for the calculation of the ultimate period and gain.
18. The gain margin is approximately the ratio of the ultimate gain to the PID gain.

CHAPTER 5

Effect of Controller Dynamics

5.1 INTRODUCTION

The effect of proportional-integral-derivative (PID) execution rate and filter time has been the subject of much confusion. Simulation tests often show little to no effect of execution rate on loop dead time because of the timing of the change in setpoint or disturbance in the test setup. Filter tests can be deceptive as well if the unfiltered process variable (PV) is not trended. Finally there are two effects at play, each dependent upon the tuning settings. A difference in tuning settings can change the conclusions reached on the effect of PID execution rate and filter time.

If the PID execution rate and filter time are relatively fast compared to the reset time, the first effect directly computable via the equation for the integrated error is small. If the PID is detuned, the second effect due to the increase in dead time can be negligible. An innovative method is developed for showing when the execution rate and filter time will start to cause an oscillation from an increase in dead time.

The PID execution rate and filter time can be too fast. Based on the amplitude of noise and deadband, threshold sensitivity and resolution in the measurement and valve or variable speed drive (VSD), an increase in the execution rate or filter time can be estimated. The objective here is to increase the measurement signal to noise ratio and prevent excessive changes in valve position or drive speed that are undesirable in terms of maintenance and disruption to related loops.

Execution rate is the time between executions and is not the number of executions per second. Subsequently this book uses the term execution time instead of execution rate.

Disturbances and controller tuning are generally larger contributors to process variability than PID execution time and filter time. A simple diagnostic procedure can find out which PID and which tuning setting is creating problem or making the situation worse for a deficiency in the automation system design or implementation.

5.1.1 PERSPECTIVE

Tuning has a profound effect on the practical limit to control loop performance. Chapter 3 detailed how the peak error, integrated error, and rise time were inversely proportional to controller gain and the integrated error was proportional to reset time.

The equation for integrated error also showed the incremental increase by the PID execution time and filter time. These parameters can potentially further decrease performance by creating additional dead time. However, the consequences are not well recognized leading to misconceptions.

Test cases using simulations are often misleading because they do not have the spectrum of dynamics and timing randomness. A test on the effect of PID execution time may not show an increase in dead time if the step change arrives before instead of after a PID execution. A test of an increase in filter time will not show an appreciable increase in dead time if the other time constants in the loop are a negligible source of dead time.

The results for closed loop tests depend upon the tuning. If the controller is sufficiently detuned, no deterioration is observable beyond what is predicted by the equation for the integrated error. Equations for the implied dead time from tuning are developed to help guide the practitioner. The implied dead time minus the original dead time is the margin of dead time available for execution time and filter time that does not trigger an oscillatory response.

The implied dead time is a back calculation of dead time from the actual tuning settings. Note that most PID controllers are not tuned as aggressively as permitted by the actual loop dead time. Consequently the implied dead time is almost always greater than the actual dead time. The major insight is that a sluggishly PID tuned will perform as badly as a loop with an actual dead time equal to the implied dead time.

A transmitter damping setting has the same effect as a signal filter. For wireless devices, it is desirable to use the transmitter damping rather than a signal filter in the Distributed Control Systems (DCS) to reduce noise. The transmitter damping can be set to keep measurement noise less than the wireless trigger level to prevent unnecessary updates, prolonging battery life.

A large signal filter time can have a particularly devious effect. If the signal filter becomes the largest time constant in the loop, the filter time is the open loop time constant in the tuning equations. An increase in the signal filter time will enable an increase in the controller gain. Oscillations that already existed may have smaller amplitude from the attenuation of the actual oscillations. Operations may think the performance is actually better as the signal filter is increased smoothing out the picture. One clue to an excessive signal filter is an increase in dead time and the period of oscillation from an increase in the fraction of other time constants converted to dead time.

The unfiltered PV should be displayed to better reveal the actual noise. Also a faster execution of a module to provide this PV is warranted to prevent aliasing particularly when the PID execution time is increased to reduce the reaction to noise. The module execution time must be significantly less than half the noise period to prevent a larger than actual period and smaller than actual amplitude from aliasing.

An increase in execution time and signal filter time will increase the ultimate period, which is the best indicator of adverse results of increased dead time. Since the ultimate oscillation method of tuning is not practical, most practitioners are unaware of changes in the ultimate period. The relay oscillation method of auto tuning computes the ultimate period and gain. This auto tuner will show the effect. The effect of execution time may not be noticeable if a change in module execution time alters auto tuner evaluation of results and is less than other variations in real world applications.

While it is great being able to compute probable effects, in actual applications the unexpected should be expected. Plant operations have a great way of humbling the most knowledgeable process control engineer. There are no experts in the control room.

To help deal with the unknowns and unrealized consequences, some diagnostic techniques are offered in terms of determining if there is a tuning problem and what is the mostly likely correction needed.

Life is a balance. Proper tuning also involves a balance, in terms of the contribution of the proportional, integral, and derivative modes. Each mode has relative advantages and disadvantages as outlined in Appendix B. Problems develop when one mode dominates to the detriment of the overall objective. Diagnostic techniques and smart reset logic can maintain a balance.

5.1.2 OVERVIEW

The equation for the integrated error shows the effect of PID execution time and signal filter time for a given set of tuning settings. If an increase in the dead time from the execution time and filter time causes the total dead time to exceed the implied dead time from the tuning settings, the controller needs to be re-tuned to prevent the start of an oscillatory behavior.

The dead time from the PID controller is a fraction of the signal filter time plus half the execution time plus latency, which is the time delay from the start of the execution till the time when the output is sent. In older DCS, the time to complete all the calculations from the PV input to the PID output might not be completed until the end of the execution time interval, which corresponds to a latency that is equal to the execution time. In some cases, the DCS controller had negative free time and the output would not be available till the start of the next execution. The latency as a fraction of the execution time in a modern DCS is usually nearly zero except for high execution rates (e.g., 0.1 second).

The fraction Y of a small time constant converted to dead time is estimated by Equation 4.1a that is based on the ratio of the small to largest time constant in the loop. Hopefully, the largest time constant is in the process to slow down the actual process excursion from fast disturbances. If the signal filter becomes the largest time constant, the PID is no longer seeing the response of the actual PV and an illusion of better control may occur leading to further detrimental increases in the filter time.

If the true rate of change of the process is slow, a rate limit can be used instead of a filter to screen out noise. The rate limit block offers the advantage of not adding any dead time to the control loop. For example, the temperature of a large liquid volume may not change faster than 0.2 degrees per minute. A rate limit set for 0.1 degrees per minute will be able to screen out most measurement noise. The measurement without the rate limit must be available to operations and maintenance for calibration and troubleshooting.

The increase in loop dead time from an increase in the PID execution time or signal filter time will show up as an increase in the ultimate period. If the signal filter time is large enough to be considered the secondary time constant, the effect is even greater for integrating and runaway processes as revealed by the equations for the ultimate period in Chapter 4. Whereas the rate time can be set equal to the secondary time constant to compensate the detrimental effect, the best practice is not to create the secondary time constant in the first place by avoiding an excessive signal filter.

The pattern of a PID's response to a disturbance or a setpoint change can be used to diagnose what tuning setting is too large or small causing an unbalance in the contribution of the proportional, integral, and derivative modes. Rules of thumb are developed to point to the setting and direction of change needed.

The rules can be automated. Since the integral setting is the most frequent culprit, smart reset action logic is outlined to take advantage of pattern recognition. The user can choose to allow the logic to only increase the reset time, which is the most common correction needed and always leads to greater stability.

Equations are developed to detail the calculations needed for saturated PID outputs to get the output off of the output limit at the best time to minimize rise time but prevent overshoot. A deadband can be increased to provide more protection against overshoot.

5.1.3 RECOMMENDATIONS

1. The PID execution time should be large enough to minimize resolution or threshold sensitivity errors in slow loops.
2. The PID execution time should be small enough not to appreciably increase the peak or integrated error.
3. Use the equation for the integrated error to determine if the PID execution time and signal filter time are too large.
4. Estimate the increase in peak error as the PID execution time or filter time multiplied by the maximum rate of change of the process variable.
5. Minimize the use of signal filters particularly on processes with a fast integrating response (e.g., gas pressure) or a runaway response (exothermic reactor).
6. If the filter time becomes the largest time constant in the loop, use Chapter 6 equation to compute actual PV amplitude from measured amplitude.
7. For wireless devices, set the transmitter damping setting to keep measurement noise less than the trigger level to prolong battery life.
8. Set the signal filter just large enough so fluctuations in the controller output from noise do not cause a change in valve position or prime mover speed. For slow processes (e.g., liquid composition, temperature, and level) consider the use of a rate limit block instead of a filter to screen out noise without adding dead time. Display the measurement without the rate limit for operations and maintenance. Turn off the rate limit when the PID is in manual.
9. Use the pattern recognition techniques to determine if one of the PID settings is considerably out of balance with the other tuning settings.
10. Use an unfiltered PV in a module with a fast execution rate to show the actual amplitude and period of the noise for process and sensor analysis.
11. Make sure the execution time and filter time are large enough to improve the signal to noise ratio by reducing the reaction of the proportional and derivative modes to backlash, threshold sensitivity limit, resolution, and noise.
12. If the output is saturated, determine if the reset time is too small or large causing the output to come off of the output limit too late or soon, respectively. A smaller time constant keeps the contribution from the integral mode larger than the integral mode for longer time causing the output to be late coming off its limit.

The execution time must be less than half the oscillation period to prevent aliasing. The measured oscillation amplitude is less than the actual values when aliasing occurs. For two samples per oscillation, the samples would have to be exactly at the maximum and minimum peaks to see the period and amplitude, an unlikely event. Making sure the execution time is significantly less than the reset time and total loop dead time, is generally a more stringent requirement on the maximum execution time. Also while aliasing is undesirable from the viewpoint of analysis, aliasing of oscillations less than the ultimate period is not detrimental because the best a PID can do is not respond to the oscillation. For this situation, aliasing may actually help reduce the PID reaction to noise.

5.2 EXECUTION RATE AND FILTER TIME

PID execution time and filter time have a primary effect on integrated error well documented by the equations developed in Chapter 3. Not well recognized and adequately quantified is the secondary effect of execution rate and filter time on the behavior of the PID. An innovative term, "implied dead time", is developed that details the secondary effect not only for these controller dynamics but also for automation system dynamics in general. The term answers the general question: "When does an automation system add too much delay?"

A large signal filter can be insidious in deception. The response may look better in terms of the size of the deviation appearing smaller on a trend chart. The PID gain may even be able to be increased in some cases furthering the deception. Section 6.10 shows the equation to compute actual PV amplitude from the measured amplitude.

5.2.1 FIRST EFFECT VIA EQUATION FOR INTEGRATED ERROR

The update time considered here is the time between updates in any discontinuous signal. In this chapter the digital controller update time is the PID module or block execution time (Δt_x), whichever is largest. The digital controller scan time can be considered to be negligible if there is oversampling of input signals to prevent aliasing. The equations and principles developed here can be applied to the update time from a variety of sources including wireless device default update rate and analyzer cycle time (see Chapter 6). Equation 5.15 gives the PID controller dead time (θ_c) as a function of execution time (Δt_x) and latency, which in our case is a calculation time (Δt_c). Note that the terms "update rate" or "execution rate" is a time interval between updates or executions rather than the number of executions or updates per second.

The starting point is Equation 5.1 for the integrated error as a function of controller gain (K_c) and reset time (T_i) derived from the PI controller's response to a load upset in Chapter 3 (Equation 3.4a):

$$E_i = \frac{T_i + \Delta t_x + \tau_f}{K_c * K_o} * E_o \tag{5.1}$$

5.2.2 SECOND EFFECT VIA EQUATIONS FOR IMPLIED DEAD TIME

Note that Equation 5.1 shows the first effect of execution time on integrated error but does not give an indication of the additional effect on tuning and stability. If the execution time is appreciable compared to the implied loop dead time causing more oscillation, the increase in loop dead time from the execution time and latency should be estimated. The implied dead time depends upon controller tuning besides the original dead time. The innovative term is based on the observation that a slowly tuned controller behaves similar to a loop with a larger than actual dead time. The following derivation provides the equations needed to estimate the allowable execution time without an additional appreciable effect on integrated error for unmeasured disturbances.

The Lambda tuning equation for self-regulating processes to set a degree of transfer of variability in terms of an original dead time (θ_o) is:

$$K_c = \frac{T_i}{K_o * (\lambda + \theta_o)} \tag{5.2}$$

If Lambda is set equal to an implied dead time (θ_i) and the reset time is realized per Lambda tuning to be equal to the open loop time constant (largest time constant in the loop), Equation 5.2 becomes Equation 5.3 that expresses the controller gain as a function of the implied dead time and other dynamics instead of Lambda and reset time.

$$K_c = 0.5 * \frac{\tau_o}{K_o * \theta_i} \tag{5.3}$$

A detuned controller (e.g., lower controller gain) has the same load rejection capability as a loop with more dead time. The maximum performance for a given dead time is suggested by using the tuning for maximum controller gain in Equation 5.1 for the IAE. Thus, we can find out how much dead time is implied in a detuned controller by setting the detuned controller gain equal to Equation 5.3 for maximum controller gain.

If you set Equation 5.2 equal to Equation 5.3, and set the reset time equal to the open loop time constant ($T_i = \tau_o$) per Lambda tuning, and cancel terms, you end up with the following equation for implied dead time (θ_i) as a function of the Lambda (λ). Note that Lambda (closed loop time constant) for self-regulating processes is the Lambda factor multiplied by the open loop time constant ($\lambda = \lambda_f * \tau_o$).

$$\theta_i = 0.5 * (\lambda + \theta_o) \tag{5.4}$$

The amount of dead time available in the implied dead time for additional delay in the controller (θ_c) is:

$$\theta_c = \theta_i - \theta_o \tag{5.5}$$

If you substitute Equation 5.4 in Equation 5.5, then the max allowable control delay that causes no appreciable increase in the IAE for self-regulating processes is:

$$\theta_c < 0.5 * (\lambda + \theta_o) - \theta_o \tag{5.6}$$

Integrating processes exhibit the unusual behavior where a proportional-integral (PI) controller gain can be too low as well as too high for a given reset time. When the gain is too low, the approach back to setpoint is much slower than expected and incredibly slow rolling oscillations can develop. If the gain is too high, instability can result where the oscillations are more rapid and increase in amplitude. Equation 5.7 shows the inequality that must be satisfied to prevent a slow protracted response from too low a gain. Equations 1.5c through 1.5e give the inequalities that predict when slow oscillations start and get so severe the oscillations appear almost undamped. These equations reveal the counterintuitive relationship where if you increase the controller gain, you can decrease the reset time.

$$K_c * T_i \geq \frac{4}{K_i} \tag{5.7}$$

Equation 5.8 is the lowest integral time (reset time) that still satisfies the inequality preventing the start of a protracted response.

$$T_i = \frac{4}{K_i * K_c} \tag{5.8}$$

The Lambda tuning equation for controller gain is Equation 5.9 where Lambda is now the closed loop arrest time for integrating processes.

$$K_c = \frac{T_i}{K_i * (\lambda + \theta_o)^2} \tag{5.9}$$

Substitution of Equation 5.8 in Equation 5.9 gives Equation 5.10 without the reset time.

$$K_c = \frac{4 / K_c}{K_i^2 * (\lambda + \theta_o)^2} \tag{5.10}$$

Multiplication of both sides by controller gain to cancel this gain in the numerator and subsequently taking the square root, you end up with Equation 5.11 for controller gain.

$$K_c = \frac{2}{K_i * (\lambda + \theta_o)} \tag{5.11}$$

If the *near-integrator* or *pseudo-integrator* gain which is the open loop gain divided by the primary process time constant ($K_i = K_o / \tau_p$) is substituted into Equation 5.11, you have Equation 5.12 where the result of the substitution is set equal to Equation 5.3 for smooth max load rejection in terms of an implied dead time (θ_i).

$$K_c = \frac{2 * \tau_p}{K_o * (\lambda + \theta_o)} = 0.5 * \frac{\tau_p}{K_o * \theta_i} \tag{5.12}$$

Equating the two expressions and canceling out terms gives Equation 5.13 for the implied dead time.

$$\theta_i = 0.25 * (\lambda + \theta_o) \tag{5.13}$$

Substitution of Equation 5.13 in Equation 5.5 yields Equation 5.14, the max allowable update delay that causes no appreciable increase in the IAE for integrating processes.

$$\theta_c < 0.25 * (\lambda + \theta_o) - \theta_o \tag{5.14}$$

Equation 5.14 has the same form as Equation 5.6 with a coefficient of 0.25 instead of 0.5 implying a much greater sensitivity to update delay by integrating processes.

Now that we know the maximum allowable dead time from the controller, we need to know what this corresponds to in terms of controller execution time. The dead time from a digital PID controller (θ_c) is the sum of half the PID module or block execution time (Δt_x), whichever is larger, plus the latency and the equivalent dead time from the controller filter time (Equation 5.15). The fraction (Y) of the controller signal filter time converted to dead time can be computed from Equations 4.1a through 4.1c. The latency is the calculation time (Δt_c) between a PID measurement input received and a PID controller output sent. Normally this latency is negligible for a digital PID unless the execution time is small (e.g., 0.1 second) or there are time consuming calculations being done on the PID input signal from an analog input channel and the PID output signal delaying the output sent to a valve or VSD via an analog output channel.

$$\theta_c = 0.5 * \Delta t_x + \Delta t_c + Y * \tau_f \tag{5.15}$$

where

E_i = integrated error from load disturbance (e.u. * sec)
E_o = open loop error (load disturbance error with controller in manual) (e.u.)
K_c = controller gain (dimensionless)
K_i = integrating process gain (%/% per sec) (1/sec)
K_o = open loop gain (%/%) (dimensionless)
T_i = integral time (reset time) (sec)
Y = fraction of filter time convert to dead time as per Equations 4.1a through 4.1c (0 to 1)
λ = Lambda (closed loop self-regulating time constant or integrating arrest time) (sec)
θ_c = controller dead time (sec)
θ_i = implied loop dead time from PID tuning (sec)
θ_o = original loop dead time (sec)
τ_f = signal filter time constant (sec)
τ_o = open loop time constant (largest time constant in loop) (sec)
τ_p = primary process time constant (sec)
Δt_c = PID Module input to output channel calculation time (latency) (sec)
Δt_x = PID Module or block execution time, whichever is larger (sec)

An alternative to a signal filter for processes with slow rates of change seen in the chemical industry is the simple insertion of a rate limit function block on the PV before it becomes the PID input. The rate limit is set to be faster than the fastest actual process PV rate of change. The rate limit may be set to be faster in one direction than the other, which is useful for temperature processes that respond faster to heating than cooling. PV rate limits are particularly effective for batch and level loops since signal lags have a greater detrimental effect on integrating processes. Level loops are prone to noise due to turbulence. In one case rotating gear teeth of a finisher passed through the beam path to a nuclear level detector. There is essentially no delay or detrimental effect on the PID response of a rate limit block to true processes changes. The PV without the rate limit is displayed for operations and maintenance and may be configured to be off when the PID is in manual, remote output, and output tracking.

5.3 SMART RESET ACTION

Smart reset action can get the PID output off the output limit from an estimate of a rise time that is the time the PV takes to get within a specified percent deadband ($\%DB$) about setpoint. The rise time (T_r) is simply the current error taking into account the deadband (E_{db}) divided by ramp rate (R_r). The ramp rate and rise time are used to compute the relative contribution of each PID mode. To get the output off of the output limit, the reset time is increased to make contribution of the integral mode less than contribution of the proportional and derivative modes. Equations 5.16a through 5.16h show the derivation of the calculation that sets the reset time based on the user choice of deadband and the consequential error that causes PID output to come off limit.

$$E_{db} = \left| \%SP - \%PV \right| - \%DB \tag{5.16a}$$

$$P_{PID} = K_c * \left| E_{db} \right| \tag{5.16b}$$

$$I_{PID} = K_c * (1/T_i) * \left| E_{db} / 2 \right| * T_r \tag{5.16c}$$

$$D_{PID} = K_c * (T_d) * \left| E_{db} / T_r \right| \tag{5.16d}$$

$$R_r = \left| \Delta\%PV / \theta_o \right| \tag{5.16e}$$

$$T_r = \left| E_{db} / R_r \right| \tag{5.16f}$$

$$I_{PID} \le P_{PID} + D_{PID} \tag{5.16g}$$

$$T_i \ge (T_r / 2) / (1 + T_d / T_r) \tag{5.16h}$$

where

$\%DB$ = deadband for smart reset (%)
D_{PID} = derivative mode contribution to PID output (%)
E_{db} = error less deadband (%)
K_c = controller gain
I_{PID} = integral mode contribution to PID output (%)
P_{PID} = proportional mode contribution to PID output (%)
$\%PV$ = process variable (%)
R_r =.ramp rate (%/sec)
$\%SP$ = setpoint (%)
T_r = rise time (sec)
T_d = derivative time (rate time) (sec)
T_i = integral time (reset time) (sec)
$\Delta\%PV$ = delta PV in total loop dead time for ramp rate calculation (%)
θ_o = total loop dead time (sec)

Once an output has come off of the output limit, smart action can prevent overshoot or faltering of approach to setpoint whether due to a setpoint change or a load disturbance. If the PV predicted two dead times into the future that is going to be within or exceeding the deadband, the reset time is increased over the next dead time. If the ramp rate is faltering or even reversing direction in the approach, a single correction in made to the reset time where the reset time is decreased by a fraction (e.g., 0.2). No further correction is made until the situation occurs again for another setpoint change. The low limit on the reset time must be set carefully and conservatively since a lower reset time is destabilizing.

An essential detail in the implementation of adaptive reset is the use of a dead time block to compute a continuous train of ramp rates and future values every execution of the dead time block with a good signal to noise ratio. The dead time block creates an old PV one loop dead time in the past. The old PV subtracted from the new PV is a delta PV that when divided by the dead time is the ramp rate. The delta PV multiplied by a time interval and added to the new PV is a predicted PV that time interval into the future. The time interval for self-regulating processes can be as small as one dead time because the process is decelerating. For integrating, the time interval needs to be at least 1.5 times the dead time to help prevent overshoot. For runaway processes, the time interval may need to be at least twice the dead time to prevent overshoot from acceleration.

5.4 DIAGNOSIS OF TUNING PROBLEMS

If the oscillations amplitude is not constant either initially or continually, PID tuning is creating or aggravating a problem. Since growing oscillations represent instability and must be addressed, the more common case is decaying oscillations (underdamped response). The following steps can help locate and mitigate or eliminate the problem.

1. If the oscillations start to decay but then settle into a fixed amplitude oscillation, the source is a batch operation, or a limit cycle from deadband, threshold sensitivity limit, or resolution limit in the valve or measurement. PID tuning is making the problem worse by creating a larger amplitude decaying oscillation. The causes of a limit cycle also increase the total loop dead time for a process input disturbance. While the fixed amplitude oscillation cannot be eliminated by tuning, the test results of Chapters 6 and 7 shows the initial oscillation can be reduced by increasing the reset time.
2. If PID tuning does not affect the period of the constant amplitude oscillation, the source is a batch operation such as a centrifuge rather than a limit cycle.
3. Note that if another disturbance occurs before the loop settles into the limit cycle, the oscillations may not be recognizable as a limit cycle. For very fast process responses without filtering, the square wave will appear in the measurement even if the problem originates in the valve. Otherwise the limit cycle source is the valve for a readback square wave and is the measurement for a PV square wave. Note that the data history needs to be done without interpolation in order to show a square wave. If the limit cycle amplitude decreases as the PID gain increases, the cause is backlash
4. If an oscillation in the PID output is upsetting another loop regardless of the cause, the PID output response can be made smoother by decreasing the PID gain for self-regulating processes and increasing the PID reset time for near-integrating, true integrating, and runaway processes.
5. If the oscillations go away when the PID is in manual, then the source is the tuning in another PID affecting this PID. If this is the case, PID loops are individually put in manual until the oscillations disappear to locate the culprit.
6. If there is no periodic disturbance or limit cycle, the problem is improper tuning.

7. If the period of the oscillation is less than the ultimate period (e.g., three-fourths) or there are multiple oscillations in the approach to setpoint, the cause is most likely a PID rate time that is too large.
8. If the period of the oscillation is slightly larger than the ultimate period for a negligible rate time or there is a single hesitation (faltering) in the approach to setpoint, the cause is most likely a PID gain setting that is too large.
9. If the period of the oscillation is appreciably larger than ultimate period (e.g., twice) or there is an overshoot of the setpoint, the cause is most likely a PID reset time setting that is too small.
10. If the period of the oscillation is much larger than the ultimate period (e.g., ten times), the cause is the product of the PID gain and reset time being too low as expressed by Equation 1.5e for a near-integrating, true integrating, or runaway process. While the fix as per the equation is making either the PID gain larger or the reset time larger, increasing the PID gain may not be advisable because of the possibility of a miss-diagnosis, noise, or abrupt PID output changes being disruptive. The best correction in most cases is an increase in the reset time.

5.5 FURNACE PRESSURE LOOP EXAMPLE (NEAR-INTEGRATING)

More information on the source of the solutions can be obtained by going to the portion of the chapters denoted by the equation numbers shown in the solutions.

Given:

(Information from Example 4.6.4)

a. Set point is 5 inches water column (w.c.) gage.
b. Measurement range is 0 to 10 inches w.c. gage.
c. Furnace volume is 10,000 ft^3 (V_f).
d. Quench volume is 1,000 ft^3 (V_q).
e. Scrubber volume is 1,000 ft^3 (V_s).
f. Furnace inlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_f).
g. Quench inlet flow resistance pressure drop is 5 inches w.c. (ΔP_q).
h. Scrubber inlet flow resistance pressure drop is 10 inches w.c. (ΔP_s).
i. System outlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_o).
j. Flue gas flow is 1,000 scfm (F_f).
k. Atmospheric pressure is 408 inches w.c. (P_a).
l. Disturbance is a rapid 20 percent increase in inlet pressure (ΔP).
m. Inlet pressure (discharge of forced draft fan) is 15 inches w.c. (P_i). (New information)
n. PID controller filter time is 0.2 second (τ_f)
o. PID controller execution time is 0.2 second (Δt_x)

Find: ultimate limit for peak and integrated errors for an unmeasured disturbance.

Solution:

a. Calculate the new dead time and ultimate period:

From the solution to Example 4.6.4:

$\tau_s = 1.5$ sec
$\theta_p = 0.2$ sec
$K_i = 0.14$ per sec
$E_o = 0.4$ inches per sec

The dead time from the PID controller is the fraction of the signal filter converted to dead time and half of the module execution time since the result is available immediately after start of execution so that latency time is zero ($\Delta t_c = 0$).

$$\theta_c = 0.5 * \Delta t_x + \Delta t_c + Y * \tau_f = 0.5 * 0.2 + 0.0 + 0.5 * 0.2 = 0.2 \sec \tag{5.15}$$

The new total loop dead time is the controller dead time added to the process dead time

$$\theta_o = \theta_p + \theta_c = 0.2 + 0.2 = 0.4 \sec$$

The new ultimate period for the total loop dead time is:

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o}\right)^{0.65}\right] * \theta_o = 4 * \left[1 + \left(\frac{1.5}{0.4}\right)^{0.65}\right] * 0.4 = 5.8 \text{ sec} \tag{4.15a}$$

b. Calculate the ultimate limit for peak and integrated errors:

$$E_x = \frac{1.8 * T_u}{2 * \pi} * E_o = \frac{1.8 * 5.8}{2 * \pi} * 0.4 = 0.66 \; \textit{inches w.c.} \tag{4.20b}$$

$$E_i = \frac{T_u}{2 * \pi} * T_u * E_o = \frac{5.8}{2 * \pi} * 5.8 * 0.4 = 2.2 \; \textit{inches w.c.} * \sec \tag{4.19b}$$

Conclusions: The peak error increased by 65 percent and the integrated error increased by 140 percent since the original loop dead time of 0.2 second was doubled to 0.4 second by the inclusion of the controller dead time. The controller execution time should be decreased to its minimum (0.1 second) and the filter time eliminated if at all possible for gas pressure control particularly when the integrating process gain is high and the scale range is low.

5.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in load flow at the process input (ΔF_L). Table 5.1 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The test cases use aggressive tuning settings computed with the short cut method to maximize disturbance rejection. Since these settings are right on the edge of causing an oscillation, the second effect of controller dynamics occurs besides the first effect.

Figures 5.1a to 5.1f show the effect of PID execution time (Δt_x) on the ability of the PID controller to deal with a load disturbance. DCS terminology calls the time in between

Table 5.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
5.1 a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 10\%$	Execution time
5.1 c, d	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Execution time
5.1 e, f	True integrating	0.01 1/sec	10	—	$\Delta F_L = 20\%$	Execution time
5.2 a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 10\%$	Filter time
5.2 c, d	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Filter time
5.2 e, f	True integrating	0.01 1/sec	10	—	$\Delta F_L = 22\%$	Filter time

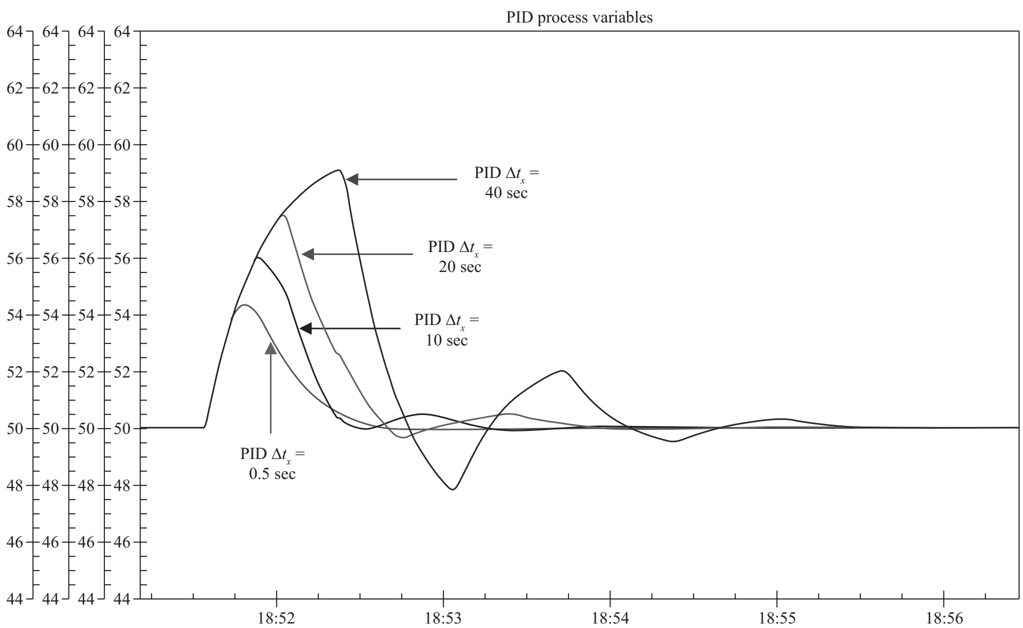


Figure 5.1a. Effect of PID execution time (Δt_x) on *actual* PV for *moderate self-regulating* process.

executions an execution rate. For these test cases, the PID execution rate was the same as the module execution rate. In some DCS, the PID execution rate can be set as a multiple of the module execution rate. The time in the module execution before the start of the PID execution and the time to the finish the PID algorithm execution are both negligible so that the latency is zero. The filter time is zero in these tests.

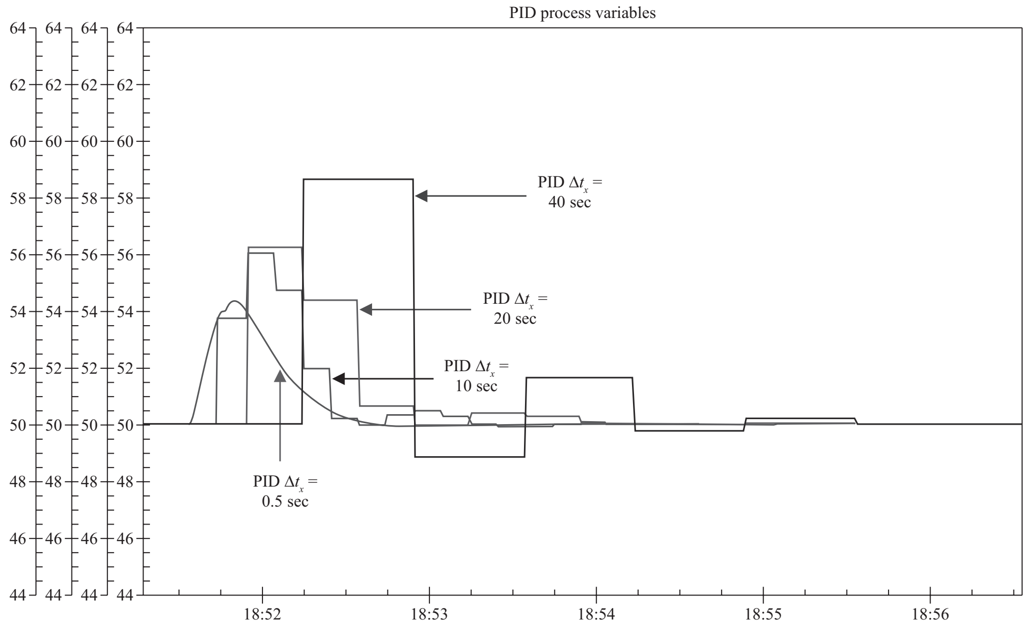


Figure 5.1b. Effect of PID execution time (Δt_x) on *measured* PV for *moderate self-regulating* process.

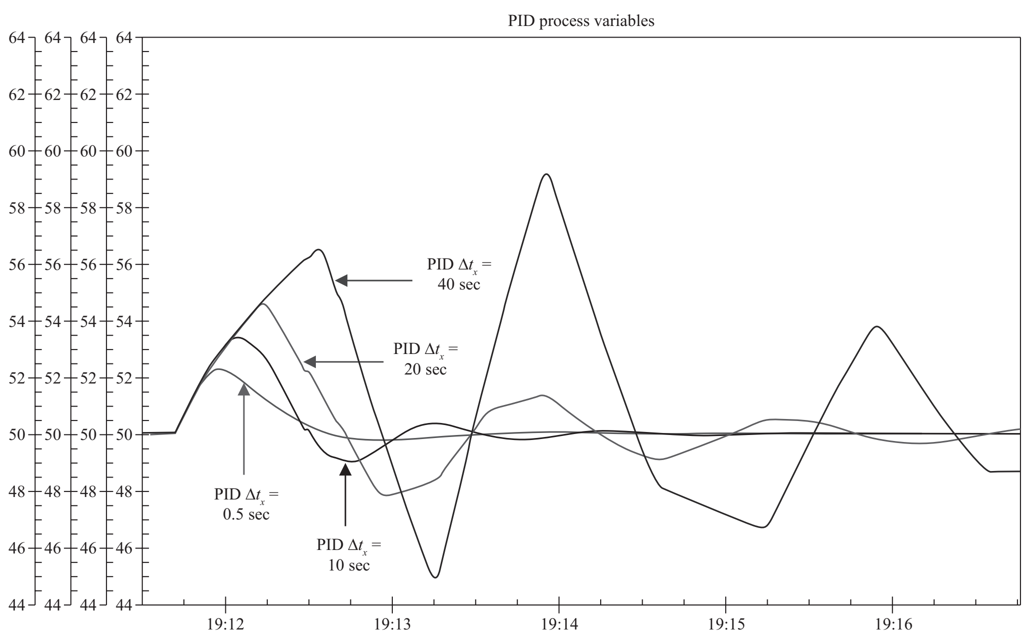


Figure 5.1c. Effect of PID execution time (Δt_x) on *actual* PV for *near-integrating* process.

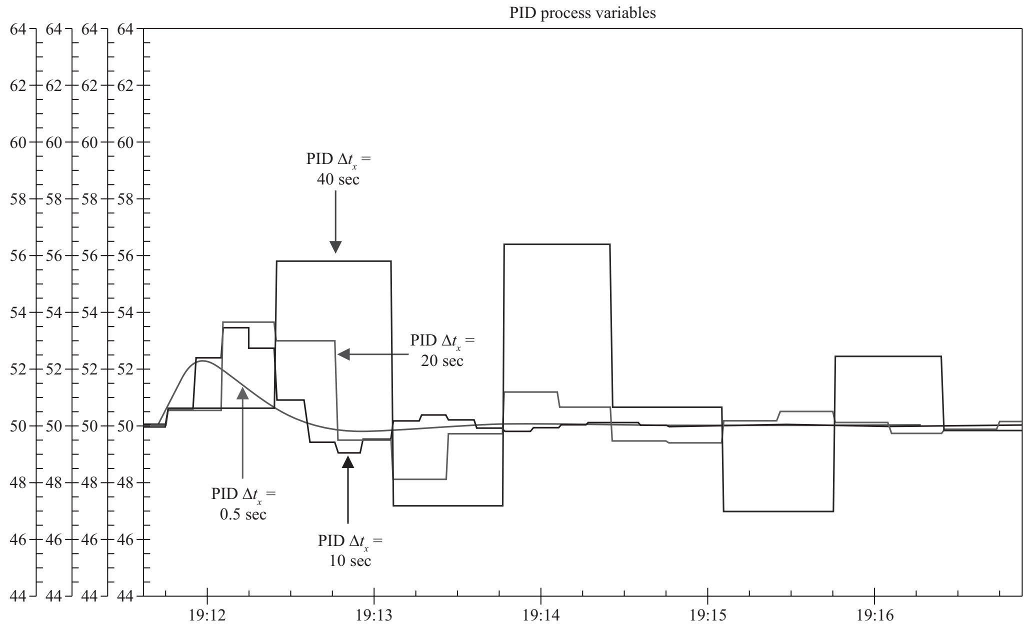


Figure 5.1d. Effect of PID execution time (Δt_x) on *measured* PV for *near-integrating* process.

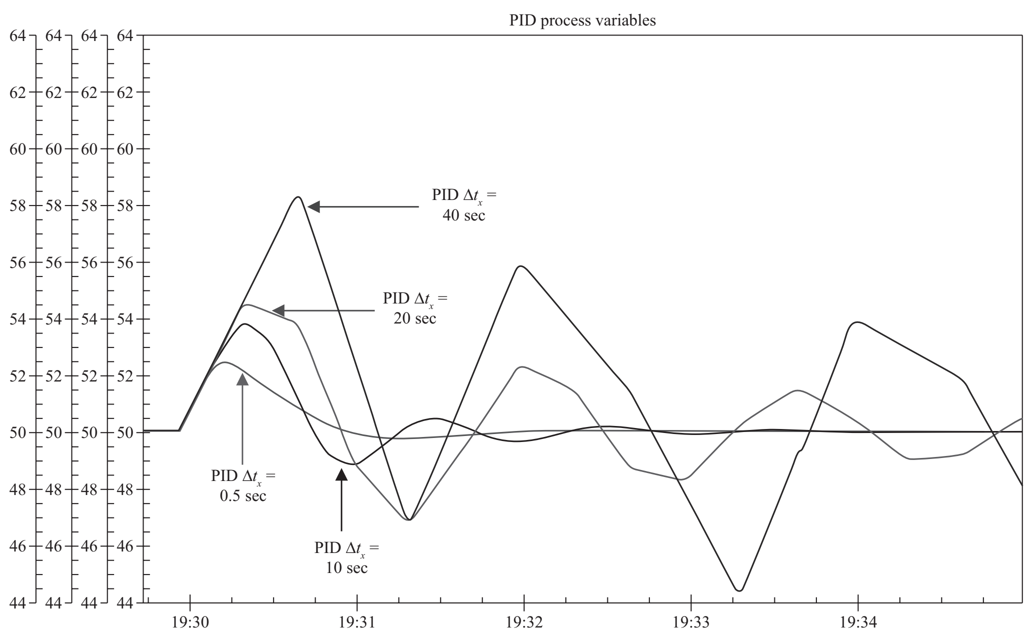


Figure 5.1e. Effect of PID execution time (Δt_x) on *actual* PV for *true integrating* process.

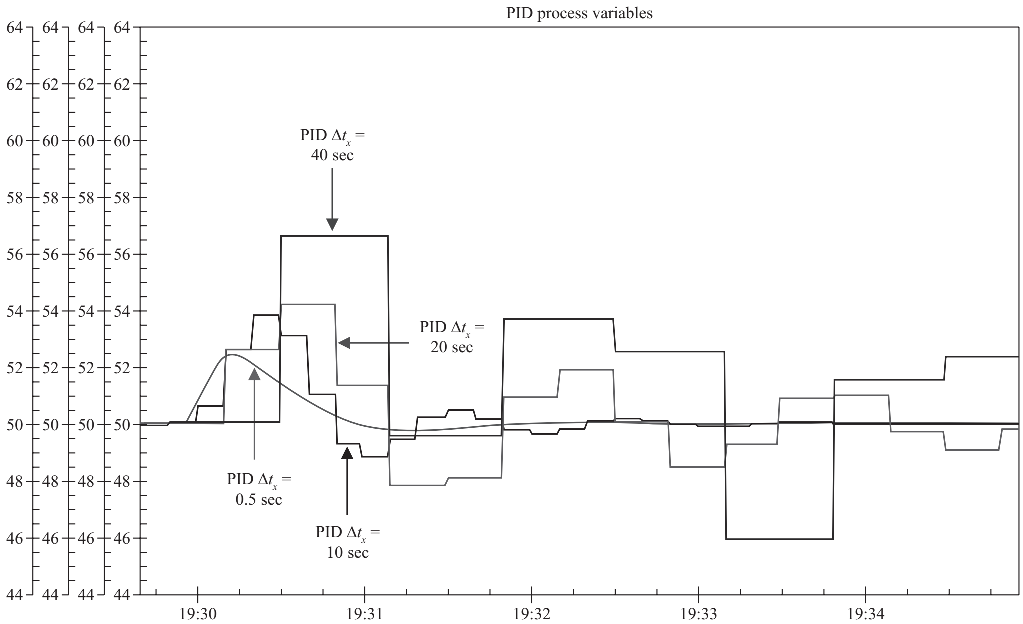


Figure 5.1f. Effect of PID execution time (Δt_x) on *measured* PV for *true integrating* process.

The additional dead time is approximately half the PID execution time since there is no latency. The base case uses a fast execution time of 0.5 seconds which contributes about one-fourth of a second to the total loop dead time of 10 seconds. The slowest execution of 40 seconds doubles the existing dead time.

The effect of execution time becomes more destabilizing as the process loses negative feedback (loses self-regulation). For the near-integrating process and the slowest PID execution the second peak is greater than the first due to timing but the peak to peak amplitude of subsequent peaks decay. For the true integrating process and the slowest PID execution, the oscillation peak to peak amplitude decay is more gradual giving the appearance of nearly undamped oscillations.

For the fast execution time of 0.5 seconds, the response of the actual and measured PV is smooth and rounded. Ramps and a saw tooth in the actual PV and steps and a square wave in the measured PV are noticeable for an excessively large execution time if the data history does not have interpolation.

Figures 5.2a to 5.2f show the effect of PID filter time (τ_f) on the ability of the PID controller to deal with a load disturbance. The effect is similar to a sensor time constant or transmitter damping setting for analog and digital transmitters with a negligible update time (e.g., electronic and Fieldbus communications). The PID execution time is 0.5 seconds in these tests.

The additional dead time is a fraction of the time constant based on the ratio of the filter time to the process time constant for the moderate self-regulating process. The 20 second filter time is equal to the process time constant for this process. Based on Figures 5.2a and 5.2b the additional dead time is 2.5, 4.0, and 6.0 seconds for the 5, 10, and 20 second filter times, respectively. For the near and true integrating processes, the effect is better approximated

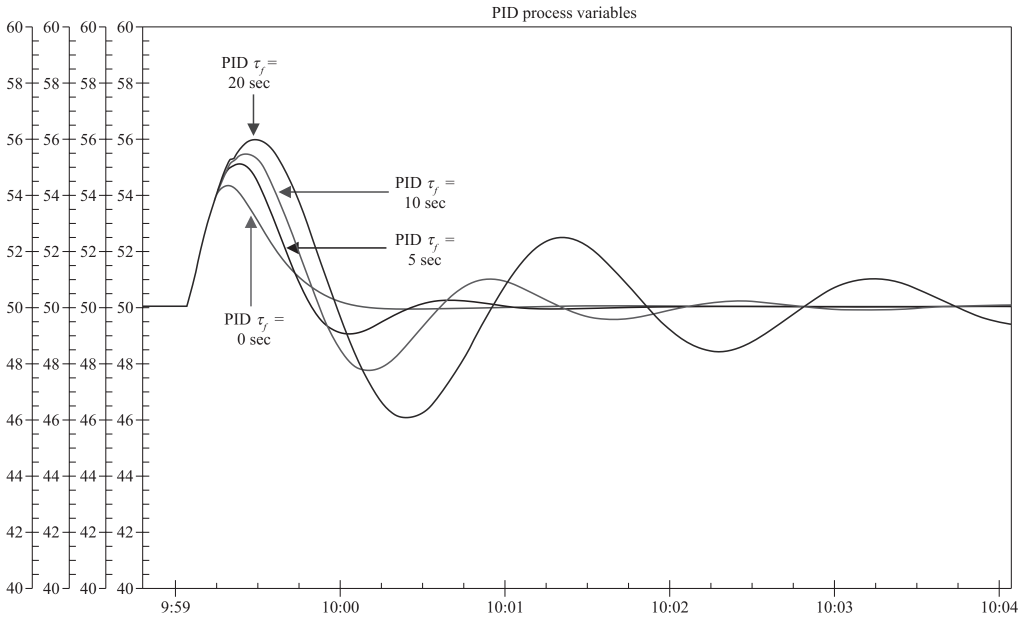


Figure 5.2a. Effect of PID filter time (τ_f) on *actual* PV for *moderate self-regulating* process.

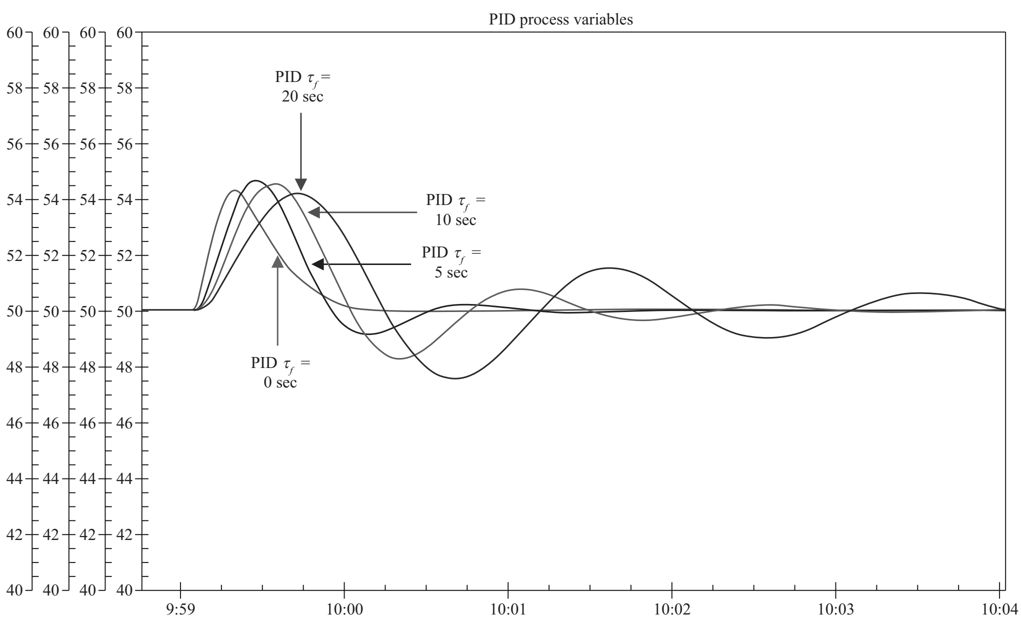


Figure 5.2b. Effect of PID filter time (τ_f) on *measured* PV for *moderate self-regulating* process.

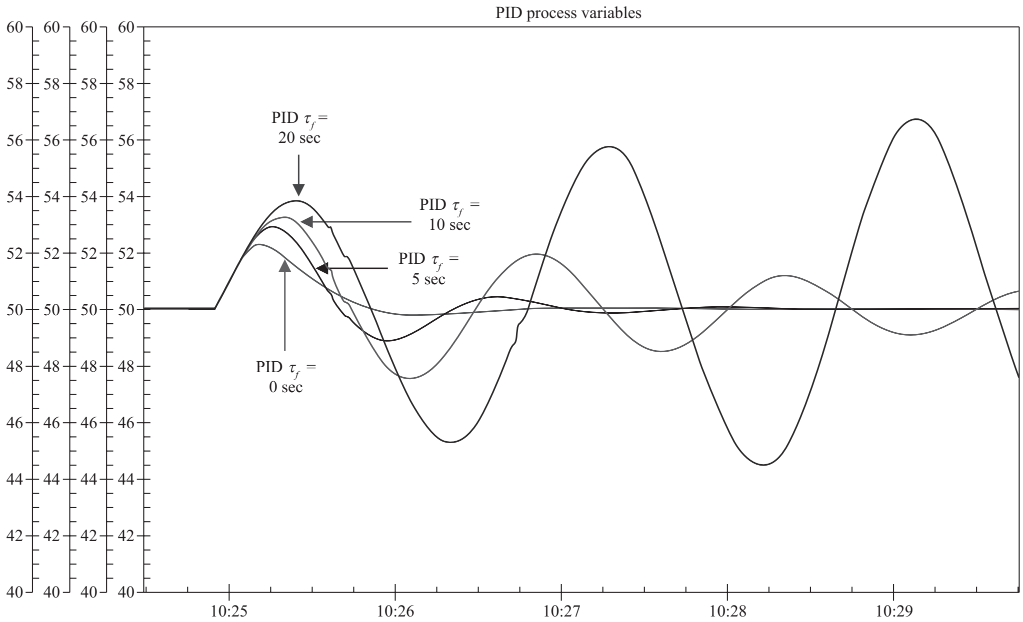


Figure 5.2c. Effect of PID filter time (τ_f) on *actual* PV for *near-integrating* process.

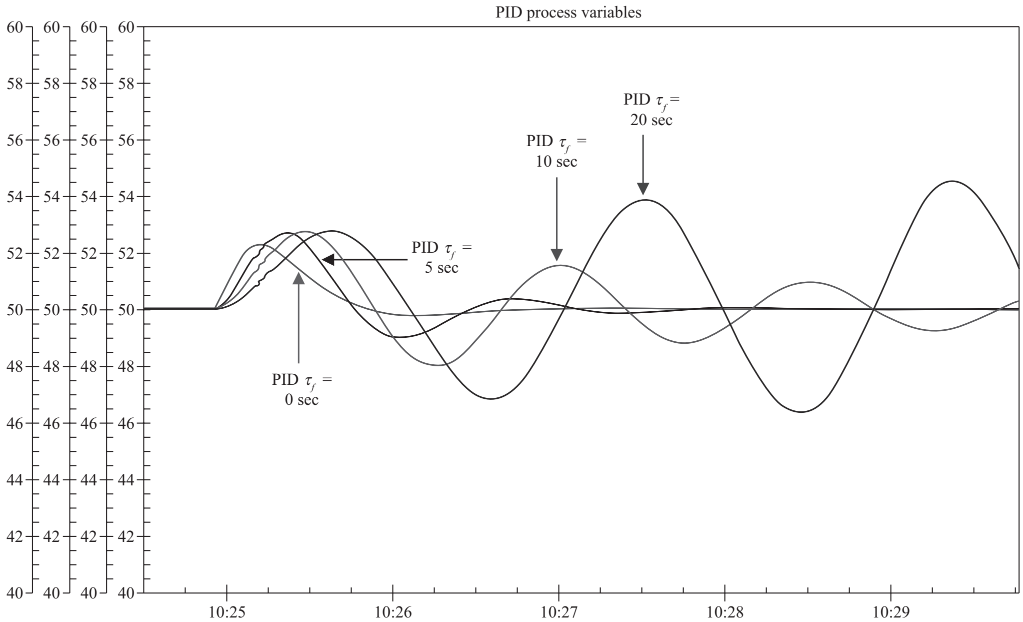


Figure 5.2d. Effect of PID filter time (τ_f) on *measured* PV for *near-integrating* process.

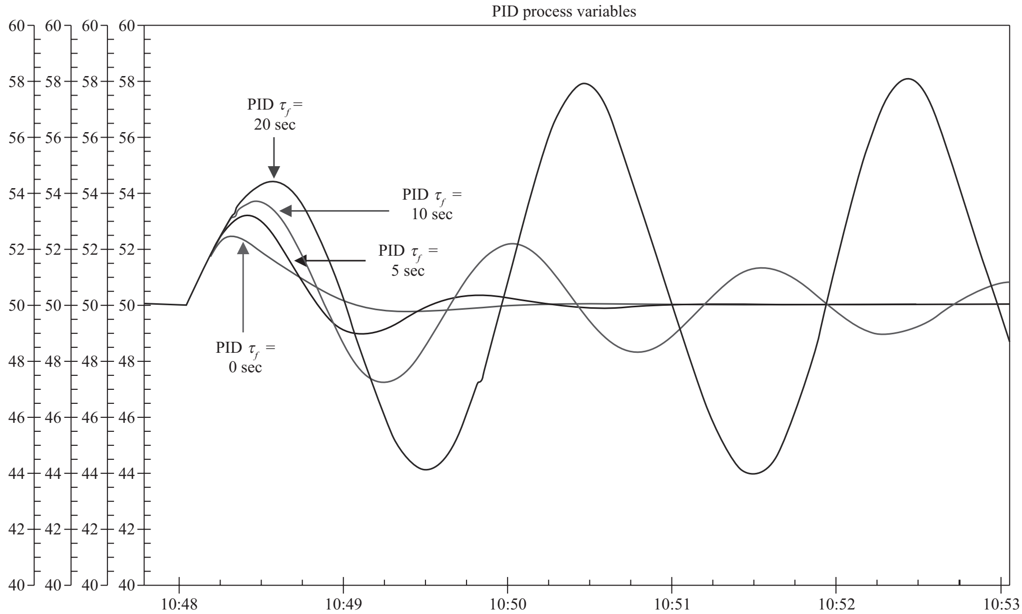


Figure 5.2e. Effect of PID filter time (τ_f) on *actual* PV for *true integrating* process.

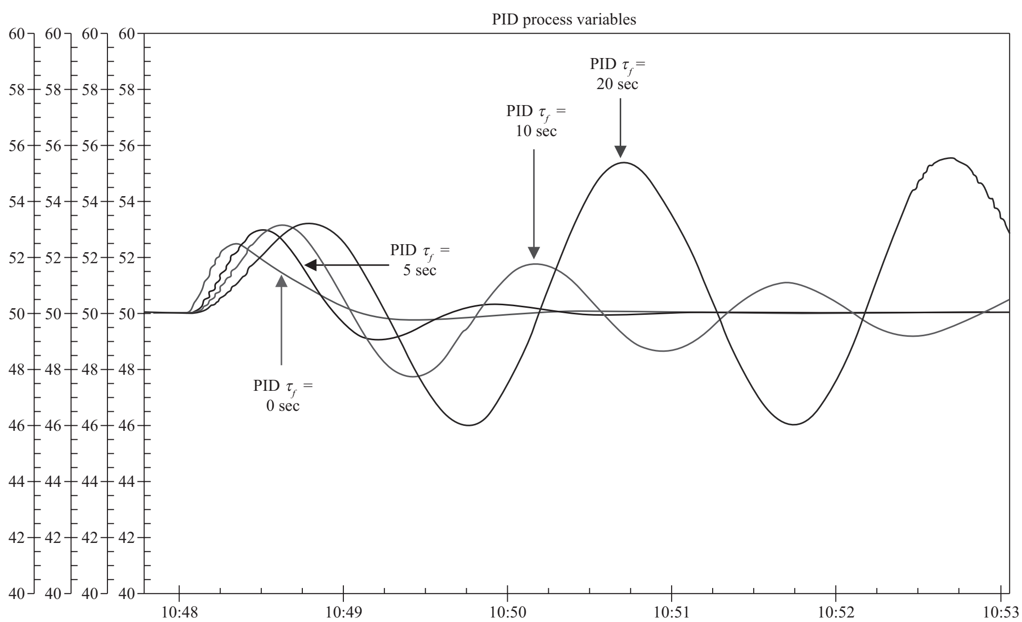


Figure 5.2f. Effect of PID filter time (τ_f) on *measured* PV for *true integrating* process.

by taking the filter time as a secondary time constant and computing via Equation 4.15a the increase in the ultimate period from 40 seconds to 65, 80, and 100 seconds, respectively.

As was the case for PID execution time, the effect of a filter time increases as the process loses negative feedback. Also for a given value, the effect of a filter is greater than execution time in terms of destabilizing the process. A filter time of 20 seconds causes the near and true integrating processes to be unstable. Unlike the effect of execution time, the measured PV oscillation is smooth and rounded.

The measured PV amplitude starts to show the attenuating effect of filtering for the larger filter time of 20 seconds. Equation 6.7a can be used to estimate this attenuation. If the PID was detuned so that the oscillations do not start for a larger filter time, the attenuation and deception is greater. A large filter time can give the appearance of being beneficial due to a smaller measured peak for a disturbance. Note that the equation for peak error uses the largest time constant wherever this time constant exists in the loop. If this open loop time constant is in the measurement, trend charts of the measured PV will give the appearance of tighter control. The unfiltered PV should always be trended to give a better indication of actual PV to better show loop performance and noise.

Some consultants advocate never using a PID filter. This book takes the approach that the PID filter time should be minimized and be set just large enough to keep the peak to peak amplitude of the PID output fluctuations from noise just less the valve deadband, threshold sensitivity, and resolution limit so that the valve does not respond to noise upsetting loops and causing excessive valve dither and wear. Note that in fast processes what appears to be noise are actually disturbances. The addition of a filter can be disastrous. If a loop requires a PID execution time of 0.1 second or less, the PID filter time should be zero and the transmitter damping minimized. Loops for surge prevention, furnace pressure control, and liquid or polymer pressure control are in this category. Liquid and polymer flow control could benefit as well from a faster controller if tight flow control is needed and possible by virtue of low measurement noise and a valve pre-stroke dead time and stroking time less than 0.1 and 1 second, respectively.

Too small of an execution time can cause a large bump in the PID output from the derivative mode's reaction to a step change in the measurement from a threshold sensitivity limit or resolution limit. In 1980s, DCS the replacement of temperature transmitters with wide range thermocouple (TC) input cards caused such a big bump in liquid batch reactor temperature PID output from a resolution limit that derivative action could no longer be used. The TC cards had 12 bits (one bit for sign and 11 bits for resolution) that resulted in a resolution limit of 0.4°F for an 800°F TC card span. Since these reactors have a large secondary time constant and can be highly exothermic, the loss of derivative action was a major concern. The temporary fix until narrow span transmitters were installed, was to increase the execution time so that the step change in temperature was not computed by derivative action as occurring in a few seconds. A better fix available today is to use an enhanced PID so that the time interval used in the derivative calculation is the elapsed time from the last change.

A signal filter may also be needed to help smooth out the step in the PID output from the proportional mode's response to a step change in the PV from a measurement threshold sensitivity limit or resolution. In today's DCS and automation systems, the use of 16 bit input cards and the avoidance of mechanical sensors has largely eliminated this issue.

While step changes from the proportional mode reaction to disturbances are largely gone, the tendency to set all PID execution times as one second leads to excessive bumps in the PID output from the derivative mode. The derivative filter can help mitigate this problem as the alpha parameter (faction of rate time that is the filter time) is increased. The problem tends to be worse in loops that have potentially the greatest benefit from rate action. Temperature loops on liquid reactors and distillation columns can have a large gain and rate time setting. If a change that normally takes minutes to become large enough to change the PV seen by the PID due to the inherently slow process response is considered to have unrealistically occurred in the PID execution time (e.g., 1 second), the derivative mode reaction (product of gain and rate time multiplied by computed PV rate of change) is excessive.

KEY POINTS

1. If the additional dead time from a PID execution time and filter time does not cause the new total dead time to be larger than the implied dead time from tuning, only the first effect is at play and Equation 5.1 predicts the deterioration in the integrated error for an unmeasured load disturbance.
2. If the total dead time is larger than the implied dead time from tuning, the loop will become oscillatory.
3. The deterioration is greater for near-integrating and true integrating processes and potentially dangerous for runaway processes. The increase in ultimate period should be estimated.
4. The deterioration for moderate self-regulating and dead time dominant processes is less than 5 percent for a PID execution time and filter time less than the 10 and 5 percent of the PID reset time, respectively and is solely from the first effect (no second effect).
5. The effect of a PID filter time is similar to the effect of a sensor time constant or transmitter damping setting if the measurement update time is negligible.
6. The peak error and oscillation amplitude of the actual PV is larger than the measured PV if the filter time is greater than the original loop dead time.
7. The oscillation period will be larger and the amplitude smaller from aliasing when the module execution time is greater than half of the oscillation period.
8. A trend of the unfiltered PV from a fast module execution shows the actual size of disturbances and noise important for automation system and process analysis.
9. A minimal filter time prevents valve dither and VSD speed changes from noise.
10. An enhanced PID or a PID execution time much larger than the percent measurement threshold sensitivity or resolution limit divided by the percent change per second in the actual PV is needed to prevent an excessive bump in the PID output from the derivative mode's computation of the PV rate of change.
11. A signal filter is needed for older DCS and mechanical sensors to prevent a step change in the PID output from the proportional mode's response to a step change in the PV from a sensor threshold sensitivity or analog input resolution limit. Alternatively, an analog output or secondary PID setpoint rate limit can be added to provide move suppression if the PID has external reset feedback of the PV of the output or PID block and has the positive feedback implementation of integral action. Older DCS may not have these features.
12. A rate limit block can screen out noise without adding dead time or a secondary time constant to the loop. A rate limit is particularly effective for large liquid volumes because the true rate of change of the process is slow and a secondary time constant is devastating to a near-integrating and true integrating process. The PV without the rate limit is displayed for operations and maintenance.

CHAPTER 6

Effect of Measurement Dynamics

6.1 INTRODUCTION

Measurements provide the window into the process. Smart transmitters have made the view clearer. Installation and operating condition effects are compensated for enabling the transmitter to have an installed accuracy close to the sensor capability. The main remaining considerations are sensor nonidealities such as drift, lag, noise, rangeability and threshold sensitivity and discontinuous updates from analyzers and to a lesser extent from wireless devices. The transportation delay in getting the process to the sensor also needs to be accessed and minimized.

6.1.1 PERSPECTIVE

Filters from signal processing and transmitter damping adjustments have an effect similar to the signal filters in controllers detailed in Chapter 5. The equation for the integrated error shows the effect of a measurement filter for given settings of controller gain and reset time. There are additional effects in terms of an increase in loop dead time and deception if the measurement filter becomes the largest time constant in the loop as discussed in Chapter 5. Here we present a simple equation to estimate the attenuating effect of time constant. The equation can be used as the basis for setting transmitter damping settings to keep noise less than default trigger level of wireless transmitters. The equation can also provide an estimate of the actual amplitude from the filtered amplitude.

There are many sources of measurement lags and pure delays. Since we live in a digital world and are entering into wireless control, we need to consider the additional dead time from discontinuous updates (periodic updates rather than continuous updates from analog devices with inline sensors). The practitioner needs to know when these are a concern.

The same approach developed in Chapter 5 is used to estimate the implied dead time from the existing tuning settings. If the increase in dead time from measurement dynamics causes the total dead time to become greater than the implied dead time, the controller will become oscillatory unless retuned. The deterioration in loop performance from the required slower tuning settings can be estimated from the equation for the integrated error.

For at-line analyzers, the total dead time from the sample transportation delay, analyzer cycle time, and multiplex time have a major effect on loop performance because of the large dead time introduced. The loop may in fact become dead time dominant where the dead time is much larger than the process time constant.

As a process has progressively more internal nonself-regulation, a measurement time constant (lag) has a greater effect especially when this lag becomes large enough to be considered a secondary time constant. Thus, measurement lags are more detrimental in integrating than self-regulating processes and are more detrimental in runaway than integrating processes. For runaway processes, the window of allowable controller gains can close making the process unstable for all tuning settings.

Resolution and threshold sensitivity limits can create a dead time or cause a kick in the controller output from derivative action. If the rate of change of the process variable is extremely slow (e.g., temperature on large volumes), the dead time and kick can be large because such processes tend to have a large PID gain and a rate time setting. Even worse is the triggering of a false update from noise.

The wide range and 12 bit A/D of thermocouple (TC) input cards in 1980s vintage Distributed Control System (DCS) and sadly in some modern day heating, ventilation, and air conditioning (HVAC) systems have a resolution limit of about 0.25°C causing deterioration in the signal to noise ratio and kicking in the output so severe that derivative action cannot be used and the controller gain had to be decreased, further degrading temperature control. Fortunately, this problem is mostly history but may offer some guidance for wireless transmitter settings.

A small judiciously set threshold sensitivity limit can have a beneficial effect when properly used to screen out noise. The addition of this limit prevents limit cycles caused by resolution limits, deadband, and stick-slip by preventing noise triggering an update of the enhanced proportional-integral-derivative (PID). The limit can also enable a higher controller gain by preventing noise from being amplified by the PID gain causing dither (high frequency oscillations) of the valve or variable speed drive (VSD). For near-integrating, integrating, and runaway processes, noise can be the primary constraint as to how high a controller gain can be used.

6.1.2 OVERVIEW

The effects of measurement dynamics are similar to the effects of controller dynamics except there are often many more sources of lags and delays and the consequences are generally more severe. Delays are the result of transportation and discontinuous updates from periodic sampling and cycle times in digital signals, wireless devices, and analyzers. The dead time from discontinuous signals is the update time plus any latency. For analyzers, the latency is the cycle time since the result is available at the end of the analysis cycle time. The resulting dead time is 1.5 times the cycle time. There can be additional dead time from sample transportation and from multiplexing.

Equations can estimate the increase in dead time. If the total dead time as a result of an increase in measurement lags or delays becomes larger than the implied dead time for the present tuning settings, oscillations will develop and the controller needs to be tuned for a slower response.

Equations can also estimate the kicking in a PID output from discontinuous updates. The equations show that wireless trigger level should be small enough and the update rate large enough to increase the signal to noise ratio for slow processes.

A simple equation enables the setting of a filter just large enough to reduce the adverse effects of noise and to predict the behavior of the actual process variable before filtering.

Equations in Chapter 4 for the ultimate period can be used to determine the adverse effect of the dynamics on the ultimate limit to loop performance. More specifically, equations are given for runaway processes to show how the ratio of the maximum to minimum controller gain changes with temperature measurement lag. The ratio should be around 10 to allow for nonlinearities and nonidealities. For a ratio of 1 the maximum and minimum controller gains are the same and the window of allowable controller gains closes. The open loop unstable process is closed loop unstable regardless of tuning.

6.1.3 RECOMMENDATIONS

1. The update time interval should be large enough (update rate slow enough) to minimize resolution or threshold sensitivity errors (particularly in slow loops). The goal here is to make the real change in the process approach or exceed the resolution or threshold sensitivity to improve the signal to noise ratio.
2. The update time interval should be small enough (update rate fast enough) to minimize the peak and integrated errors (particularly in fast loops). The goal here is to minimize the increase in total loop dead time.
3. Use the equation for the integrated error to determine if the digital device and wireless update rate and measurement filter time are too large.
4. Estimate the worst case increase in peak error as the sum of the digital device and wireless update rate (update time interval) plus the measurement filter time multiplied by the maximum rate of change of the process variable.
5. If the measurement lag becomes the largest time constant in the loop, use given equation to compute actual process variable amplitude from measured amplitude.
6. Set the transmitter damping just large enough so fluctuations in the controller output from noise do not cause a change in valve position or prime mover speed or an unnecessary wireless transmitter update.
7. Estimate the increase in total loop dead time from measurement lags, delays, and discontinuous updates (e.g., analyzers, digital devices, and wireless).
8. If the discontinuous update time becomes larger than the process 63 percent response time (process dead time plus time constant), use an enhanced PID with external reset feedback.
9. If the measurement lag becomes large enough to be considered a secondary lag in an integrating or runaway process, use the equations for the ultimate period to estimate the severe implications for tuning and the ultimate limit to performance.
10. If the measurement lag approaches or exceeds the loop dead time and cannot be reduced, set the PID rate time equal to the lag to compensate for the lag.
11. Consider whether a small threshold sensitivity limit would help increase the controller gain or enable an enhanced PID to stop limit cycles.
12. For wireless devices on slow processes, set the wireless update rate slow enough and the trigger level small enough where the actual process change in the update time is larger than the trigger level.
13. For wireless devices, make sure the wireless trigger level is lower than the measurement tolerance to meet environmental or pharmaceutical regulatory requirements and if permitted higher than sensor or process noise or the threshold sensitivity limit of the sensor.
14. For runaway processes, estimate the effect of the sensor lag on the ratio of the maximum to minimum controller gain.

The wireless default update rate must be less than half the oscillation period to prevent aliasing. Making sure this time interval between measurement updates is significantly less than the reset time and total loop dead time is generally a more stringent requirement on the default update rate.

6.2 WIRELESS UPDATE RATE AND TRANSMITTER DAMPING

The wireless update rate is the time interval between transmissions and consequently the update time. The transmitter damping setting is a signal filter. The wireless update rate has the same effect on as the PID execution rate and the transmitter damping has the same effect as a PID filter on loop performance. The primary effect on integrated error is well documented by the equations developed in Chapter 3. Not well recognized and adequately quantified is the secondary effect of update rate and damping on the behavior of the PID. An innovative term "implied dead time" developed in Chapter 5 is used to answer the general question: "when does a measurement add too much delay?"

A large transmitter damping setting or sensor lag can be insidious in deception. The response may look better in terms of the size of the deviation appearing smaller on a trend chart. The PID gain may even be able to be increased in some cases furthering the deception. Section 6.10 develops the equation to compute actual process variable amplitude from the measured amplitude to understand and predict the deception.

6.2.1 FIRST EFFECT VIA EQUATION FOR INTEGRATED ERROR

The update time considered here is the time between updates in any discontinuous signal. In this chapter the wireless default update rate is termed measurement update time (Δt_{wx}). The digital controller scan time can be considered to be negligible if there is oversampling of input signals to prevent aliasing.

The effect of a measurement time constant is many folds and can be quite insidious. For a small measurement time constant and update time that does not require detuning of the PID, the detrimental effect is straightforward and can be estimated by Equation 6.1 which is Equation 5.1 with measurement time constants ($\tau_{m1} + \tau_{m2}$) and wireless update time (Δt_{wx}) added in the numerator. However the amplitude attenuation described in Section 6.10 from the filtering effect of a measurement can hide process variability leading users and control systems to think the loop is better when in fact the process has deteriorated.

$$E_i = \frac{T_i + \Delta t_x + \tau_f + \tau_{m1} + \tau_{m2} + \Delta t_{wx}}{K_c * K_o} * E_o \tag{6.1}$$

6.2.2 SECOND EFFECT VIA EQUATIONS FOR IMPLIED DEAD TIME

Note that the Equation 6.1 shows the first effect of update time on integrated error but does not give an indication of the additional effect on tuning and stability. If the update time is appreciable compared to the implied loop dead time causing more oscillation, the increase in loop dead time from the update time and latency should be estimated. The implied dead time depends upon controller tuning besides the original dead time. The innovative term is based on the observation that a slowly tuned controller behaves similar to a loop with a larger than actual dead time.

We can reuse Equation 5.6 to estimate the amount of measurement delay that causes no appreciable need to retune the controller for self-regulating processes.

$$\theta_m < 0.5 * (\lambda + \theta_o) - \theta_o \tag{6.2}$$

Similarly, we can reuse Equation 5.14 to estimate the amount of measurement delay that causes no appreciable need to retune the controller for integrating processes.

$$\theta_m < 0.25 * (\lambda + \theta_o) - \theta_o \tag{6.3}$$

Equation 6.3 has the same form as Equation 6.2 with a coefficient of 0.25 instead of 0.5 implying a much greater sensitivity to measurement delay by integrating processes.

Now that we know the maximum allowable dead time from the controller, we need to know what this corresponds to in terms of measurement from a multitude of sources.

Equation 6.4a gives the measurement dead time (θ_m) from measurement lags 1 (τ_{m1}) and 2 (τ_{m2}), delays 1 (θ_{m1}) and 2 (θ_{m2}), wireless devices (θ_w), analyzers (θ_a), and threshold sensitivity limits. Note that the terms "update rate" is a time interval between updates rather than the number of updates per second.

$$\theta_m = \theta_w + \theta_a + Y * \tau_{m1} + Y * \tau_{m2} + \theta_{m1} + \theta_{m2} + \theta_{mx} \tag{6.4a}$$

The dead time from noncontinuous updates is half the wireless update time (Δt_{wx}) plus the wireless latency (Δt_{wz}) which is the time from the start of an update time interval to when a result is communicated as an output. Normally this latency is negligible.

$$\theta_w = 0.5 * \Delta t_{wx} + \Delta t_{wz} \tag{6.4b}$$

If a change in measurement is greater than the wireless trigger level, an update will occur before the next scheduled transmission set by the default update rate. Fortunately, the enhanced PID developed for wireless can handle extremely large and variable update times as described in Section 2.2.10 by executing a single feedback correction when a new result is detected as a change in the process variable.

If the total loop dead time from a wireless device or analyzer exceeds the implied dead time, the peak and integrated errors will be seriously degraded.

For self-regulating processes, the use of an enhanced PID may eliminate the need for the PID to be retuned as the wireless update time or analyzer cycle time and multiplex time increases. For loops where the implied dead time is greater than the primary process time constant, the PID tuning is no longer a function of the discontinuous update time. If the total loop dead time is much greater than the 63 percent process response time (process dead time plus process time constant), the enhanced PID gain can be set as large as the inverse of the open loop self-regulating process gain. While the enhanced PID prevents the loop from becoming oscillatory, the errors for a load disturbance are still increased. Equation 6.1 can be used for a rough estimate of the integrated error by adding the update time interval between wireless results to the numerator.

For near-integrating, true integrating, and runaway processes, the enhanced PID cannot prevent the degradation in loop performance and the need for retuning if the total loop dead time becomes larger than the implied dead time. The enhanced PID can still help in terms of dealing with a large variable update time and a communication failure or sensor failure (e.g., coated or broken pH electrode).

6.3 ANALYZERS

The latency for an analyzer is the cycle time since the analysis result is available at the end of the cycle time. Thus the dead time from an analyzer cycle time is 1½ the analysis cycle time ($1.5 * \Delta t_{ax}$).

Since analyzers are a large investment in capital and maintenance, multiple sample points are serviced by a single analyzer particularly for parallel trains of unit operations. The time the analyzer takes to get another sample from a given sample point is the multiplex time. The analyzer dead time is increased by half of the multiplex time ($0.5 * \Delta t_{az}$).

$$\theta_a = 1.5 * \Delta t_{ax} + 0.5 * \Delta t_{az} \tag{6.4c}$$

At-line analyzers can have quite a range of analysis cycle times from a few seconds for a mass spectrometer to 30 minutes for a chromatograph. For at-line analyzers there is also sample transportation delay that is the sample tubing volume divided by the sample flow rate. Since the flow rate needed by the analyzer is extremely small, transportation delays often simply known as sample delays can be quite large particularly if the analyzer is located in a central analyzer house to provide a better climate for analyzer operation and maintenance. Often a much higher flow recirculation line is used so that the sample line is as short as possible to reduce the sample delay.

Off-line analyzers have extremely large and variable analysis cycle times because the operator may not immediately take a sample to the lab, and technician in the lab may not do the analysis right away and quickly enter the result into the data historian. In some cases, the sample must be sent off to a special lab. The off-line analysis time can vary from hours to days.

The effect of analyzer cycle time and multiplex time can be mitigated by the use the enhanced PID developed for wireless. Results can be obtained that are similar to those discussed for wireless transmitters at the end of the last section.

Off-line analyzers can only be used if the open loop integrating process gain is so slow that the excursion is acceptable between worst case analysis intervals. Off-line analyzers are not generally used for closed loop control of runaway processes due to the risk of excessive delay, human error, and missing results.

where

K_c = controller gain (dimensionless)
K_o = open loop self-regulating process gain (dimensionless)
E_i = integrated error for load disturbance (e.u.)
E_o = open loop error for load disturbance (e.u.)
T_i = integral time (rate time) (sec)
Δt_{ax} = analyzer cycle time (sec)
Δt_{az} = analyzer multiplex time (sec)

Δt_{wx} = time interval between wireless updates (wireless default update rate) (sec)
Δt_{wz} = latency in wireless updates (calculation time)
λ = Lambda (closed loop self-regulating time constant or integrating arrest time) (sec)
θ_a = analyzer dead time (sec)
θ_i = implied loop dead time from PID tuning (sec)
θ_m = total measurement dead time (sec)
θ_{m1} = measurement delay 1 (sec)
θ_{m2} = measurement delay 2 (sec)
θ_{mx} = measurement delay from resolution or threshold sensitivity limit (sec)
θ_o = original loop dead time (sec)
θ_w = wireless dead time (sec)
τ_{m1} = measurement time constant 1 (sec)
τ_{m2} = measurement time constant 2 (sec)
Y = Fraction of a small time constant converted to dead time (Equation 4.1a)

6.4 SENSOR LAGS AND DELAYS

In this section and elsewhere in the book, *lag* is used as the shorter term for *time constant* and *delay* as the shorter term for *dead time*. Figures, text, and tables will use these terms interchangeably.

For pressure and differential pressure transmitters, the estimate of measurement dynamics needs to include impulse line or the diaphragm seal and capillary system sensing delays and lags. The dynamics of liquid filled impulse lines are negligible. For gas filled impulse lines, the dynamics can be estimated from tables for pneumatic signal lines. The delay and lag for 20 meters of half inch tubing is about 0.04 and 0.14 seconds. The delay and lag for 200 meters increase to 0.64 and 2.56 seconds, respectively. Most impulse lines are much shorter, making the dynamics negligible. Chemical seals can appreciably slow down a pressure response. The delay and lag for a capillary system increases as the diameter of the seal and the length of the capillary increase. For example, a three inch diaphragm seal has an 86 percent response time that is about 0.2 and 2 seconds for a 1.5 and 6 meter long capillary, respectively. The 86 percent response time is one delay and two time constants. The delay is usually about one-fourth the lag. Using this relationship, for the 1.5 meter long capillary, the delay was 0.02 and the lag was 0.09 seconds. For the 6 meter capillary, the delay and lags are about 10 times larger. Capillary systems can be longer than 6 meters. These numbers show the advantage of minimizing capillary length or eliminating the use, where possible, of chemical seals. For pressure measurements, the use of an integral mounted pressure transmitter is advisable (transmitter mounted directly on equipment or pipe pressure connection). For differential pressure measurements, integral smart mounted separate transmitters for the low and high pressure connections can be used if the total pressure is less than 10 times the differential pressure.

The fraction (Y) of the measurement lags converted to dead time can be computed from Equations 4.1a through 4.1c. The lags for most flow and pressure sensors are less than 0.1 second. The lags for temperature and pH sensors are much larger.

Table 6.1 shows the time constants for a new pH electrode. The word *new* means the glass surface is in factory condition and is completely clean. For this best case scenario, the pH electrode time constant varies between 0.25 and 6 seconds. A longer time constant occurs

Table 6.1. New pH electrode time constants

Direction	Magnitude (pH)	Buffering	Velocity (fps)	Time constant (seconds)
Positive	0.5	No	5	1.2
Negative	0.5	No	5	2.8
Positive	1.0	Yes	5	0.25
Negative	1.0	Yes	5	0.5
Positive	1.5	No	5	1.8
Negative	1.5	No	5	6.2
Positive	3.0	Yes	4	0.75
Negative	3.0	Yes	4	1.5
Positive	3.0	Yes	2	1.5
Negative	3.0	Yes	2	3
Positive	3.0	Yes	1	2
Negative	3.0	Yes	1	4
Positive	3.0	Yes	0.5	3
Negative	3.0	Yes	0.5	12

when there is no buffering (e.g., no weak acid or weak base and salts) that would moderate the dissociation of hydrogen ions and the slope of the titration curve. The table also shows the time constant for a decrease in pH is about twice the time constant for an increase in pH except at low velocities. As the velocity decreased from 5 to 0.5 feet per second (fps) the time constants increased by more than the order of magnitude.

The pH electrode time constants can become over 100 seconds for what appears to be an imperceptible 1 millimeter coating. Even more horrendous is the slowing down of the pH electrode due to aging of the glass surface. Figure 6.1 shows that the 86 percent response time of a clean electrode increased from about 30 seconds to 30 minutes due to premature aging of the glass surface from high temperature exposure, reducing the number of active sites. Since the 86 percent response time is one dead time plus two time constants and the dead time is negligible, the response time for the aged electrode simply corresponds to an electrode time constant of about 15 minutes. New high temperature glass electrode technologies prevent this aging so that a clean electrode stays as fast as a new electrode.

The use of three electrodes and middle signal selection will inherently prevent the measurement from using a slow electrode, whether due to a coating or aging of the glass. The installation enables the online identification of the increase in the electrode lag by noting the time difference between the ramp of the slow electrode and the middle value.

The measurement delay must include any transportation delays of the fluid concentration, temperature, or pH from a vessel or point of mixing of streams to the sensor. This delay

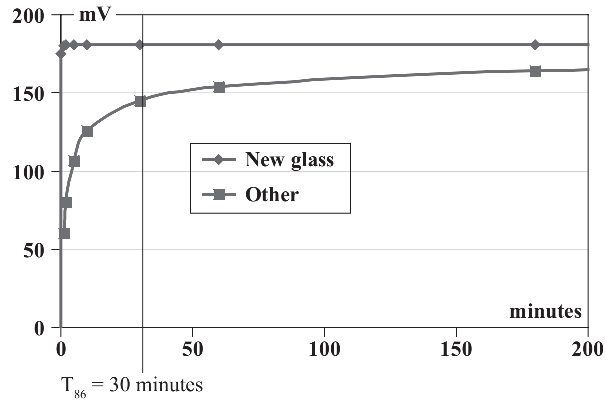


Figure 6.1. Increase in response time of clean pH electrode due to premature aging of glass surface from high temperatures.

is estimated as either the pipe distance divided by the pipe velocity or the pipe system volume divided by the pipe flow rate. The more general relationship is that the transportation delay for volumes without back mixing is the residence time (volume divided by flow). While normally associated with pipes, plug flow also exists in extruders, coils, dip tubes, and tubes in heat exchangers.

For example, a batch operation could not automatically cutoff reagent flow because the pH electrodes to measure vessel pH were mounted downstream of a heat exchanger in the recirculation line at a platform on the top of the vessel. The intent was reasonable in terms of increasing the life of electrodes by lowering the temperature and making them accessible via a platform. The additional delay from the residence time in the heat exchanger and piping caused the pH detection of end point to be late. The best location for the pH electrodes is about 20 pipe diameters downstream of the recirculation pump discharge. A similar criterion can be used for temperature sensors on the discharge of the tube side downstream of shell and tube heat exchangers.

Table 6.2 shows the bare element temperature sensor speed of response depends upon the element type and size. The time constant of a bare sheathed and *grounded* TC is about one-fourth the time constant for sheathed and *insulated* TC of the same size. The insulation in the tip helps prevent ground loops but slows down the response. The one-fourth inch resistance temperature detector (RTD) is slightly slower than the one-fourth inch sheathed and insulated TC. Bare sensors are rarely installed in chemical processes due to concerns about damage to the sensor and the hazard of removing the sensor from the process stream. Thermowells are used for sensor and personnel protection.

The response of thermowells is characterized by two interactive thermal lags. Table 6.3 shows the equivalent noninteractive time constants. Except for the last case of high liquid velocity (10 fps) and an exceptionally tight fit of the element in the thermowell (0.005 inch), the sum of the two time constants are much larger than the sensor time constant. Thus, the decision to use a TC instead of an RTD or an insulated TC instead of a grounded TC should not be based on the relative speed of the sensor.

An oil fill can make the thermowell response much faster but the oil may not be in the tip unless the thermowell is installed with the tip facing down. Also, high temperatures can cause

Table 6.2. Bare temperature element time constants

Bare element type	Time constant (seconds)
Thermocouples	
1/8 inch sheathed and grounded	0.3
1/4 inch sheathed and insulated	4.5
1/4 inch sheathed and grounded	1.7
1/4 inch sheathed and exposed loop	0.1
Resistance temperature detectors (RTD)	
1/16 inch	0.8
1/8 inch	1.2
1/4 inch	5.5
1/4 inch dual element	8.0

Table 6.3. Thermowell and thermocouple assembly time constants

Fluid type*	Fluid velocity (fps)	Annular clearance (inch)	Annular fill	Time constants (seconds)
Gas	5	0.04	Air	107 and 49
Gas	50	0.04	Air	93 and 14
Gas	152	0.04	Air	92 and 8
Gas	300	0.04	Air	92 and 5
Gas	152	0.04	Oil	22 and 7
Gas	152	0.02	Air	52 and 9
Gas	152	0.005	Air	17 and 8
Liquid	0.01	0.01	Air	62 and 17
Liquid	0.1	0.01	Air	32 and 10
Liquid	1	0.01	Air	26 and 4
Liquid	10	0.01	Air	25 and 2
Liquid	10	0.01	Oil	7 and 2
Liquid	10	0.055	Air	228 and 1
Liquid	10	0.005	Air	4 and 1

* The gas is saturated steam and the liquid is organic.

degradation of the oil creating tars, reducing the effectiveness of the fill and making removal of sensing element difficult.

The focus should be on thermowell construction and fluid velocity to make the temperature measurement fast enough. Not shown in the table is the effect of lower conductivity materials and a larger thermowell mass. The effect is minimal for metal thermowells (less than five seconds) because the conductance lag of metals is still much less than the lag of the air gap. However, the use of ceramic protection tubes for furnaces creates a large lag from particularly low material conductivity. The already slow response for a gas stream seen in Table 6.3 coupled with the large ceramic lag makes a measurement so slow as to cause problems. Optical pyrometers have a negligible lag and may be a better choice if an accuracy of 30°C is acceptable and the changes in emittance of the target and intervening gases are negligible or can be compensated for by a two color pyrometer.

During a test for a sensor time constant, the 86 percent response time (time to 86 percent of the final change) should be measured that is two time constants if the dead time is negligible. Waiting for the 98 percent response time that is four time constants is problematic and less useful. Noise and differences in threshold sensitivity can cause dramatically different results. Also, many sensors such as pH electrodes have a slow protracted response after 86 percent of the final value is reached causing an estimation of the time constant that is unreasonably long. Most of the PID correction occurs before the 86 percent response time. However, calibrations must use at least 99.9% of the final response.

6.5 NOISE AND REPEATABILITY

The technical definition of repeatability is the maximum deviation in readings for successive full scale changes in signal in the same direction. As a practical matter, most people think of repeatability as the difference in measurements for the same value of the process variable. Noise can be confused with repeatability because of a similar observed random deviation from the true value. Normally, noise has a higher frequency and is more effectively filtered by transmitter damping or signal filtering.

The source of noise can be the process or electromagnetic interference (EMI) and ground potentials. EMI noise is generally at a high frequency (e.g., 60 Hertz). Some of the principle causes are improper shielded and grounded signal cable, power cables, and older inverters for VSDs. The newer Pulse Width Modulated (PWM) drives and isolation transformers have reduced the noise problem. Ground potentials are particularly problematic for pH electrodes causing a dramatic dip in pH. The use of solution grounds and wireless transmitters has greatly reduced pH measurement noise.

For gas pressure and differential measurements, the suction or kinetic energy from high velocities at connections and probes can introduce sensor noise. Differential head flow meters on gas stream can be affected. These flow meters also suffer from the distortion of the velocity profile with rapid changes showing up as noise. Flow tubes and venture meters tend to have less noise than orifice meters.

Process noise in concentration, pH, and temperature measurements is primarily due to imperfect mixing and changes in phase. There are temperature and concentration gradients in even highly back mixed volumes, including the ideal case of an axial agitated liquid reactor with baffles and excellent geometry. The pH electrode is the best detector of concentration inconsistency because a fluctuation of 10^{-7} in hydrogen ion concentration can cause 1 pH measurement noise at 7 pH. For steep titration curves, this corresponds to 10^{-7} mixing uniformity requirement.

Volumes that do not have back mixing (e.g., axial mixing) are called plug flow. Gas furnaces and reactors, extruders, pipes, inline liquid reactors, and static mixers are nearly plug flow. The plug flow can be beneficial in terms of a constant residence time and hence reaction time. There is considerable noise since there is no consistency axially (length wise or time wise). Static mixers by design have some radial mixing but otherwise plug flow volumes have little consistency radially (cross sectional).

The tip of an electrode or thermowell should be near the pipe centerline because the temperature and composition vary over the cross section of the pipe. For highly viscous fluids, the error is pronounced. The temperature measurement in extruder outlets and the pH measurement in static mixer outlets with a sulfuric acid reagent are particularly sensitive to the depth of insertion of the sensor tip due to the effects of the high viscosity of polymers and of 98 percent sulfuric acid.

The thermowell immersion length needs to be long enough to minimize the change in tip temperature from heat transfer from the process to the pipe wall (e.g., immersion length >6 pipe diameters). For example, a temperature sensor with a two thermowell diameter immersion length was found to be measuring more the Therminol™ temperature in the piping jacket than the polymer temperature. Temperature sensors should be installed in elbows and if this not possible, the sensor should be angled in small diameter pipelines.

Bubbles in liquid streams and droplets in gas streams cause noise when they hit a sensor. Bubbles from air, oxygen, and carbon dioxide spargers in bioreactors and chemical reactors can cause dissolved oxygen and pH signals to become noisy. Droplets of water at a desuperheater outlet cause a noisy temperature measurement. Ammonia bubbles at a static mixer outlet cause a noisy pH measurement.

Changes in phase can occur in impulse lines. The calibration of a differential pressure is based on fixed conditions of fluid phase and density. Bubbles can enter the liquid impulse lines or be created in the lines by vaporization. Similarly droplets can enter gas impulse lines or be created in the lines by condensation. Heat tracing turning on and off will create fluctuations in fluid density in impulse lines.

The use of diaphragm seals with capillary systems in place of impulse lines creates different sources of noise. Often not recognized is the noise from unsecured capillary systems moving in the wind. Changes in ambient conditions also introduce extraneous changes in capillary fill density and hence differential pressure transmitter output. Day to night temperature changes are slow but sun, clouds, rain, and sleet can cause temperature changes and shifts in signals that are fast compared to the response of large liquid volumes. Smart transmitters have improved greatly in compensating for changes in ambient conditions and fluid conditions at the transmitter diaphragm but dealing with changes in the capillary system is more problematic because there are no temperature sensors in the capillary or at the diaphragm seal.

The signal to noise ratio and accuracy of differential pressure measurements can be greatly improved by the use of integral direct mounted pressure transmitters. By not using impulse lines and capillary systems, extraneous changes are eliminated, maintenance is reduced, and smart transmitters can directly compensate for changes in ambient and process conditions enabling an installed accuracy close to the manufacturer's specifications. For differential pressure measurements, the accuracy of separate pressure measurements must be assessed in the digital computation of the differential pressure.

6.6 THRESHOLD SENSITIVITY AND RESOLUTION LIMITS

A resolution limit is the smallest input change that will cause an output change. When the input change exceeds this limit, the output change is an integer multiple of the resolution limit. Thus, the output response is a series of steps with a size that is an integer multiple of the resolution limit. In contrast when a change in input exceeds the threshold sensitivity limit, the change in output equals the change in input. The result is a step size that matches the input change. Resolution limits come from digital devices. However, the ISA standard for valve response testing expresses stick-slip as a resolution limit when really a threshold sensitivity limit is more at play where the valve does not move with a fixed step size.

The 1980s vintage DCS had only 12 bits in the input cards. Since one bit is for sign, there were only 11 bits for resolution. If you divide 100 percent by two to the 11th power, you get a resolution limit of 0.05 percent which doesn't seem bad until you consider the practice at the time was to use wide range TC input cards rather than transmitters to save on field hardware costs. These cards caused a series of steps of about 0.25°C that caused such severe bumps in the PID output from rate action that derivative could not be used on reactors where derivative is most needed to compensate for secondary time constants from thermal lags. Today's DCS has 16 bit input and output cards.

When a step change in the PV occurs, the derivative mode thinks the entire change occurred in one PID execution. Exothermic liquid reactors have a slow temperature response and use a PID gain greater than 10 and a rate time greater than 120 seconds to prevent a runaway. Since the step from a resolution limit is divided by the PID execution time and multiplied by the product of the PID gain and rate time, a 0.25°C step can cause a full scale bump in the PID output.

Resolution and threshold sensitivity limits both cause limit cycles if there is any integral action in the process or the controller. The amplitude depends upon the process gain. For pH control on the steep part of a titration curve, the process gain is so large, the limit cycle can violate pH limits. Such pH processes are very sensitive to valve resolution and threshold sensitivity limits.

Most sensors and analyzers have a threshold sensitivity limit larger than the resolution limit in today's smart transmitters. The wireless default trigger level is really a threshold sensitivity limit. The default update rate helps prevent a limit cycle if the trigger level is set too large.

The best solution is to use transmitters and valves with the best resolution and threshold sensitivity. Limit cycles and bumps in the PID output from rate action can be eliminated by the use the enhanced PID with external reset feedback.

The dead time from a measurement sensor resolution or threshold sensitivity limit can be approximated as a measurement sensitivity limit (S_m) divided by rate of change of the process variable ($\Delta\%PV / \Delta t$). For fast large disturbances, this dead time is negligible.

Equation 6.4d estimates the dead time from a measurement sensitivity limit.

$$\theta_{mx} = \frac{S_m}{\Delta\%PV / \Delta t} \tag{6.4d}$$

A small threshold sensitivity limit can have a beneficial effect if it is set just large enough to prevent PID reaction to noise. If the PID does not see noise, the controller gain and rate time settings can be increased to the full extent permitted by loop dynamics. For most near-integrating, true integrating, and runaway processes, the main limit to how high you can set the gain and rate time is amplification of noise. For the enhanced PID developed for wireless, the elimination of extraneous updates due to noise can prevent limit cycles from valve stick-slip and oscillations from large analyzer cycle times.

The step in the PID output from a resolution limit ($\Delta\%CO_x$) can be estimated by Equation 6.5a which shows the effect of the input card span ($\Delta\%PV_m$) and the controller scale ($\Delta\%PV_c$).

$$\Delta\%CO_x = \frac{\Delta\%PV_m}{2^{n-1}} * \frac{100\%}{\Delta\%PV_c} * K_c * \left(\frac{T_d}{T_f} + 1.0\right) \tag{6.5a}$$

For most DCS with built-in filter that is one-eighth of rate time and no additional filter:

$$\Delta\%CO_x = \frac{\Delta\%PV_m}{2^{n-1}} * \frac{100\%}{\Delta\%PV_c} * K_c * 9 \tag{6.5b}$$

6.7 RANGEABILITY (TURNDOWN)

Errors are specified as a percent of span or as a percent of reading of the measurand (measured variable) and the transmitter output. The difference in performance for the two specifications can be dramatic particularly in terms of rangeability (turndown). For example, a 0.1 percent of span accuracy transmitter for a linear process relationship would have a maximum error of 0.5 percent at 20 percent output (five times larger than a 0.1 percent of reading accuracy at 100 percent output).

For the differential head flow meter, the deterioration in accuracy with turndown is much more drastic because the flow is proportional to the square root of the differential pressure measurement. A 0.1 percent of span accuracy transmitter would have a maximum error of 2.5 percent at 20 percent output (25 times larger than the error at 100 percent output). To increase rangeability, a second transmitter with a reduced span (e.g., 20 percent) is added and the computed signal automatically switched to use the second transmitter when the flow drops to 20 percent. However, the effect of noise is still a factor since a given pressure noise amplitude causes a flow noise that is 25 times larger at 20 percent than 100 percent flow. To really achieve much better than 4:1 turndown with differential head meters, extraordinary measures need to be taken in terms of reducing pressure noise from the process, impulse lines, and capillary lines.

Magnetic flow meters (magmeters), turbine meters, and vortex meters have a low velocity limit that is a percent of the manufacturer's velocity specified for a given meter size. The application maximum velocity is less than the meter maximum velocity for the size selected. Consequently, the actual rangeability achieved is less than what is stated in the literature. Fortunately these meters have an accuracy that is a percent of flow.

For example if the process maximum velocity is 60 percent of the size maximum, the 30:1 rangeability stated for a vortex meter becomes 18:1. The meter coefficient and rangeability is also affected by kinematic viscosity and upstream and downstream piping. The straight run requirements are similar to those of a differential head meter.

The 100:1 rangeability of a magmeter is reduced to 60:1 for the same situation as the vortex meter. If the fluid conductivity approaches a minimum, the signal to noise ratio and rangeability deteriorates. However, the meter is not affected by viscosity and only requires a straight run of five pipe diameters upstream and downstream of the meter.

The Coriolis meter has the greatest rangeability (200:1) and is not affected by process conditions. This is the only true mass flow meter. All of the other flow meters are volumetric and can only provide a mass flow measurement for a known composition and hence density relationship to temperature. The Coriolis meter can supply an extremely accurate actual density measurement for any composition including a mixture of phases (newer special meters). The Coriolis meter offers an inferential measurement of the concentration of components or phases (e.g., bubbles). There are essentially no upstream and downstream piping requirements. The straight tube version developed for slurries has significantly less accuracy and rangeability.

The use of flow as a lower loop in a cascade control system has many benefits in terms of isolating valve nonlinearities and pressure disturbances from the upper process loop and enabling flow feedforward control. However, precautions must be taken when the flow drops below the rangeability of the flow meter where the flow measurement gets erratic or drops out (vortex meter). The best solution is to do a bumpless switch of the controlled variable in the lower loop from the flow meter signal to a computed flow based on valve position and the installed flow characteristic. The computed flow does not need to be accurate to sustain feedback and feedforward control since the upper PID will correct the lower PID setpoint for errors in the lower loop PV.

6.8 RUNAWAY PROCESSES

Runaway processes are extremely sensitive to the measurement accuracy and dynamics of the key process variable, such as exothermic reactor temperature. A higher actual than measured temperature can cause an acceleration in a positive temperature excursion and trigger undesirable side reactions. A large thermowell time constant can cause uncontrollable consequences. If the temperature measurement time constant approaches the positive feedback time constant, the denominator for the ultimate period detailed in Equation 4.21c approaches zero. A huge increase in the ultimate period causes the ultimate gain defined by Equation 4.22b to cross the minimum PID gain defined by Equation 4.22f closing the window of allowable controller gains.

A more extensive study of exothermic reactors quantified of the effect of not only the thermal lag in a temperature measurement but also the thermal lag of a heat transfer surface on the ratio of maximum to minimum PID gains. The result is expressed in the following series of equations documented in the Hydrocarbon Processing article "Consider Reactor Control Lags" (Luyben and Melcic 1978).

The minimum PID gain is as noted in Equation 4.22f the inverse of the reactor's open loop runaway process gain for a cascade control system of reactor temperature to jacket temperature. This open loop gain is the product of the jacket temperature gain, reactor process gain (positive feedback process gain as per Appendix F), and reactor temperature measurement gain. For cascade control the jacket temperature gain is simply the jacket temperature controller setpoint scale span divided by the span of the PID algorithm output (e.g., 100 percent).

$$K_{c\min} = \frac{1}{K_j * K_p^{'} * K_m} \tag{6.6a}$$

The maximum controller gain depends upon the ratio of the heat removal time constant to the positive feedback process time constant (R_h) and the ratio of the reactor temperature measurement time constant to the positive feedback time constant (R_m). If the sum of these ratios approaches 1.0, the maximum PID gain is the minimum PID gain.

$$K_{c\max} = \frac{1 + (1 - R_m - R_h) * \left[\frac{R_m + R_h}{R_m * R_h} - 1 \right]}{K_j * K_p^{'} * K_m} \tag{6.6b}$$

For industrial processes, a ratio of maximum to minimum controller gain greater than twenty ($R_c > 20$) is advisable to kept the window of allowable controller gains large enough considering tuning errors and changes in dynamics with heat transfer surface coefficient and feed rates.

$$R_c = 1 + (1 - R_m - R_h) * \left[\frac{R_m + R_h}{R_m * R_h} - 1 \right] \tag{6.6c}$$

where

K_c = controller gain (dimensionless)
$K_{c\max}$ = maximum controller gain (dimensionless)
$K_{c\min}$ = minimum controller gain (dimensionless)
K_j = jacket gain (jacket temperature scale span/100%) (°C/%)
$K_p^{'}$ = reactor positive feedback process gain as per Appendix F (°C reactor/°C jacket)
K_m = measurement gain (100%/reactor temperature scale span) (%/°C)
R_c = ratio of maximum to minimum controller gain (dimensionless)
R_h = ratio of heat removal to reactor positive feedback time constant (dimensionless)
R_m = ratio of measurement to reactor positive feedback time constant (dimensionless)

6.9 ACCURACY, PRECISION, AND DRIFT

The *accuracy* of a measurement is the maximum deviation of the mean measured value from the true value over a stated range of the measurement (e.g., 1,000 to 4,000 lb/hr flow), measurand (e.g., 5 to 100 inches orifice differential pressure for differential head flow meters), or process condition (e.g., 0.1 to 10 fps fluid velocity for magmeters and vortex meters). The maximum error is the worst case of the systematic and random errors in the measurement. Systematic errors are constant errors for a given true value. The most common systematic error is a bias of the mean of measured values and the true process value as shown in Figures 6.2a and 6.2b. The bias error can be due to an offset (intercept) error or a span (slope) error.

Measurement span is the difference between the lower and upper values of a measurement range. For example, a temperature transmitter calibrated for a temperature range of 50°C to 150°C has a span of 100°C. While technically a span error should be corrected by a span adjustment, a zero or offset adjustment is sufficient unless the process setpoint significantly changes. This strategy is the basis of pH standardization where a pH measurement is made to agree with a lab sample by an offset adjustment. A span error will change the open loop gain but the effect is generally much less than that of nonlinearities in the loop.

For an offset error, the bias is constant. For a span error, the bias increases with measurement. A bias can also originate from a fixed nonlinearity error that is a function of the true value. An offset

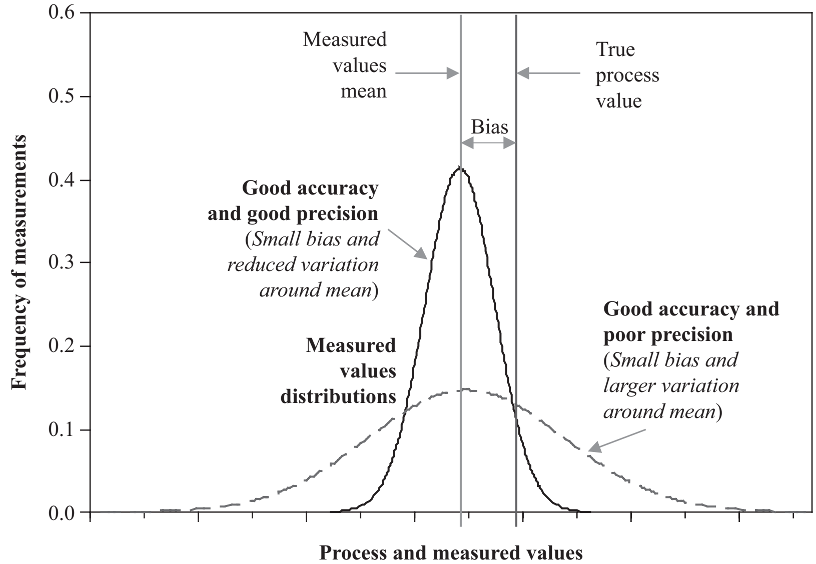


Figure 6.2a. *Good* accuracy measurement with good and poor precision.

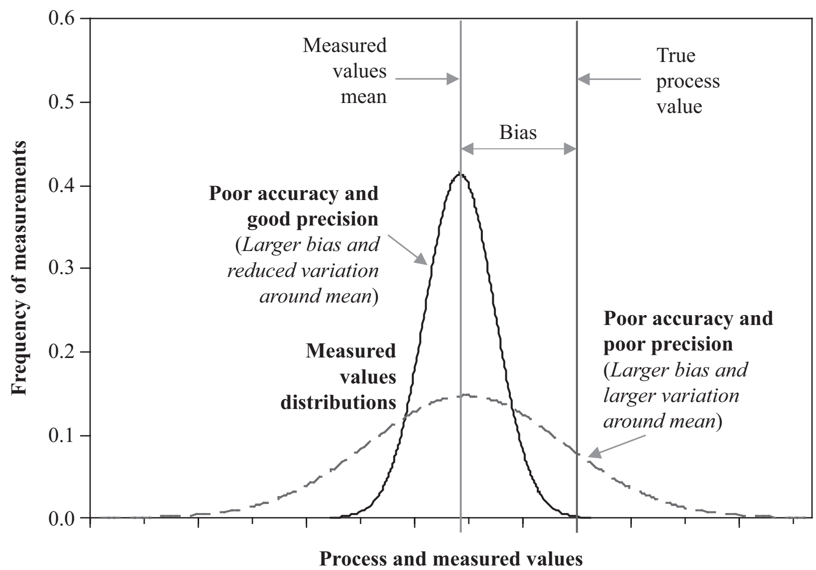


Figure 6.2b. *Poor* accuracy measurement with good and poor precision.

error can be compensated for by a corresponding bias in the setpoint of the loop with the measurement. While the bias can achieved automatically by an upper loop in a cascade control system or manually by process engineers or operators, there is still a problem in terms of steady state process analysis. Material and energy balances will not close. The amount of energy and material going into a phase in a volume will not equal the energy and material going out. The error in product, raw material, and utility flows will cause an error in the computation of process efficiency.

Accuracy is often expressed in terms of a tolerance, which is the allowable deviation in the error from a true value. While a sensing element tolerance can result from randomness in the manufacturing process the result is often a bias error in the measurement for a given sensor. As a result the device's tolerance consists of mostly a systematic error in the measurement. Systematic errors can be largely eliminated by linearization, calibration, and by set point changes. Random errors show up as variability in the process and cannot be eliminated by calibration or manual changes in the set point of the measurement loop. However, the effect of random errors on important process variables downstream can be attenuated by a back mixed volume. An upper loop on a downstream volume can control the average of the fluctuations so the key variables are closer to the desired value. For example, the effect of random errors in a reagent flow measurement on vessel pH can be minimized by the cascade of the output of a primary pH loop that automatically adjusts the secondary reagent flow set point. In the section on repeatability, the ability of back mixed volumes to attenuate random errors is discussed.

The *precision* of a measurement is the standard deviation of the total random error. The principal sources of random errors are repeatability and noise that have a statistical distribution as shown in Figure A10 in Appendix A. Sensitivity and resolution limits and hysteresis and deadband create errors that appear to be random due to changes in the start and stop points and direction of the change in the true value of measurement. For modern measurements with digital transmitters and no mechanical mechanism, the random error from resolution, sensitivity, hysteresis, and deadband are generally negligible.

Drift affects the accuracy but not the precision of measurement. Transmitter drift has been slowed by smart digital devices by orders of magnitude to the point where the electronic drift is negligible. What is left is sensor drift. Even here new sensor designs have dramatically reduced the drift. For example, coplanar pressure and differential pressure transmitters have an accuracy of 0.025 percent of span and 10 year stability. Furthermore, the installed accuracy remains essentially the specification accuracy, because of the compensation of ambient and process effects. In the 1970s, a typical electronic pressure transmitter had a bench top accuracy of 0.25 percent, an installed accuracy of one to two percent, and a stability specification of one to two years.

Thermocouples and pH electrodes are the sensors with significant drift. Table 6.4 indicates the drift from thermocouples can range from 1 to 20 degrees a year. The potential drift (figuratively and virtually) for pH electrodes requires greater consideration because the drift can dramatically increase from 0.2 pH per year for pH extremes (e.g., pH < 2 or pH > 10, high temperatures (e.g., $>50°C$), high salt concentrations (e.g., >1 percent), low water concentrations (e.g., <60 percent), poisoning (e.g., cyanides and sulfides), plugging of reference, and low fluid conductivity (e.g., condensate and boiler feedwater). For worst case conditions, the drift can accelerate with time "in-service" to be as large as 0.2 pH per day. Technicians may end up chasing calibration adjustments. The result is poor performance and excessive maintenance. Three pH electrodes with middle signal selection can help enormously particularly if the electrodes have different vintages. Middle signal selection will inherently prevent the measurement from using the electrode with the greatest drift. Middle signal selection also greatly reduces maintenance since calibration or replacement can be delayed until one electrode fails or the offset of two electrodes from the middle is excessive.

The manual correction of drift by process engineers and operators adjusting setpoints is difficult at best because the bias is a moving target. The direction may be predictable but the rate of drift is generally not. The drift has to be very slow for manual correction.

Table 6.4. Accuracy, range, and size of temperature sensing elements

Criteria	Thermocouple	Platinum RTD	Thermistor
Repeatability (°C)	1–8	0.02–0.5	0.1–1
Drift (°C/year)	1–20	0.01–0.1	0.01–0.1
Sensitivity (°C)	0.05	0.001	0.0001
Temperature Range (°C)	–200–2,000	–200–850	–100–300
Signal Output (volts)	0–0.06	1–6	1–3
Power (watts at 100 ohm)	1.6×10^{-7}	4×10^{-2}	8×10^{-1}
Minimum Diameter (mm)	0.4	2	0.4

6.10 ATTENUATION AND DECEPTION

Time constants act as filters attenuating oscillations. Equation 6.7a (Equation 4.12a for a Bode Plot amplitude ratio) shows that the attenuation of a measurement oscillation decreases as the measurement time constant (τ_f) and frequency of the oscillation (w_o) decreases. For a measurement time constant greater than the period, Equation 6.7a simplifies to Equation 6.7c via the use of Equation 6.7b to get from the frequency domain to the time domain for better visualization on trend charts. Equation 6.7c shows the measurement oscillation amplitude is the process amplitude attenuated by a factor that is proportional to the oscillation period and inversely proportional to the time constant.

$$A_m = A_p * \frac{1}{\sqrt{1+(\tau_m * w_o)^2}} \tag{6.7a}$$

Converting the oscillation from the frequency domain to the time domain:

$$\omega_o = \frac{2 * \pi}{t_o} \tag{6.7b}$$

For a measurement time constant greater than the oscillation period ($\tau_m > t_o$):

$$A_m = A_p * \frac{t_o}{2 * \pi * \tau_m} \tag{6.7c}$$

where

A_p = process oscillation amplitude (e.u)
A_m = measurement oscillation amplitude (e.u)
t_o = period of oscillation (sec)
τ_m = measurement time constant (sec)
w_o = frequency of oscillation (radians/sec)

A time constant can be beneficial or detrimental. A single large time constant in the process can reduce the variability of process inputs to the point of being negligible in the process output. This effect is rarely included in the analysis of potential benefits in reducing variability in upstream loops. For a back mixed liquid volume, the process time constant is essentially the residence time (volume/flow). This time constant is so large that limit cycles and damped oscillations and peak errors from control loops disappear. Furthermore, the process time constant can greatly slow down step disturbances giving the PID ample time to correct for them before there is an appreciable error. Plus, the maximum allowable PID gain is proportional to this time constant for near-integrating process tuning rules. The downsides of a large process time constant occur for disturbances downstream of the time constant, large setpoint changes, and tuning tests that are looking for a steady state. The process time constant slows down the correction from the manipulated flow and the time to steady state. A feedforward signal with a lead time equal to the process time constant can help. The time to reach setpoint and to complete a tuning test are mitigated by the use of a high PID gain, proportional action on error, and an integrating process tuning test that needs to only see the initial ramp rate.

These same benefits of a large time constant in terms of slowing down disturbances and attenuating oscillations from the process inputs can mislead practitioners into using time constants in the measurement resulting is a delayed and attenuated view of the real world. If the measurement time constant becomes the largest time constant within the loop, the deception is incredibly destructive. The amplitude of an oscillation and the peak error from a disturbance decrease and the PID gain can be increased (if the reset time is less than four dead times) as the measurement time constant is increased. Note that the equations for peak and integrated error and tuning do not know where the time constant is in the loop. As the largest time constant increases, the PID gain can be increased and the observed errors decreased. Equation 6.7c needs to be used to convert the filtered peak error as measured to the actual process variable errors. Four actual examples show the spectrum and severity of the deception.

An automation engineer presenting his paper at an ISA Conference said he almost was not allowed to do the paper because his company thought the technical advantage he discovered was so important it should be kept secret as a proprietary knowledge. The presenter had increased the measurement time constant to be by far the largest time constant in the loop. Since he was only trending the filtered measurement, the amplitude of the observed variability was drastically reduced.

A biochemist partially withdrew a temperature sensor in a bioreactor thermowell. The considerable resulting air gap in the tip made the temperature trend smoother. He was so proud he showcased his discovery and decided to run all of his batches this way.

A process engineer noted that a recently installed temperature sensor in a massive block of metal on the extruder outlet showed a dramatic reduction in the temperature variability. The other extruder temperature installations were accordingly modified to move the temperature sensors from the polymer melt to a block of metal. The trend charts looked better but operations eventually realized something was wrong when customers complained about product quality.

The trend charts of temperatures during the startup of a new plant were incredibly smooth but plant performance was horrible. Upon removal of one temperature sensor, sand was found in the tip of the thermowell. The E&I construction crew had installed open thermowells before the piping was sand blasted.

Even seemingly small measurement time constants can be problematic. Figure 6.3 shows that a time constant of 1.7 seconds will hide compressor surge oscillations that typically have a period of one to two seconds. Several surges will have occurred before the suction flow has dropped enough to indicate a problem. Each compressor surge cycle reduces compressor

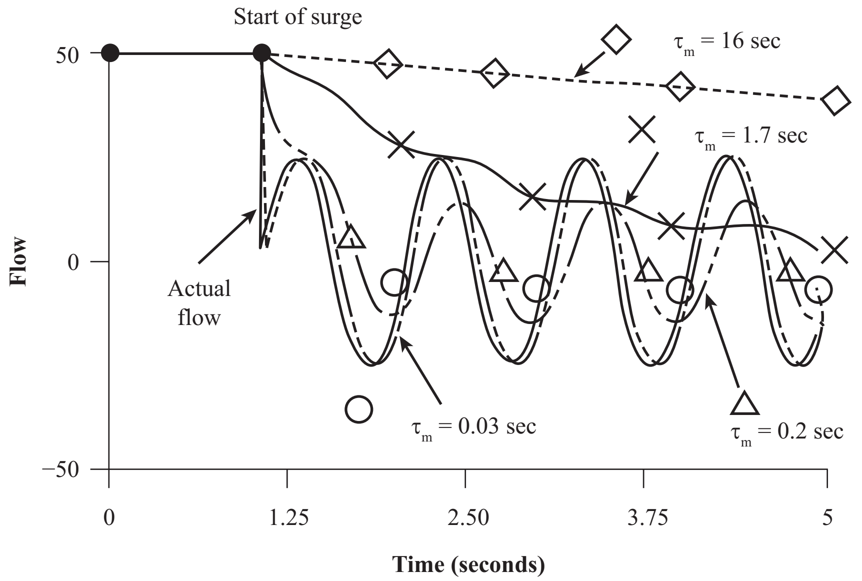


Figure 6.3. Effect of transmitter damping setting on PID PV for compressor surge oscillation.

efficiency from high axial thrust, radial vibration, and temperature damaging seals and the rotor. The compressor may trip before the anti-surge control system opens the vent or recycle valve sufficiently.

Measurement time constants including damping adjustments in pressure and differential pressure transmitters should be minimized (e.g., <0.2 sec) for compressor, furnace, and liquid polymer pressure control. All of the transmitters in the warehouse should have their damping settings minimized. One automation engineer lamented that the transmitters from one supplier had a default damping setting of 1.2 seconds. He found he had to reduce the setting to 0.2 seconds to prevent a compressor trip but whenever a transmitter was replaced on the weekends or night shift, a compressor would shut down.

People are not the only source of an excessive measurement time constant. A new clean glass pH electrode on a static mixer or inline pH control system has a measurement time constant larger than the process time constant. As the electrode ages or gets coated, the pH trends may look smoother.

Temperature sensors in fluidized gas reactors, reformers, and furnaces have a measurement time constant from the thermowell that is larger than the process time constant for the gas volume, which has a small residence time and little back mixing. As thermowell design and installation gets worse, the peak temperatures from hot spots and disturbances appear smaller.

The key to knowing whether a measurement time constant is too large is spotting the increase in dead time and settling time. A slow reset cycle (oscillation that is about 10 times the dead time) may develop if the reset time was set based on a fast sensor.

The installation of faster measurement can be discredited because the amplitude of the variability on a trend chart will be larger after the improvement in measurement design or installation. Such was the case for the installation of a much faster pressure transmitter on a phosphorous furnace. Even though the number of electrode seal blows and furnace trips had decreased, operations were concerned about the appearance on the trend charts. Fortunately, the old transmitter was kept for indication only that showed the actual process variability had decreased. The lesson here is to keep the old measurement or make the calculation of the filtered measurement based on the old measurement time constant using Equation 6.7c to show a "before" and "after" trend.

6.11 EXAMPLES

More information on the source of the solutions can be obtained by going to the portion of the chapters denoted by the equation numbers shown in the solutions.

6.11.1 WASTE TREATMENT pH LOOP (SELF-REGULATING PROCESS)

Given:

(Information from Example 4.6.1)

a. Set point is pH 7.
b. Measurement range is pH 0 to 14.
c. Minimum influent flow is 10 gallons per minute (gpm) (F_{il}).
d. Normal influent flow is 22 gpm (F_{in}).
e. Maximum influent flow is 100 gpm (F_{ih}).
f. Influent concentration is 32 percent by weight HCl (10.17 normality) (C_i).
g. Influent disturbance is rapid increase of 2.2 normality HCl concentration (ΔC).
h. Reagent concentration is 20 percent by weight NaOH (7.93 normality) (C_r).
i. Vertical tank liquid volume is 1,000 gallons (V).
j. Axial blade agitator diameter is 1 foot (D).
k. Axial blade agitator speed is 120 rpm (N_s).
l. Axial blade agitator discharge coefficient is 1 (N_q).

(New information)

m. Volume of pipeline from tank to pH electrodes is 30 gallons (V_e).
n. Recirculation turbulent flow in pipeline to electrodes is 15 gpm (F_e).

Find: ultimate limit for peak and integrated errors for each of the individual tanks for an unmeasured disturbance entering the first tank.

Solution:

a. Calculate the new dead time and ultimate period:
From the solution to Example 4.6.1:

$$\tau_p = 19 \text{ min}$$
$$\theta_p = 1 \text{ min}$$
$$E_o = 2.2 \text{ normality}$$

The dead time from the pH measurement is the transportation delay to the electrode per equation for injection plug flow in Table 4.1 plus the fraction of the electrode time constant converted to dead time. The transportation delay is the volume of pipeline to the pH electrode divided by the recirculation flow rate. From Table 6.1 the time constant of a clean electrode for a large pH decrease in a system with no buffering is about six seconds.

$$\theta_m = V_s / F_s + Y * \tau_e = 30/15 + 1.0 * 0.1 = 2.1 \min$$

The new total loop dead time is the measurement dead time added to the process dead time

$$\theta_o = \theta_p + \theta_m = 1 + 2.1 = 3.1 \min$$

The new ultimate period for the total loop dead time is:

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o}\right)^{0.65}\right] * \theta_o = 2 * \left[1 + \left(\frac{19}{19 + 3.1}\right)^{0.65}\right] * 3.1 = 11.8 \min \tag{4.5a}$$

b. Calculate the ultimate limit of concentration peak error for individual loops on each vessel based on the new dead time and ultimate period:

Calculate peak error of 1st vessel from E_o step disturbance:

$$E_{x1} = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o = \frac{1.8 * 11.8}{2 * \pi * 19} * 2.2 = 0.39 \textit{ normality} \tag{4.14b}$$

Calculate the peak error of 2nd vessel from peak error of 1st vessel taking into account the disturbance is slowed down by the first tank volume and the open loop error for the loop on the 2nd vessel can be approximated as the peak error from the first vessel:

$$E_{x2} = (1 - e^{-\theta_o/\tau_o}) * \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_{x1} = 0.15 * \frac{1.8 * 11.8}{2 * \pi * 19} * 0.39 = 0.0104 \textit{ normality}$$

Calculate peak error of 3rd vessel from peak error of 2nd vessel:

$$E_{x3} = (1 - e^{-\theta_o/\tau_o}) * \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_{x2} = 0.15 * \frac{1.8 * 11.8}{2 * \pi * 19} * 0.0104 = 0.000222 \textit{ normality}$$

c. Calculate the ultimate limit of pH peak errors for individual loops on each vessel using pH definition:

$$E_{xpH1} = 7 - \log(E_{x1}) = 7 - \log(0.39) = 6.6\, pH$$

$$E_{xpH2} = 7 - \log(E_{x2}) = 7 - \log(0.0104) = 5.0\, pH$$

$$E_{xpH3} = 7 - \log(E_{x3}) = 7 - \log(0.000222) = 3.5\, pH$$

d. Calculate the ultimate limit of concentration integrated errors for individual loops on each vessel:

The integrated error in terms of the peak error is approximately the ratio of Equation 4.13a to 4.14a:

$$R = \frac{E_i}{E_x} = \frac{\dfrac{T_u}{K_u * K_o}}{\dfrac{1.8}{K_u * K_o}} = \frac{11.8}{1.8} = 6.56$$

$$E_{i1} = R * E_{x1} = 6.56 * 0.39 = 2.56 \; normality * \min$$

$$E_{i2} = R * E_{x2} = 6.56 * 0.0104 = 0.068 \; normality * \min$$

$$E_{i3} = R * E_{x3} = 6.56 * 0.000222 = 0.0015 \; normality * \min$$

e. Calculate the ultimate limit of pH integrated errors for individual loops on each vessel:

$$E_{ipH1} = R * E_{xpH1} = 6.56 * 6.6 = 43.3 \; pH * \min$$

$$E_{ipH2} = R * E_{xpH2} = 6.56 * 5.0 = 32.8 \; pH * \min$$

$$E_{ipH3} = R * E_{xpH3} = 6.56 * 3.5 = 23.0 \; pH * \min$$

Conclusions: The pipeline transportation delay increases the dead time in the loop by a factor of 3. The peak error in concentration units increased by a factor of 3, 8, and 23 for vessels 1, 2, and 3, respectively. The greater increase in peak error for tanks 2 and 3 is due to the increase in peak error from the upstream tank and the reduced filtering effect due to the increase in ultimate period. The accumulated error in concentration units increased by a factor of 8, 24, and 68 for vessels 1, 2, and 3, respectively. The factors are greater for the accumulated error because the increase in ultimate period increases the recovery time. The increase in pH errors will not be as great as a result of the exponential relationship between normality and pH. The electrode time constant was entirely converted to effective dead time, since the electrode time constant was less than five percent of the residence time. The additional error due to transportation delay is a common problem because the electrodes are frequently located in a recirculation pipeline instead of inside the tank for easier access for maintenance and higher velocities. Unfortunately the location chosen is near the reagent valves at a platform on the tank top that is the farthest point from the tank discharge nozzle (the pipeline volume, V_e, is large). The transportation delay can be reduced either by locating the electrodes closer to the pump or by increasing the recirculation flow.

6.11.2 BOILER FEEDWATER FLOW LOOP (SELF-REGULATING PROCESS)

Given:

(Information from Example 4.6.2)

a. Set point is 100,000 pounds per hour (pph).
b. Measurement range is 0 to 200,000 (pph).
c. Disturbance is rapid 20 percent increase in flow (ΔF).
d. Pipe diameter is 4 inches or 0.33 foot (D_p).
e. Fluid velocity is 5 feet per second (fps) (V_f).
f. Pipe friction factor is 0.01 (C_f).
g. Pipe wall modulus of elasticity is 500,000,000 lb/ft2 (E_p).
h. Pipe wall thickness is 0.34 inch or 0.03 foot (H).
i. Fluid density is 62 lb/ft3 (ρ).
j. Fluid modulus of elasticity is 5,000,000 lb/ft2 (E_f).
k. Acceleration due to gravity is 32 ft/(sec*sec) (G).
l. Pipe length from valve to transmitter or discharge is 150 feet (S_p).
m. Ratio of valve pressure drop to pipe pressure drop is 0.7 (R).

(New information)

n. Transmitter damping (time constant) is set to 1 sec (τ_m)

Find: ultimate limit for peak and integrated errors for an unmeasured step disturbance.

Solution:

a. Calculate the new dead time and ultimate period:
From the solution to Example 4.6.1:

$$\tau_p = 1\text{ sec}$$
$$\theta_p = 0.1\text{ sec}$$
$$E_o = 11\ \text{ lb/sec}$$

The dead time from the flow measurement is the fraction of the transmitter time constant converted to dead time. From Figure 4.2 we see the fraction is about 0.28 for equal time constants.

$$\theta_m = Y * \tau_m = 0.28 * 1 = 0.28\text{ sec}$$

The new total loop dead time is the measurement dead time added to the process dead time

$$\theta_o = \theta_p + \theta_m = 0.1 + 0.28 = 0.38\text{ sec}$$

The new ultimate period for the total loop dead time is:

$$T_u = 2 * \left[1 + \left(\frac{\tau_o}{\tau_o + \theta_o}\right)^{0.65}\right] * \theta_o = 2 * \left[1 + \left(\frac{1}{1 + 0.38}\right)^{0.65}\right] * 0.38 = 1.38\text{sec} \tag{4.5a}$$

b. Calculate the ultimate limit of peak and integrated errors

$$E_o = \Delta F = 40000\ pph = 11\ lb/\text{sec}$$

$$E_x = \frac{1.8 * T_u}{2 * \pi * \tau_o} * E_o = \frac{1.8 * 1.38}{2 * \pi * 1} * 11 = 4.4\ lb/\text{sec} \tag{4.14b}$$

$$E_i = \frac{T_u}{2 * \pi * \tau_o} * T_u * E_o = \frac{1.38}{2 * \pi * 1} * 1.38 * 11 = 3.3\ lb \tag{4.13b}$$

Conclusions: The loop period and peak error both increased by a factor of 3.5. The accumulated error increased by a factor of 11.8 which is approximately the square of the factor for the peak error validating the general concept that the ultimate limit to the accumulated error is proportional to the dead time squared. The predicted errors are approaching those experienced in the field but are still somewhat smaller because the effects of valve dynamics and measurement noise have not been included.

6.11.3 BOILER DRUM LEVEL LOOP (INTEGRATING PROCESS)

The effect of measurement dynamics is negligible because the actual PID gain is far below the maximum allowable mostly due to the adverse effects of shrink and swell.

6.11.4 FURNACE PRESSURE LOOP (NEAR-INTEGRATING PROCESS)

Given:

(Information from Example 4.6.4)

a. Set point is 5 inches water column (w.c.) gage.
b. Measurement range is 0 to 10 inches (w.c.) gage.
c. Furnace volume is 10,000 ft3 (V_f).
d. Quench volume is 1,000 ft3 (V_q).
e. Scrubber volume is 1,000 ft3 (V_s).
f. Furnace inlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_f).
g. Quench inlet flow resistance pressure drop is 5 inches w.c. (ΔP_q).
h. Scrubber inlet flow resistance pressure drop is 10 inches w.c. (ΔP_s).
i. System outlet flow resistance pressure drop is 2.5 inches w.c. (ΔP_o).
j. Flue gas flow is 1,000 scfm (F_f).
k. Atmospheric pressure is 408 inches w.c. (P_a).
l. Disturbance is a rapid 20 percent increase in inlet pressure (ΔP).
m. Inlet pressure (discharge of forced draft fan) is 15 inches w.c. (P_i).

(New information)

n. Transmitter damping (time constant) is set to 1 sec (τ_m)

Find: ultimate limit for peak and integrated errors for an unmeasured disturbance.

Solution:

a. Calculate the new dead time and ultimate period:
From the solution to Example 4.6.4:

$\tau_s = 1.5$ sec
$\theta_p = 0.4$ sec
$K_i = 0.14$ per sec
$E_o = 0.4$ inches per sec

The dead time from the flow measurement is the fraction of the transmitter time constant converted to dead time. From Figure 4.2 we see the fraction is about 0.36 for equal time constants.

$$\theta_m = Y * \tau_m = 0.36 * 1 = 0.36 \text{ sec}$$

The new total loop dead time is the measurement dead time added to the process dead time

$$\theta_o = \theta_p + \theta_m = 0.1 + 0.36 = 0.46 \text{ sec}$$

The new secondary time constant is the original secondary time constant (1.5 sec) plus the fraction of the transmitter time constant not converted to dead time.

$$\tau_s = 1.5 + (1 - Y) * \tau_m = 1.5 + 0.64 * 1 = 2.14 \text{ sec}$$

The new ultimate period for the total loop dead time is:

$$T_u = 4 * \left[1 + \left(\frac{\tau_s}{\theta_o}\right)^{0.65}\right] * \theta_o = 4 * \left[1 + \left(\frac{2.14}{0.46}\right)^{0.65}\right] * 0.46 = 6.8 \text{ sec} \tag{4.15a}$$

b. Calculate the ultimate limit for peak and integrated errors:

$$E_x = \frac{1.8 * T_u}{2 * \pi} * E_o = \frac{1.8 * 6.8}{2 * \pi} * 0.46 = 0.9 \textit{ inches w.c.} \tag{4.20b}$$

$$E_i = \frac{T_u}{2 * \pi} * T_u * E_o = \frac{6.8}{2 * \pi} * 6.8 * 0.46 = 3.4 \textit{ inches w.c.} * \text{sec} \tag{4.19b}$$

Conclusions: The peak error increased by 225 percent and the integrated error increased by 380 percent. Transmitter damping settings must be minimized for gas pressure control, particularly when the integrating process gain is high and the scale range is low.

6.11.5 EXOTHERMIC REACTOR CASCADE TEMPERATURE LOOP (RUNAWAY PROCESS)

Given:

(Information from Example 4.6.4)

a. Set point is 150 F.
b. Measurement range is 100 to 200 F.
c. Reactant feed flow is 2,000 pph (W_r).
d. Reactant mass is 3,000 pounds (M_r).
e. Reactant heat capacity is 0.5 Btu/(lb*F) (C_r).
f. Heat transfer coefficient*area is 8,000 Btu/(hr*F) (UA).
g. Change in heat generation with temperature is 12,000 Btu/(hr*F)($\Delta Q/\Delta T$).
h. Coolant mass is 400 pounds (M_c).
i. Coolant heat capacity is 1 Btu/(lb*F) (C_c).
j. Coolant flow is 80,000 pph (W_c).
k. Disturbance is a 20 percent increase in coolant temperature (ΔT).
l. Coolant temperature is 100 F.
m. Continuous reactor.

(New information)

n. Temperature sensor is a resistance temperature detector (RTD) in a thermowell.
o. Annular clearance in the thermowell is 0.01 inch.
p. Sensor tip is touching bottom of thermowell.
q. Annular fill in thermowell is air.
r. Liquid velocity at thermowell is 1 fps.

a. Calculate the new ultimate period for the measurement time constants.

Even though Table 6.2 is for thermocouples, we can use the table for RTDs by just increasing the smaller time constant by two seconds to reflect the slower response of the RTD sensor inside the thermowell. This gives us time constants of 24 and 6 seconds.

$$\theta_m = Y_1 * \tau_{m1} + Y_2 * \tau_{m2} = 0.6 * 24 + 0.9 * 6 = 20 \text{ sec} = 0.33 \text{ min}$$

The new total loop dead time is the measurement dead time added to the process dead time

$$\theta_o = \theta_p + \theta_m = 0.3 + 0.33 = 0.63 \text{ min}$$

The new secondary time constant is the original secondary time constant (three minutes) plus the fraction of the sensor time constants not converted to dead time.

$$\tau_s = 3 + (1 - Y_1) * \tau_{m1} + (1 - Y_2) * \tau_{m2} = 3 + 0.4 * 0.4 + 0.1 * 0.1 = 3.16 \text{ min}$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D} \right)^{0.65} \right] * \theta_o \qquad (4.21a)$$

$$N = \left(\tau_p' + \tau_s \right) * \left(\tau_p' * \tau_s \right) = (30 + 3.16) * 30 * 3.16 = 3144 \qquad (4.21b)$$

$$D = \left(\tau_p' - \tau_s \right) * \left(\tau_p' - \theta_o \right) * \theta_o = (30 - 3.16) * (30 - 0.63) * 0.63 = 496 \qquad (4.21c)$$

$$T_u = 4 * \left[1 + \left(\frac{N}{D} \right)^{0.65} \right] * \theta_o = 4 * \left[1 + \left(\frac{3144}{496} \right)^{0.65} \right] * 0.63 = 10.9 \text{ min}$$

b. Compute the ultimate gain from the ultimate period and the primary and secondary time constants.

$$K_u = \frac{\sqrt{1 + (\tau_p' * 2 * \pi / T_u)^2} * \sqrt{1 + (\tau_s * 2 * \pi / T_u)^2}}{K_o} \qquad (4.22c)$$

$$K_u = \frac{\sqrt{1 + (30 * 2 * \pi / 10.9)^2} * \sqrt{1 + (3.16 * 2 * \pi / 10.9)^2}}{K_o} = \frac{18.7 * 2.1}{K_o} = \frac{20.8}{K_o}$$

c. Calculate the ultimate limit for peak and integrated errors

$$E_x = \frac{1.1}{0.6 * K_u * K_o} * E_o = \frac{1.8}{\frac{20.8}{K_o} * K_o} * 53 = 4.6 \ {}^oF \qquad (4.24a)$$

$$E_i = \frac{0.6 * T_u}{0.6 * K_u * K_o} * E_o = \frac{T_u}{K_u * K_o} * E_o \qquad (4.23a)$$

$$E_i = \frac{T_u}{K_u * K_o} * E_o = \frac{10.9}{\frac{20.8}{K_o} * K_o} * 53 = 27.8 \ {}^oF * \text{min}$$

Conclusions: The peak error increased by 328 percent and the integrated error increased by 488 percent. The thermowell time constant should be minimized particularly when the positive feedback time constant is less than factor of 10 greater than the secondary time constant.

6.11.6 BIOLOGICAL REACTOR BIOMASS CONCENTRATION LOOP (RUNAWAY PROCESS)

The disturbances originating from cell growth and product formation are so slow that on-line measurement lags are inconsequential except in terms of narrowing the window of allowable gains. At-line analyzer cycle times of up to eight hours are even acceptable for mammalian cell cultures whose batch times are about two weeks.

6.12 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in setpoint (ΔSP) or load flow at the process input (ΔF_L). Table 6.5 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The test cases use aggressive tuning settings computed with the short cut method to maximize disturbance rejection. Since these settings are right on the edge of causing an oscillation, the second effect of measurement dynamics occurs besides the first effect.

Figures 6.4a to 6.4i show the effect of measurement threshold sensitivity (S_m) on the ability of the PID controller to deal with a load disturbance. For these test cases, the measurement resolution limit was 0 percent.

The test reveal decreasing the PID gain (e.g., cutting the gain in half) or eliminating derivative action (setting the rate time to zero), actually made the situation worse. The peak error and settling time increased due to the increase in total loop dead time from the additional dead time from the threshold sensitivity limit. The immediate reaction to a disturbance provided by the proportional and derivative modes, drove the process variable faster reducing the amount of time to get through the measurement threshold sensitivity limit. The deterioration from less proportional and derivative action increased as the process non-self-regulation decreased. For moderate self-regulating processes, the deterioration was mostly observed as an increase in peak error. As the test moved to near-integrating and then a true integrating process, the process developed an increasingly severe damped oscillation before settling into the limit cycle.

The best tuning solution as per test results was to double the reset time. The increase in reset time did not appreciably affect the peak error but made the approach back to setpoint

Table 6.5. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
6.4a, b, c	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 10\%$	Threshold sensitivity
6.4d, e, f	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Threshold sensitivity
6.4g, h, i	True integrating	0.01 1/sec	10	—	$\Delta F_L = 20\%$	Threshold sensitivity

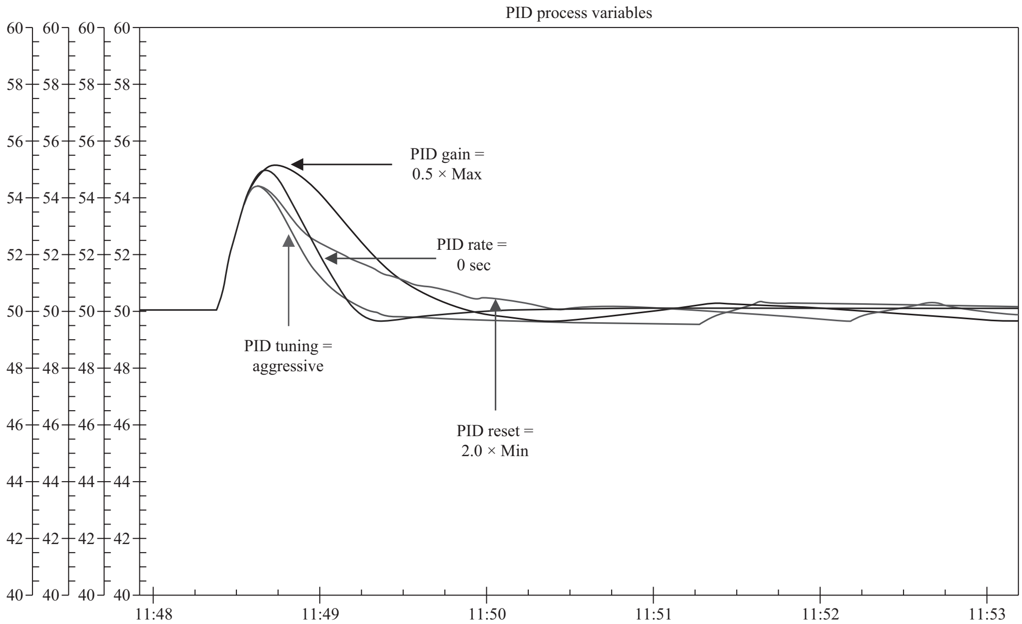


Figure 6.4a. Effect of 0.5% *threshold sensitivity* on *actual PV* for load upset to *moderate self-regulating* process.

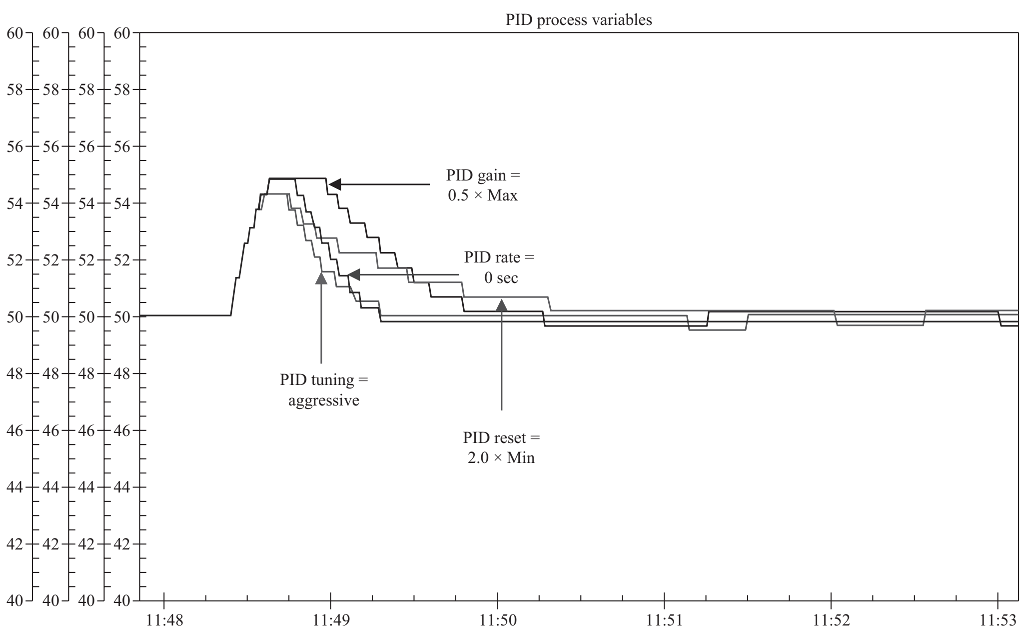


Figure 6.4b. Effect of 0.5% *threshold sensitivity* on *measured PV* for load upset to *moderate self-regulating* process.

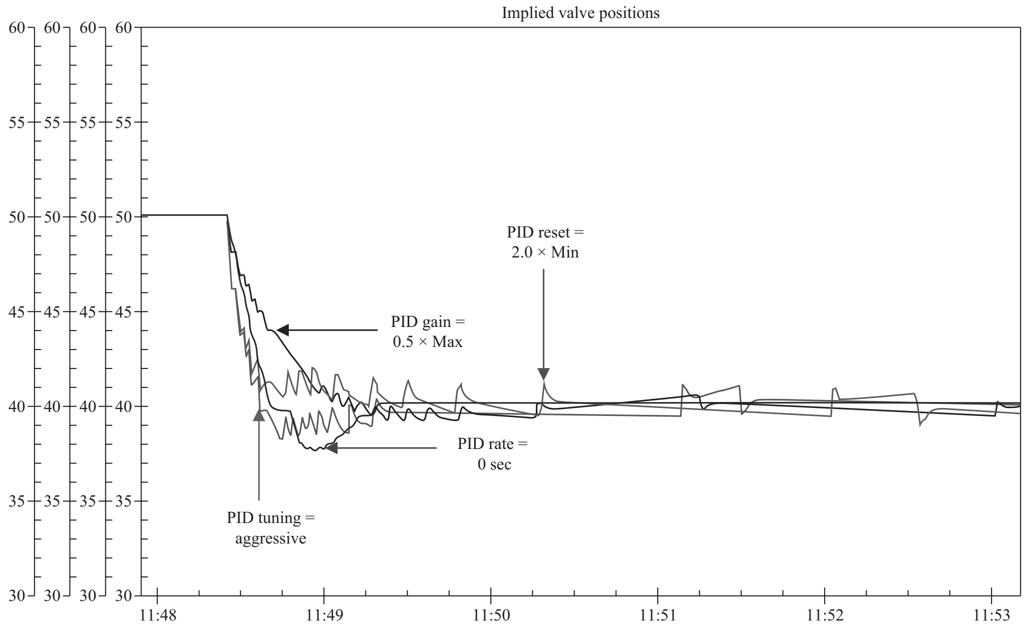


Figure 6.4c. Effect of 0.5% *threshold sensitivity* on *valve* for load upset to *moderate self-regulating* process.

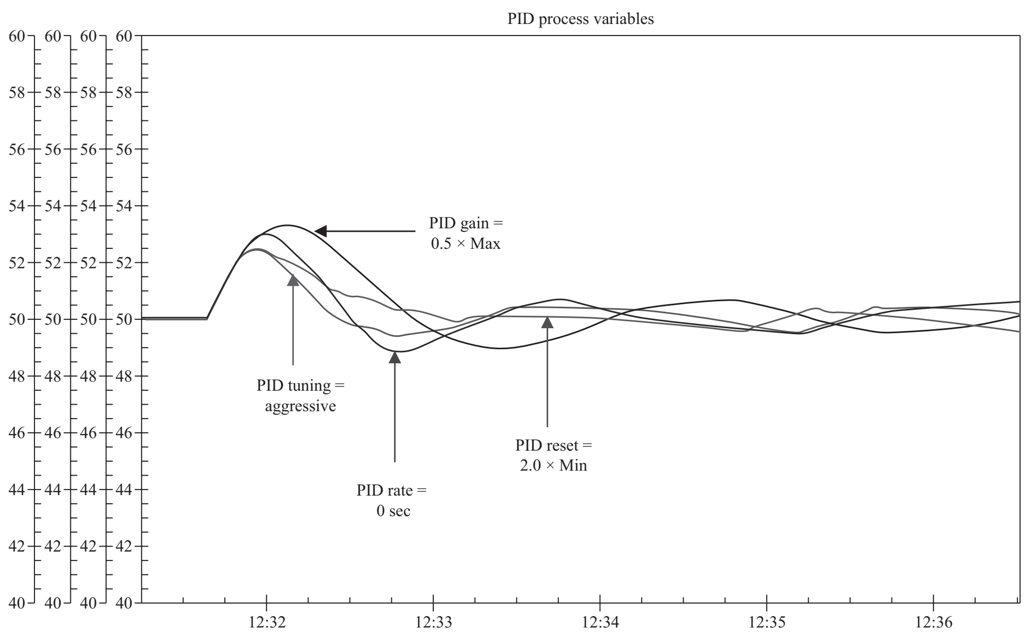


Figure 6.4d. Effect of 0.5% *threshold sensitivity* on *actual PV* for load upset to *near-integrating* process.

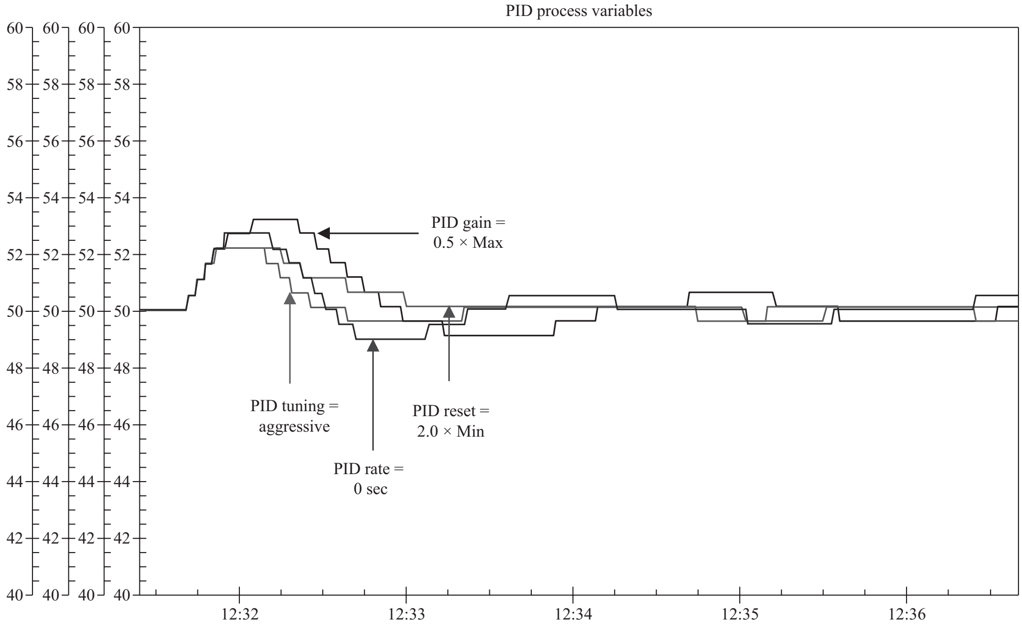


Figure 6.4e. Effect of 0.5% *threshold sensitivity* on *measured PV* for load upset to *near-integrating* process.

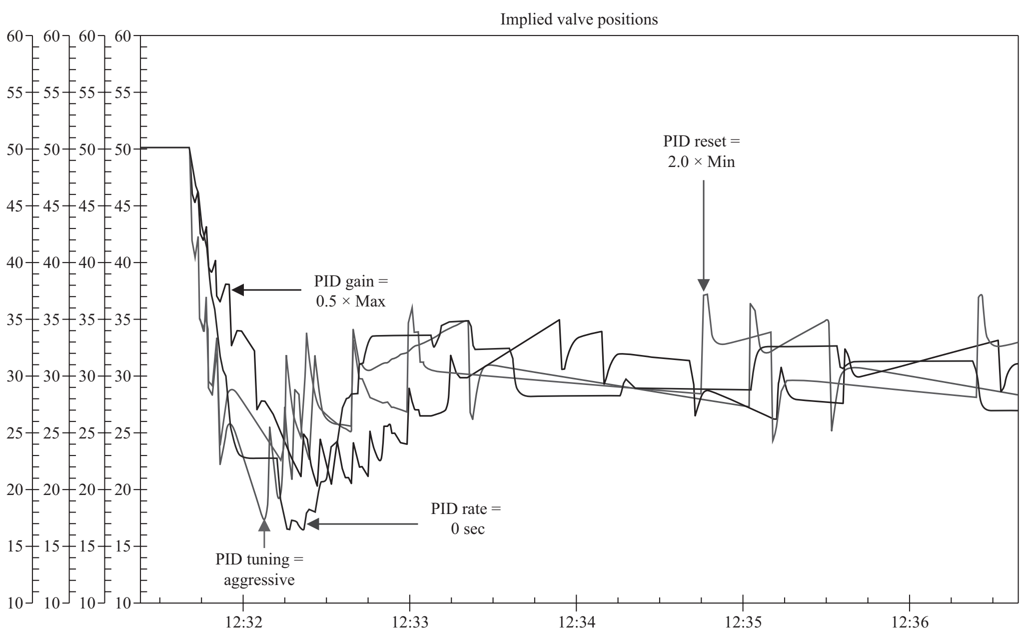


Figure 6.4f. Effect of 0.5% *threshold sensitivity* on *valve* for load upset to *near-integrating* process.

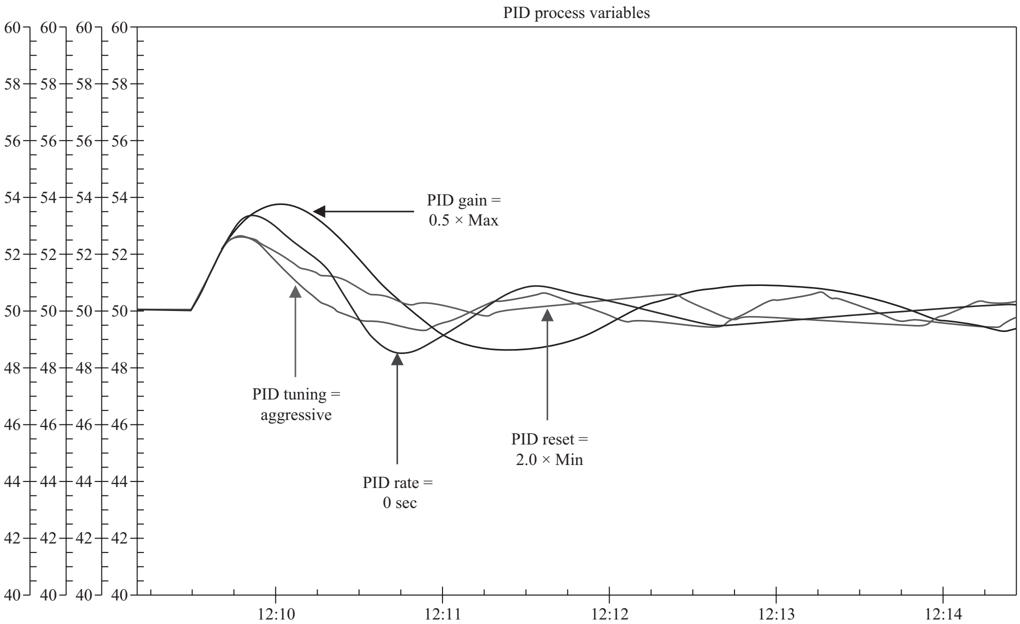


Figure 6.4g. Effect of 0.5% *threshold sensitivity* on *actual PV* for load upset to *true integrating* process.

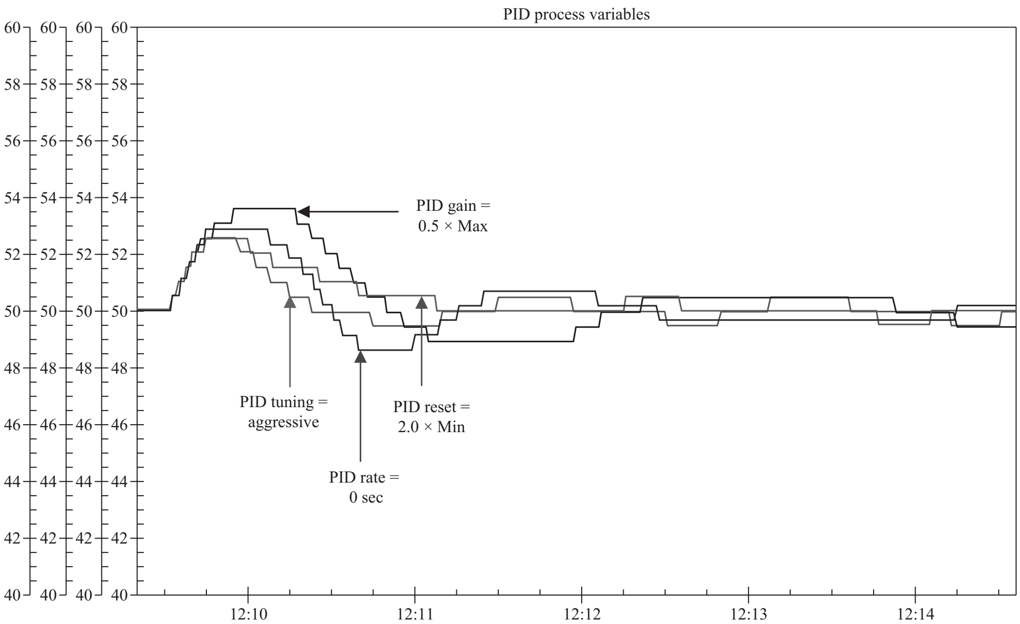


Figure 6.4h. Effect of 0.5% *threshold sensitivity* on *measured PV* for load upset to *true integrating* process.

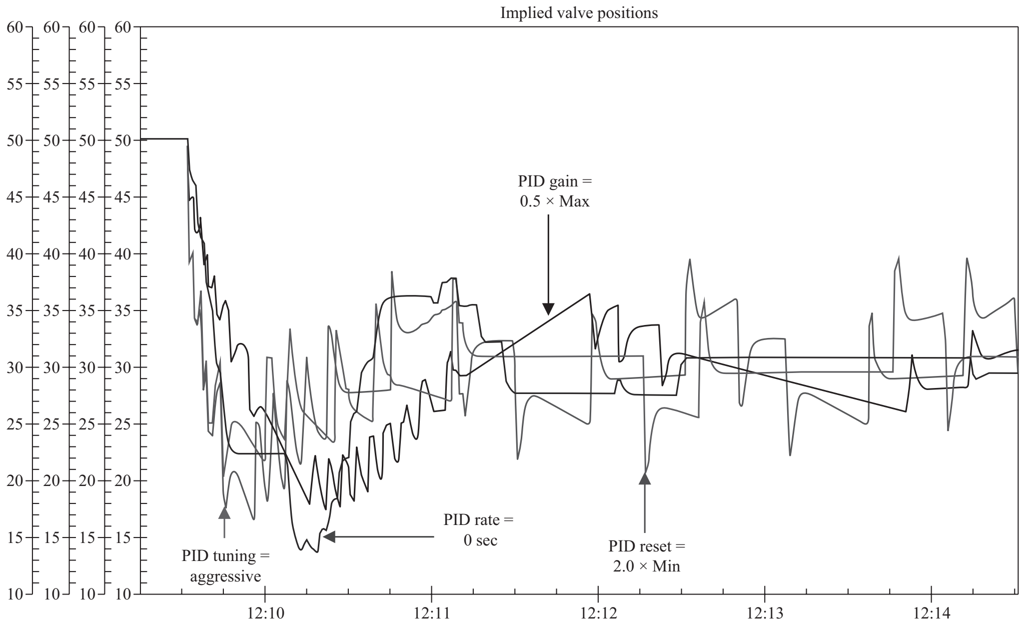


Figure 6.4i. Effect of 0.5% *threshold sensitivity* on *valve* for load upset to *true integrating* process.

smoother and increased the period of the limit cycle. There was no additional damped oscillation for this case even for the true integrating process. The gradual change in the measured process variable resulted in less frequent updates that translated to fewer changes in the actual valve position. Before the reset time was increased, excessive valve dither was observed in the recovery from the disturbance.

KEY POINTS

1. The measurement is the only window into the process. The view needs to be timely and sharp.
2. The common sources of measurement dead time in a general order of decreasing size are analyzer cycle and multiplex times, transportation delays, pH sensor lags, thermowell lags, wireless default update rates, transmitter damping, and sensor threshold sensitivity and transmitter or input card resolution limits.
3. Given a fast valve or VSD, the largest source of dead time comes from the measurement for the control of flow, pressure, pH in static mixer, temperature in gas unit operations (e.g., catalytic reactors, reformers, and furnaces), and concentration using liquid at-line analyzers.
4. Changes in phase (e.g., gas, liquid, solid) create noise in nearly all measurements.
5. Imperfect mixing is the source of noise in concentration, pH, and temperature measurements.
6. The elimination of impulse lines and capillary systems enable a smart transmitter to achieve the accuracy stated in the manufacturer's specification.

7. Drift in modern measurements have largely been eliminated except for analyzers, pH electrodes, and thermocouples.
8. Precision affects lower control loops more than accuracy or drift.
9. Accuracy, precision, and drift affect the calculation of process efficiency.
10. The best tuning compensation for a measurement threshold sensitivity limit is to increase the reset time. Decreasing the PID gain or rate will increase the dead time from the limit and unexpectedly increase valve dither in the load response.
11. Runaway processes are much more sensitive to measurement time constants. The window of allowable controller gains can close for exothermic reactors due to a large thermowell lag.
12. A measurement time constant can create an illusion of reduced variability amplitude. Trend charts may look better, but the actual process is doing worse.

CHAPTER 7

Effect of Valve and Variable Frequency Drive Dynamics

7.1 INTRODUCTION

The control valve and variable frequency drive (VFD) are the final control elements that directly affect the process by manipulating a flow. The expectation is that these elements do their job and do not adversely affect the tuning and performance of the loop. Most control text books do not include details of the final control elements in the analysis or solution. Until recently, most users were not aware of the potential problems. Here is a perspective, overview, and recommendations.

7.1.1 PERSPECTIVE

Prior to the 1980s, nearly all final control elements were control valves supplied by throttling valve manufacturers. Since these valves were originally designed for throttling service, backlash and stick-slip were minimal. Actuators and positioners with excellent threshold sensitivity were generally used. Since energy conservation and the possible use of VFD were less of a motivation, the valve to system pressure drop ratio was large enough to provide a good installed flow characteristic and rangeability. The downside was that the positioners being pneumatic devices were out of calibration typically resulting in an offset between actual position and implied valve position based on the valve signal. Since there was no readback, the user was typically unaware of positioner problems. Fortunately, the main problem being simply an offset was automatically corrected by the control loop manipulating the valve to reach setpoint.

These control valves with a throttling valve heritage typically had a deadband less than 0.4 percent, stick-slip less than 0.2 percent, and an actuator-positioner resolution or threshold sensitivity of better than 0.1 percent. Unfortunately, this capability was taken for granted and never put on the valve specification as a requirement. The lack of understanding of valve dynamics and the missing link in terms of readback set the scenario for a disaster in valve performance for decades except where company and plant standards required the time proven valve solution despite a higher price tag.

The typical control valve specification had sizing information and a leakage specification. There was no information on installed characteristic or valve dynamics. Since the main goal was making sure the valve could pass the required flow and stop the flow as completely as possible, valves designed by on-off valve (e.g., piping and isolation valve) manufacturers met the specification with a significantly lower price tag. In many cases the valves were already in the piping spec and being used manually by operators or automatically by batch sequences and safety instrumented systems (SIS).

The on-off valves had excessive deadband (e.g., 8 percent) from backlash associated with links and connections between shafts, stems, and internal closure elements (e.g., balls and discs). These valves also had excessive stick-slip (e.g., 4 percent) from the high seating-sealing friction and stem friction from packing designed for high temperatures and lower emissions. Poor resolution and threshold sensitivity (e.g., 2 percent) also originated in the accessories due to the use of piston actuators with O-rings or gear teeth instead of diaphragm actuators, high volume spool positioners and volume boosters instead of high gain relay positioners. Rotary valves were used with a nearly quick opening flow characteristic because these were a very low cost high capacity solution particularly attractive in larger line sizes. The steep slope of the flow characteristic near the closed position where the stick-slip was the highest from high seating-sealing friction associated with tight shutoff, caused larger amplitude limit cycles. The term *high performance* was often used for these valves because performance was judged in terms of tightness of shutoff and capacity. Who wouldn't want a *high performance* valve at a much lower cost? This deception has caused untold problems for over 40 years.

Some applications tried to take advantage of the control valve having low leakage by making it serve the dual purpose of isolation and throttling. The lack of position readback and the standard testing procedure of making 25 or 50 percent changes and eyeballing the stem position did not show the problem created by the use of on-off valves. The limit cycles in the process were often attributed to some other problem.

While awareness has improved by the ISA standards *Test Procedure for Control Valve Response Measurement from Step Inputs* (ISA-75.25.01) and *Control Valve Response Measurement from Step Inputs* (ISA-75.25.01), many users still can get into trouble because there is very limited response data published. What data does exist is from tests at mid position with hand tight packing to eliminate sealing and seating friction and minimize stem friction. Furthermore, original throttling valve manufacturers seeing the profits generated by on-off valve manufacturers have bought on-off valve companies. To become more cost competitive, throttling valve manufacturers may quote a control valve with an on-off heritage or use the smallest possible actuator size making the valve more susceptible to stick-slip particularly near the closed position or as packing is tightened.

To make matters more confusing, the readback of actual valve position for on-off valves posing as control valves is from the shaft position instead of actual internal closure element position. Tests by the author in separate applications revealed smart positioners saw and reported only a 0.6 percent backlash in the test results when the use of a travel gage indicated 8 percent backlash in the ball or disc position. This deception caused by the backlash in the connections between the shaft, stem, and internal closure element is widespread.

Dampers on large gas flows tend to have even more backlash than on-off valves because of greater play (slop) in the linkages. On-off actuators and positioners used on dampers tend to have a poor resolution and threshold sensitivity.

VFD do not have backlash or stick-slip. However, an excessive deadband and rate limiting is often introduced due to an over concern about hunting and motor load. Also, the standard input card receiving the proportional-integral-derivative (PID) output signal had a resolution of 0.4 percent. A special input card had to be requested from the VFD manufacturer to get the resolution as good as the resolution of the modern day Distributed Control Systems (DCS) output card.

Before a valve can move, enough air must move into or out of the actuator to create a sufficient change in force. This time interval is called a pre-stroke dead time. Once the valve starts to move there is a stroking time (time for 100 percent stroke) with a lag. For small actuator volumes this dead time and lag is less than 0.2 seconds and the stroking time is less than two seconds. Note that the slewing rate is 100 percent divided by the stroking time. The ISA standards use 86 percent response time (T86) as a test criterion, which is the time to reach 86 percent of the final response. For extremely large actuators, these times can be much larger unless volume boosters are added to the positioner output. The ISA standards combine stick-slip and resolution and threshold sensitivity limits together simply as a resolution limit. Note that the response time includes the effect of threshold sensitivity limits.

Deadband, stick-slip, resolution and threshold sensitivity limits cause limit cycles and dead time in addition to the pre-stroke dead time. The limit cycles (sustained equal amplitude oscillations) develop from the integrating action in the process or control system. For deadband, two or more integrators are required to create a limit cycle. The additional dead time can be approximated as the deadband plus stick-slip (threshold sensitivity or resolution limit), divided by the rate of change of PID output.

The VFD is often thought to provide a change in flow that is linear with speed. For this case the rangeability and linearity is impressive. Often not recognized is that when the discharge head approaches the static head as the speed is lowered, the change in flow with speed becomes larger and erratic. If the discharge head becomes less than the static head, a dangerous reversal in flow can occur. A check valve is advisable in this case if the check valve is reliable at operating conditions. As a result, the installed characteristic becomes steep and a low speed limit is added reducing the controllability and rangeability of the VFD.

While the inherent dead time in the VFD response is negligible, additional dead time is created by introduction of deadband in the drive setup or a resolution limit in the signal input card. The dead time created is the deadband plus the resolution limit divided by the rate of change of the PID output.

If the speed control is slowly tuned or moved from the field into the DCS, the speed loop is not sufficiently faster than the process loop. This violation of the cascade rule where the secondary loop is not five times faster than the primary loop, causes poor performance. If external reset feedback is not enabled on the primary loop, an insidious burst of oscillations will occur for large disturbances or setpoint changes. This same problem can occur for valves when positioners are sluggishly tuned or volume boosters are not used on large actuators. See Chapter 5 for more details on the use of external reset feedback to prevent the primary PID output from changing faster than the secondary controller process variable (speed or stroke) can respond.

7.1.2 OVERVIEW

If your company or plant has strong standards as to the use and testing and response of throttling control valves, you probably don't need to pay much attention to valve dynamics except for applications where a fast valve response is needed such as surge control or pressure control or where there are exceptionally large line sizes or high pressure drops. If you don't have the protection of these standards, you are vulnerable to undersized actuators or the use of on-off valves posing as throttling valves. You need to become educated on all the gotchas.

Since the focus of VFD manufacturers is on energy savings, there is very little understanding and guidance offered by the supplier on dynamic problems created by cascade control, deadband, speed rate limiting, and static head. Design and maintenance engineers must develop the knowledge and assume the responsibility for making sure the VFD response is fast and the installed flow characteristic is linear.

7.1.3 RECOMMENDATIONS

1. For surge control valves and for systems where the valve to system pressure drop ratio is greater than 0.8, use a linear inherent characteristic, otherwise use an inherent equal percentage characteristic.
2. Ensure the valve to system pressure drop ratio is greater than 0.2 and large enough to meet rangeability requirements.
3. Select valve type and size to avoid operation on the upper flat portion of the installed flow characteristic.
4. Keep the largest of threshold sensitivity and resolution limit, stick-slip, and half the deadband less than 0.2 percent by valve, positioner, and actuator design.
5. Add a separate on-off valve for isolation, batch charge shutoff and to minimize reagent injection delays (volume between tight shutoff valve and process pipe connection) that is coordinated with the opening and closing of the control valve.
6. Tune the positioner for a fast but non oscillatory response. Only use integral action in the positioner as a last resort to create a fast limit cycle completely attenuated out by a slow process response (e.g., temperature).
7. If a volume booster is needed for large actuators, put the booster on the positioner output and open the booster bypass just enough to stop the very fast oscillation.
8. Keep the speed control for a VFD in the field and tune the speed controller for a fast non-oscillatory response.
9. If the position or speed control is not fast enough, add a fast external reset feedback of actual valve position or VFD speed to the process PID.
10. Make sure the VFD deadband is less than 0.4 percent and the resolution limit of the input card is less than 0.2 percent and the speed rate limit is fast enough.
11. Make sure the lower speed limit will always prevent the pump or fan curve discharge head above the static head.
12. Ensure the static head is sufficiently less than the system drop to meet VFD rangeability requirements and if necessary add system resistance and pump head.

7.2 VALVE POSITIONERS AND ACCESSORIES

A valve positioner provides feedback control of valve position. The position of a shaft is the controlled variable. The setpoint is the valve signal that is the output of a process PID controller forming a cascade control system. The analog valve positioner was principally a proportional only controller relying upon a high gain (e.g., 100) for position control.

7.2.1 PNEUMATIC POSITIONERS

Pneumatic positioners persisted well into the days of electronic instrumentation and the DCS because the actuator was pneumatic. For pneumatic positioners, the analog or digital signal first has to be converted to a 3 to 15 psi by a current to pneumatic (I/P) transducer. This positioner was principally limited to proportional-only control by means of links, levers, bellows, flappers, and nozzles. The time required for even good pneumatic positioner designs to respond increased from 1 to 100 seconds or more for small changes in the signal (e.g., <0.2 percent). For spool type positioners, there may be no response for small changes. Adjustment of the proportional gain was limited or nonexistent. The proportional gain and the resolution capability depended upon the pneumatic design. A pneumatic relay design provided better sensitivity and a higher gain than the inexpensive spool design. Spool type positioners with poor shaft and stem connection designs were often used for throttling on-off valves and dampers. Pneumatic positioners were difficult to calibrate and readily lost their calibration. Pneumatic positioners do not provide a valve position signal (travel readback or feedback), alerts, and diagnostics so the user was unaware of calibration or performance problems. In order to get a position indication in the control room, a separate position transmitter had to be installed. Limit switches that indicate full open or closed position were used for batch operation, startup, and shutdown, but position transmitters were rare. It was only after the advent of electronic positioners that users started to become aware of the valves as a major source of process variability next to controller tuning. Existing pneumatic positioners should be replaced with smart positioners known as digital positioners (digital valve controllers).

7.2.2 DIGITAL POSITIONERS

The digital positioner manipulates the supply and exhaust flow to the actuator via a pneumatic relay. The exhaust flow capacity is generally higher than the supply capacity to provide faster stroking to the air failure position. The digital positioner accepts analog or digital input signals. The current to pneumatic (I/P) transducer is functionally moved from the input to the output of the controller. The I/P transducer converts the microprocessor based controller output to a pneumatic signal that is amplified by a pneumatic relay. The pneumatic relay is like a self-contained pressure controller. The relay adjusts the supply or exhaust flow to achieve an outlet pressure that matches the desired pressure set by the I/P transducer. The relay supplies air flow to increase actuator pressure and exhausts air flow to decrease actuator pressure. The relay has two outputs. In Figure 7.1b, relay output “1” is direct acting (actuator pressure increases with valve signal) and relay output “2” is reverse acting (actuator pressure decreases with valve signal). For zero percent signal output “1” is at zero psig and output “2” is at full supply pressure. To make the electrical signal failure position consistent with the air failure position, relay output “1” is connected to the diaphragm or single-sided piston. For a double acting piston, relay output “1” is connected to the piston end that drives the valve open and relay output “2” is connected to the piston end that drives the valve closed.

Digital positioners and air supply filter regulators are an essential part of every control valve. The air must be clean and dry. The air supply filter should only have to remove infrequent small particles. The air supply filter regulator is normally integrally mounted on the digital positioner. The air supply regulator maximum setting is much higher for piston actuators (e.g., 120 psig) than standard diaphragm actuators (e.g., 40 psig) but is only slightly higher than the new high pressure diaphragm actuators (e.g., 90 psig). A higher operating pressure enables a smaller volume actuator to generate the same thrust for sliding stem and torque for rotary actuators.

The digital valve controller comes with a recommended set of tuning setting values for different types of actuators and valves. These tuning sets generally have an increasing travel gain and rate setting as the size of the valve increases. The user can stay with the recommended tuning set or go to the expert mode and enter settings for more aggressive control. For example, the gain and pressure rate settings might be increased for high friction valves to reduce stick-slip. Integral action is normally not enabled. If integral action is used, an integral deadband should be set to stop limit cycles.

7.2.3 CURRENT TO PNEUMATIC (I/P) TRANSDUCERS

The current to pneumatic (I/P) transducer on the controller output was extensively used before the digital control system and digital positioner became common. The I/P output was the input to a pneumatic positioner or the input to a diaphragm actuator. A digital positioner directly accepts a current, fieldbus, HART, or wireless input so the only reason today to use an I/P is if you didn't want a positioner. The tuning flexibility and the diagnostic capability of smart digital positioners (e.g., digital valve controllers) makes them applicable and beneficial for all loops eliminating the need for an I/P.

7.2.4 SOLENOID VALVES

The most common accessory for discrete positioning is a solenoid valve to put a control valve or damper in a fixed position for safety or to achieve a sequence of flows for startup, shutdown, batch operation, or product transition. A solenoid valve is normally piped or tubed on each output of the positioner being used. For fail safe action, an open contact from a DCS, PLC, or SIS will de-energize the solenoid valve. The solenoid valve selection and installation is designed to ensure the valve goes to the safe position for electrical failure, wiring failure (open circuit), and air failure.

For the spring return pneumatic actuator in Figure 7.1a, a de-energization of the three-way solenoid valve SV-1 will exhaust the air to trip and achieve the desired electrical and air failure position. For the double acting piston actuator in Figure 7.1b, the de-energization of three-way solenoid valve SV-1 exhausts and the de-energization of three-way solenoid valve SV-2 supplies, respectively air to the sides of the piston needed to trip and achieve the desired failure position. For normal operation, the flow path (solenoid energized) is open between ports A and B. For a trip, the flow path is open between ports B and C. To provide the desired trip and fail safe position, port C of SV-1 is vented to atmosphere and port C of SV-2 is connected to a pressurized air tank reservoir. A check valve prevents a high pressure in the volume tank from causing a reverse flow into the air supply system.

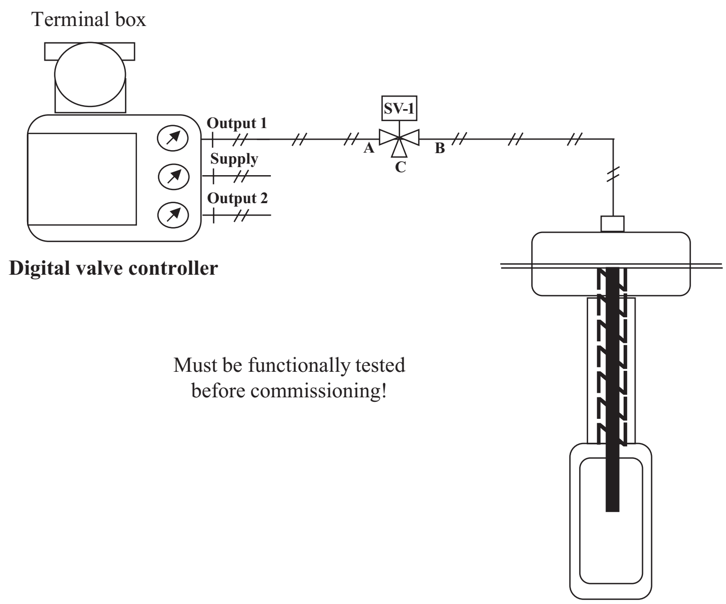


Figure 7.1a. *Diaphragm actuator* with digital positioner and *solenoid valve* (shaft retracts on electrical and air failure).

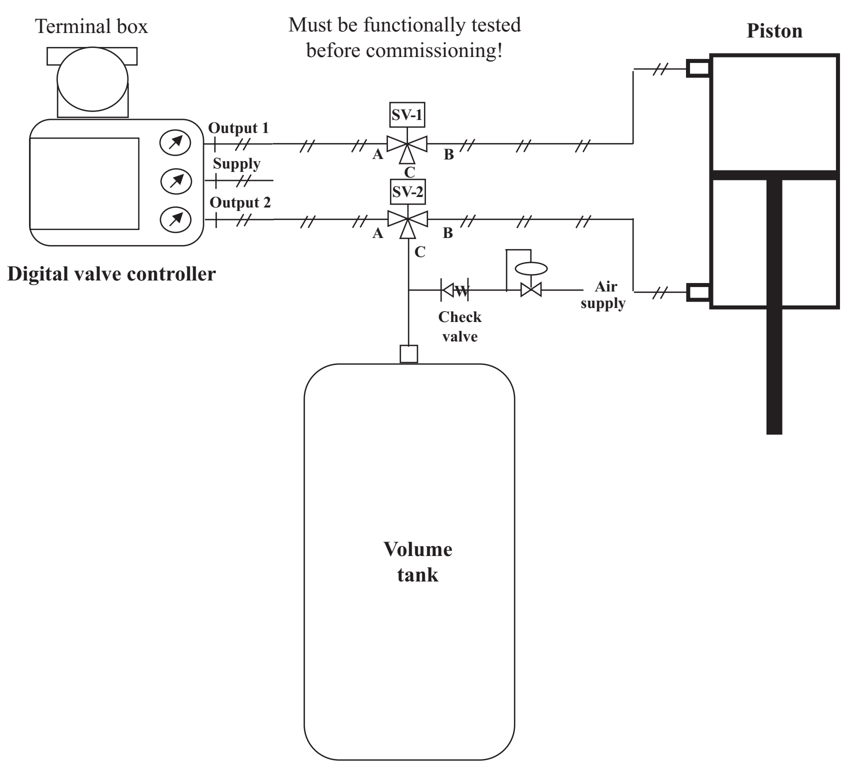


Figure 7.1b. *Double acting piston actuator* with digital positioner and *solenoid valves* (shaft retracts on electrical and air failure).

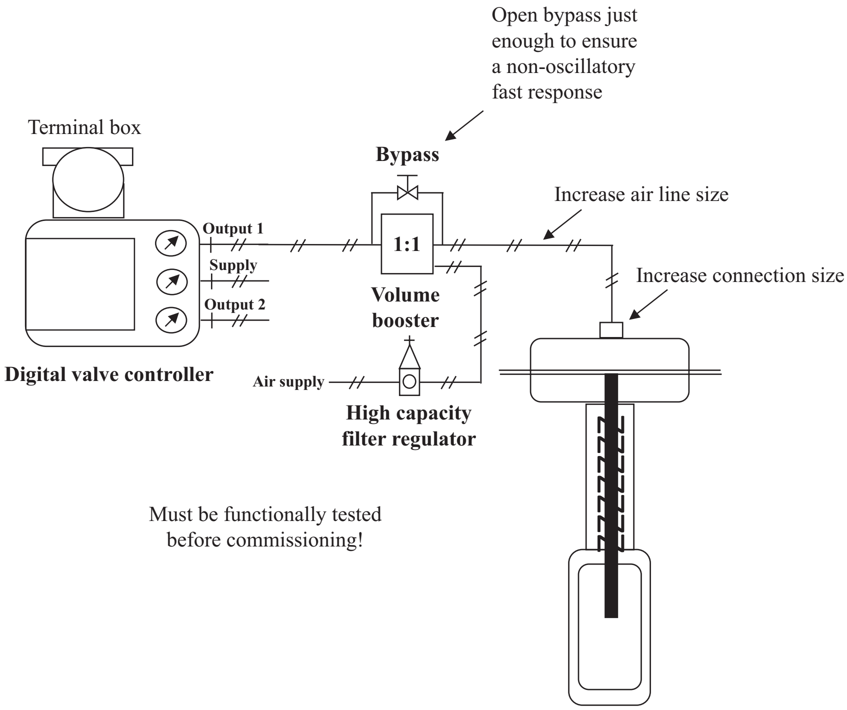


Figure 7.1c. *Diaphragm actuator* with digital positioner and *volume booster* (shaft retracts on electrical and air failure).

7.2.5 VOLUME BOOSTERS

A booster should not be used in lieu of a positioner but in conjunction with a positioner to maintain actuation stiffness, sensitivity, and a consistent input signal range for stroking the valve. The booster is mounted on each output port of the positioner. For double acting pistons, two boosters are required. The booster must have a bypass as noted in Figure 7.1c to prevent instability. If the booster does not have an integral bypass valve as shown in Figure 7.2, an external needle valve in parallel with the booster must be installed to bypass some of the air from the input to the output of the positioner. Since the positioner is designed to be looking into a relatively large volume of an actuator, if the bypass is closed or omitted, the extremely small volume of the booster inlet port will cause a rapid limit cycle. The bypass must be opened until the high frequency oscillations (e.g., 1 to 2 cps) stop. The bypass may be opened slightly further to provide a stability margin. Since the booster bypass slows down the speed of actuator, it is desirable to open the bypass just enough to ensure stability. If the tuning of the digital positioner is changed, the booster bypass may need to be adjusted accordingly. More aggressive tuning settings will often require a larger bypass flow.

The combination of a positioner and booster can dramatically reduce the pre-stroke dead time and stroking time of large actuators. The air pipe or tubing size must be increased and in some cases the diaphragm actuator casing connection enlarged. The flow capacity of accessories, such as solenoid valves and air filter regulators, must be accordingly increased. Otherwise, the true flow capability of the booster is restricted.

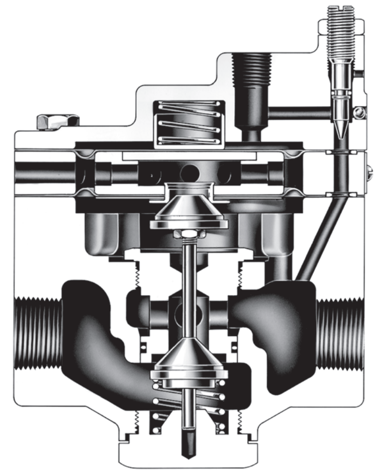


Figure 7.2. Volume booster with integral bypass.

Boosters have poor inlet (signal) port sensitivity. Consequently, the booster will not respond to small changes in signal. For slowly changing signals, a dead time is added that is proportional to the booster deadband or resolution limit divided by the rate of change of the input signal. Since most PID controllers are making small changes in output for each PID execution, we have the ironic situation where a booster that is added to speed up the response of a valve actually prevents it from moving. The large step changes (e.g., >5 percent) that are normally used to test valve response do not reveal the problem.

Boosters are designed to have an exceptionally high outlet port sensitivity to maintain the output pressure. The use of them without positioners on diaphragm actuators has resulted in large butterfly valves slamming shut due to fluid forces. In fact a person can grab the shaft of an 18 inch butterfly valve with a booster (no positioner) and move the shaft. There is a positive feedback from slight pressure changes associated with slight volume changes for a flexible diaphragm actuator. The negative feedback provided by valve position control counteracts the positive feedback in the diaphragm actuator response.

The lack of understanding the implications of high outlet port sensitivity, poor inlet port sensitivity, bench settings, and fluid-forces has led to the misguided rule that boosters instead of positioners should be used on fast loops. The digital valve controller and PID in a modern DCS can be tuned to prevent the problem observed in analog control systems from violation of the cascade rule (lower loop five times faster than upper loop).

7.3 ACTUATORS, SHAFTS, AND STEMS

Pneumatic actuators used in 99 percent of the loops are the subject of this book. The other types of actuators are just briefly addressed here. Hydraulic and motor actuators used for very high pressure drops and temperatures have high capital and maintenance costs. Hydraulic actuators are so fast that their contribution to the loop dynamics is negligible. In contrast, conventional motor driven actuators have such a low rate of change limit that the slow response is the major limitation to the loop's capability. Servo type actuators are extremely fast but are rarely used in the process industry.

7.3.1 DIAPHRAGM ACTUATORS

Diaphragm actuators offer advantages in terms of reliability, maintainability, and resolution. The diaphragm provides an air tight seal that does not wear out with travel. Since the diaphragm ends are attached and stationary, the friction associated with the movement (diaphragm flexure) is negligible. The diaphragm actuator is able to respond to small changes in air pressure.

The diaphragm actuator pressure must overcome the spring force. The traditional diaphragm actuator was limited to about 30 psig. Diaphragm actuators have been developed that use 90 psig and a positioner to provide a more compact and powerful package. Diaphragm actuator size and cost increase as valve size and pressure drop increase. For very large valves (e.g., >24 inch) or very high pressure drops (e.g., >1,000 psi), a pneumatic piston is generally needed.

7.3.2 PISTON ACTUATORS

The pneumatic piston actuator is widely used for on-off valves and for throttling large control valves and dampers. Piston actuators can be smaller than diaphragm actuators.

Pistons can be single sided where the pressure is loaded on one side of the piston to oppose the force of a spring similar in concept to a spring return diaphragm actuator. The more common approach is a double sided (double acting) piston where air pressure is loaded on both sides of the piston, There may be a small bias spring to directionally put the valve in the correct air fail position. However, to ensure an air fail position, a volume tank (reservoir) is needed as shown in Figure 7.1b.

Motion of the double acting piston depends upon a difference in air pressure across the piston. Positioners are necessary to load and balance the high pressures in these piston actuators so the question of whether to use a positioner on fast loops goes away. There are no bench settings but there is a cross over pressure that is rarely set properly. The cross over pressure is the balance point where motion ceases. The pressure on one side of the piston must differ from the cross over pressure by enough to exceed the unbalance from pressure drop and frictional forces posed by piston seals and valve stem and trim seals. If the cross over pressure is set too high, the pressure in the piston may approach the supply pressure to overcome unbalance and frictional forces. If the cross over pressure is set too low, there is not enough muscle for the piston making the valve trim more susceptible to buffeting from fluid forces and flutter of the plug or disc.

Cylindrical Piston actuators require extremely smooth surfaces, precision machining and guiding, and lubrication of the O-ring to minimize the friction between the piston and cylinder walls. Wear and damage, rust, and improper lubrication can cause excessive friction and poor resolution. The O-rings also need to be periodically replaced.

The piston actuator doesn't respond to as small a change in air pressure as a diaphragm actuator. Consequently, the minimum positioning resolution is not quite as good for a piston actuator even in the best of conditions. Also for comparable sizes and the positioner capacity, the pre-stroke time delay required to make a significant pressure change is larger for a piston actuator.

7.3.3 LINKAGES AND CONNECTIONS

The shaft of the actuator and the stem of the internal closure element more commonly known as trim (e.g. plug, ball, or disc) are normally separate. The internal element may be cast and forged with the stem or the stem may be connected during valve assembly. The actuator shaft moves the stem that moves the internal trim. While *shaft* and *stem* are more appropriate terms for the actuator and internal element, respectively, in practice the terms *stem* and *shaft* are used interchangeably. The tightness or the amount of play (slop) of the connections between the shaft, stem, and internal element, determine how well the valve responds to small changes in signal. The location and type of connection of the positioner feedback mechanism for valve travel determines whether the positioner is seeing the response of just the actuator or the actual internal trim.

Previous methods of testing valve response involved making changes in the valve signal much larger than normally made in closed loop control. Most valves will look OK for the large changes in requested position. The change in PID output for each PID execution is generally small (e.g., <0.2 percent) except during the start of an operation or process. For small changes the resolution (stick-slip) and backlash (deadband) that prevent a good response and creates sustained oscillation (limit cycle) comes into play (see Section 7.5).

Sliding stem (globe) valves have the least amount of backlash (deadband) because of the direct connection between the actuator shaft and trim stem and low trim friction. For rotary valves, connections are problematic since there is the need to convert between the linear motion of a piston or diaphragm shaft and the rotary motion of the ball or disc.

The direct connection design (Figure 7.3a) eliminates linkages for the translation of linear to rotary motion. This actuator and connection introduces the least amount of backlash (deadband). The actuator shaft must pivot as the diaphragm or piston strokes. The pivoting design for piston actuators is especially complex because of the effect of the pitch on the O-ring seal.

There are several less expensive connection designs. Unfortunately, these designs all introduce either resolution (stick-slip) or backlash (deadband) errors into the valve response that

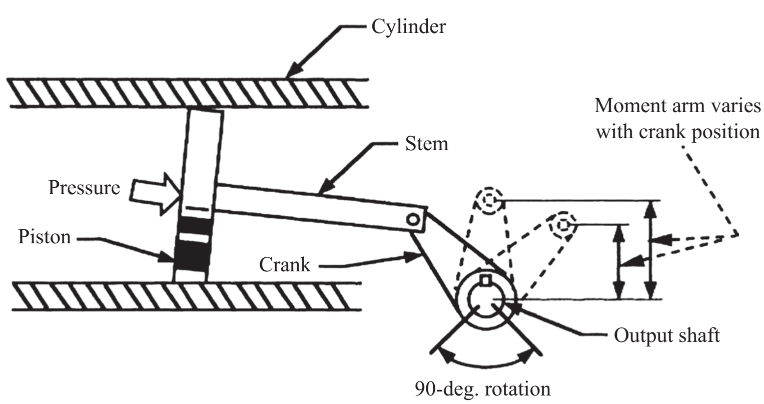


Figure 7.3a. *Direct connection* of piston actuator to rotary valve.

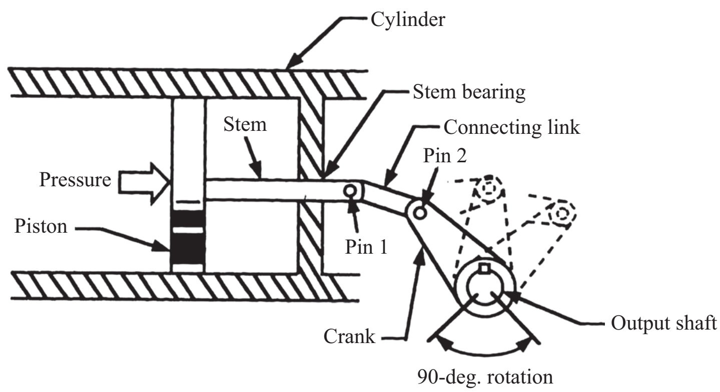


Figure 7.3b. *Link-arm connection* of piston actuator to rotary valve.

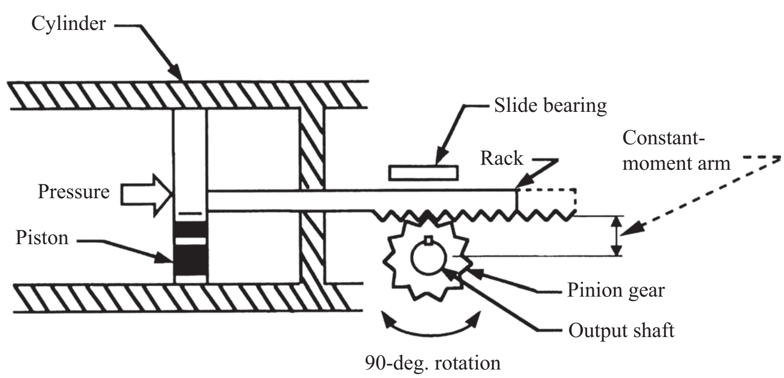


Figure 7.3c. *Rack and pinion connection* of piston actuator to rotary valve.

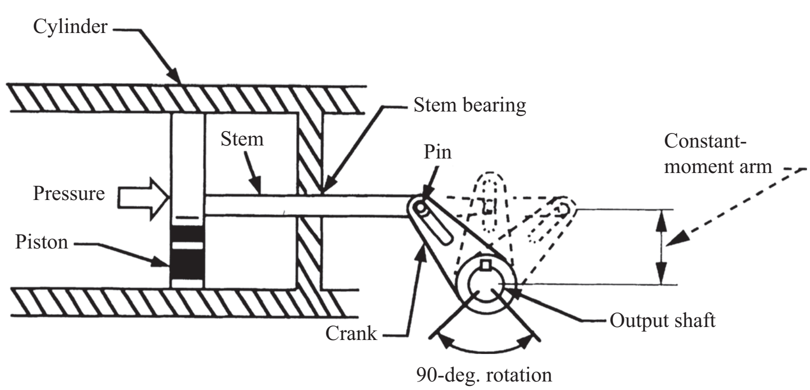


Figure 7.3d. *Scotch-yoke connection* of piston actuator to rotary valve.

increase with wear. The link arm connection design (Figure 7.3b) has backlash from the connections at pin points 1 and 2. The rack and pinion design (Figure 7.3c) has a resolution limit from teeth spacing and backlash from play in the fit between the pinion gear and rack teeth. The slide bearing to hold the rack in contact with gear adds friction. The scotch yoke design (Figure 7.3d) is perhaps the worst from a standpoint of control because the slot adds a lot of backlash.

The problems with the rotary valve connections are aggravated by high friction from stem packing and sealing of internal element (trim) surfaces. Until recently, high temperature and environmental packing designs create excessive stem friction. Additional tightening of the packing beyond factory specifications makes the friction worse. On-off valve (isolation valve) designs (e.g., full port ball or rectangular port plug valves) where the internal element is always rubbing against a seal designed for tight shutoff, create excessive friction. The result can be a twist in the shaft (shaft windup), where the actuator shaft moves but the internal element remains stationary. Eventually, the trim breaks free and jumps to a new position, which is typically beyond the requested position. This twist can also occur in the plug, ball, or disc stem. The windup increases as friction increases, stem or shaft length increases, and stem or shaft diameter decreases.

On-off valves typically use a measurement of actuator shaft rather than trim stem position as feedback to the positioner. The shaft moves but the stem does not. Positioner feedback does not see and diagnostics do not report the fact that the stem did not move. A sensitive flow measurement is needed to determine when the trim has actually moved. In the case of tight shutoff butterfly valves for compressor control and ball valves for phosphorous control, the author has witnessed a difference from backlash of 8 percent between the position indicated by a smart positioner and the actual position of the internal disc or ball. These and other problems can be avoided by a direct connection actuator design that eliminates linkages, a splined shaft to stem connection that eliminates the play of key-lock or pinned connections, a stem cast with the trim that eliminates the play of stem to trim connections, a short and large diameter shaft and stem that reduces shaft windup, and a positioner connection to the stem that provides the best position feedback.

7.4 VFD SYSTEM DESIGN

Modern variable speed drives (VSDs) for pumps are VFDs. The motor is an AC induction motor. The DC motor has a faster response but the initial cost and maintenance cost is much higher due to brushes that need to be periodically replaced. DC motors are used for servo mechanism control where a smaller size and faster speed of response is needed.

7.4.1 PULSE WIDTH MODULATION

Pulse width modulation (PWM) is predominantly used today to vary the frequency of the voltage or current to the pump motor. PWM introduces lower noise, has a higher input power factor, and has better low speed performance than older drive technologies. PWM also offers better rangeability from less cogging (torque pulsation) at low speeds. A VFD inverter converts an AC line voltage to a variable frequency voltage for an AC induction motor. Industrial motors have 3 phases staggered to provide a smoother output. The AC line voltage is rectified and filtered to create a DC voltage. The DC voltage is then inverted to a variable frequency voltage whose frequency is proportional to the drive input signal by PWM. The square wave output is finally filtered to create a sine wave. The rounding from the peak is more accurate for finer pulse widths. Insulated gate bipolar transistors (IGBT) are used to create the series of pulses of varying width. The dominant carrier frequency from PWM is proportional to the drive input signal. The resulting waveform from PWM is closer to a sine wave.

PWM inverters do not have the harmonics and sharp spikes seen in older drives that can damage bearings and create electrical noise in instrument signals. For example, the older 6-step voltage older drive technology, while inexpensive, has a number of undesirable characteristics. The motor can be pushed to its breakdown point. Shorts can cause an infinite current spike. The inverter puts out the same voltage and current at half load as at full load, which reduces efficiency. The waveform has wide and fast current variations that can damage the inverter. The inverter has numerous harmonics that increase motor losses, heat generation, and electromagnetic interference. These inverters have an output choke to prevent damage to motor insulation but the input choke was optional and often missing or insufficient. Eventually, the noise in instrument signals became bad enough that chokes were offered to meet the International Electrochemical Commission (IEC) standards. Alternatively, isolation transformers were located close to the inverter with the power cable between the inverter and transformer in hard pipe conduit to minimize the noise from this section of cable. While the PWM drive has less harmonics and spikes, the rapid rise time of the pulses for precise speed control is still a source of noise and potential damage to bearings and cables.

7.4.2 CABLE PROBLEMS

Belden Inc. has studied the radiated noise from cables between the VFD and the motor. Unshielded VFD cables can radiate 80 V noise to unshielded communication cables and 10 V noise to shielded instrument cables. The radiated noise from foil tape shielded VFD cables is also excessive. A foil braided shield and armored cable performs much better. Still a spacing of at least one foot is recommended between shielded VFD and shielded instrumentation cables. The cables should never cross. As a best practice, use separate trays to isolate VFD and instrumentation cables to avoid mistakes during plant expansions and instrumentation system upgrades.

All VFD systems have reflected waves from an impedance mismatch between the VFD and motor. The amplitude of the waves depends on the voltage magnitude and rise time from the PWM drive, the distance between the VFD and motor, and the impedance mismatch. If a reflected wave gets in-phase with the radiated wave, the voltage can double and the PVC jacketed VFD cables can be damaged. XPLE jacketed VFD cables that are capable of withstanding a high voltage impulse are recommended.

7.4.3 BEARING PROBLEMS

The electric discharge machining (EDM) effect from inverter drives can cause arcing across the lubrication gap of the bearing “almost like a series of little lightning strikes”. The strikes damage the bearing surface and deteriorate the lubricant. The damage is seen as pit marks for variable speed operation and fluting for constant speed operation. The localized high temperatures from the strikes cause reactions of oil additives and burning or charring of the oil. Pitting, fluting, and poor lubrication cause an increase in noise and vibration. Eventually the bearing fails. Mechanical solutions insulate the bearing or provide a path to ground. Ceramic coatings and balls or rollers are used for insulation. However, the insulation forces the currents to go elsewhere and possibly cause damage, such as erosion of pump impellers in the prime mover.

7.4.4 SPEED SLIP

In an AC induction motor, the rotor and hence pump shaft speed lags behind the speed of the rotating electrical field of the stator because a difference in speed is needed to provide the rotor current and consequently the torque to balance any motor losses and the load torque from pump operation. This difference in speed between the stator field and the rotor of the motor is called slip. There is a dynamic slip for large changes in the pump load (e.g., static head) or desired flow rate (speed signal). There is also a steady state slip for operation at a particular load and speed.

It is important to note that VFD speed slip is not the same as valve stroke slip. In speed slip, the speed still responds smoothly to a change in drive signal. At low speed the loss in pump efficiency and an increase in slip cause a dip in flow. Slip affects the minimum controllable speed and hence the VFD rangeability, particularly for high static heads.

In a synchronous motor, the rotor is designed to inherently eliminate slip so the rotor speed is at the synchronous speed of the stator. Synchronous motors are significantly more expensive and complicated and are used only where inherent fast and precise speed regulation is needed. Synchronous motors have been used for ratio control of reactants or additives where small transients or offsets in the speed could cause a significant variation in the product concentration.

7.4.5 MOTOR REQUIREMENTS

Since the inverter waveform is not purely sinusoidal, it is important to select motors that are designed for PWM. These *inverter duty* motors have windings with a higher temperature rating (class F). Another option that facilitates operation at lower speeds to achieve the maximum rangeability offered by the PWM drive is a higher service factor (e.g., 1.15).

To help prevent motor overheating at low speeds, larger frame sizes and line powered ventilation fans are used. In the process industry, totally enclosed fan cooled (TEFC) motors are used to provide protection from chemicals and ventilation by a fan that is run off the same power line as the motor. The fan speed decreases as the motor speed decreases. To reduce the problem from motor overheating at low speeds, an AC line power constant speed ventilation fan and a larger frame size to provide more ventilation space can be specified. Alternately, a separate booster fan can be supplied. For very large motors (e.g., 1,000 HP), totally enclosed water cooled (TEWC) motors are used to deal with the extra heat generation. For low static head pump applications, the overheating at low speeds is not a problem because the torque load decreases with flow.

7.4.6 DRIVE CONTROLS

Open loop voltage (volts/hertz) control has the simplest algorithm but is susceptible to varying degrees of slip. Most of the drives provided for pump control use this strategy in which the rate of change of flux, and hence speed, is taken as proportional to voltage. At low speeds the motor losses are larger making the difference between the computed and actual speed (slip) much larger. Some drives make a correction to the voltage to account for estimated motor losses. Ultimately these drives depend upon the DCS to correct for dynamic slip through proportional action and to correct for steady state slip through integral action in process controller(s). The rangeability is normally 40:1 for low static head applications with 0.5 percent speed regulation.

Closed loop slip control has a speed loop cascaded to a torque loop. Speed (tachometer) and torque feedback are shown to be from sensors. The torque feedback may be calculated from a current sensor. A DCS process controller output is the speed set point for the speed controller whose output is the set point to a torque controller. The speed control should stay in the field. Speed control in the DCS is too slow and coordination with the torque control compromised. The speed controller plays a role similar to the digital valve controller and the torque controller serves a similar purpose as the relay controller. The VFD supplier should have addressed the dynamics of speed to torque cascade control so that the VFD response is smooth. The rangeability is normally 80:1 for low static head applications with 0.1 percent speed regulation.

For liquid flow and liquid pressure control, the process response is so fast that slip can be corrected by the flow or pressure controller in the DCS. Inferential speed or tachometer feedback control in the DCS may require the DCS loop to be detuned because of the cascade rule and thus do more harm than good. The volts/hertz control strategy works well. For gas pressure and gas flow, there may be some benefit from closed loop speed control in the drive particularly on large volumes. For concentration, vessel pH, level, and temperature control where there is no secondary flow loop, the benefit of closed loop speed control is much more obvious because these primary loops are so slow.

7.5 DYNAMIC RESPONSE

The dynamic response of control valves and VFD with PWM is complicated by a considerable rate limit that limits the rate of change of stroke or speed (%/sec). For control valves with pneumatic actuators and VFD with PWM, the largest contributor to dead time is deadband and resolution. The extent of rate limiting and the dead time depends upon the size of the change in signal to the valve or VFD for each execution of the PID controller. The result is a time constant and dead time that depends upon the size of the disturbance and setpoint change and the tuning of the PID. For control valves there is an additional small time constant associated with the digital positioner and actuator response.

7.5.1 CONTROL VALVE RESPONSE

If the change in signal is greater than the final control element deadband, resolution limit, or threshold sensitivity, the control valve response is a pre-stroke dead time plus a rate limited exponential. Equation 7.1a shows the effective time constant of a rate limited exponential response is the time for a full scale change in stroke or speed (T_v) multiplied by the fractional change in controller output plus an inherent time constant (τ_i). The inherent time constant can be generalized to be a ratio (R_v) of the inherent time constant to the time required to make a full scale change in stroke or speed. The ratio ranges 0.02 to 0.1 for diaphragm actuators and

0.2 to 1.0 for piston actuators. Consequently, for response time of a well-designed throttling valve for small step changes in signal (e.g., 0.5 percent) is significantly faster for a diaphragm actuator. For medium step changes (e.g., 5 percent), the response times are similar. For large step changes (e.g., 25 percent), the piston slewing rate is faster because of the smaller actuator volume from the use of higher air pressure. New high pressure diaphragm actuators eliminate much of the need to go to pistons.

$$\tau_{v1} = T_v * \frac{\Delta\%CO}{100\%} + \tau_i \tag{7.1a}$$

$$\tau_i = T_v * R_v \tag{7.1b}$$

$$\tau_{v1} = T_v * \left(\frac{\Delta\%CO}{100\%} + R_v\right) \tag{7.1c}$$

The pre-stroke dead time (Equation 7.2a) and stroking time (Equation 7.2b) are proportional to size of the actuator as determined by the (X_v) factor and (Y_v) factor in Table 7.1a for diaphragm and Table 7.1b for piston actuators 7.2, respectively. The dead time and stroking time are inversely proportional to the effective flow coefficient defined by Equation 7.2c from the individual flow coefficients in Table 7.2 for a positioner and volume booster in series. The tables are typical values from one manufacturer. Actual values should be sought from the valve supplier.

$$\theta_{vz} = \frac{X_v}{C_v} \tag{7.2a}$$

$$T_v = \frac{Y_v}{C_v} \tag{7.2b}$$

$$C_v = \sqrt{\frac{C_{v1}^2 * C_{v2}^2}{C_{v1}^2 + C_{v2}^2}} \tag{7.2c}$$

where

$\Delta\%CO$ = change in PID controller output to correct for load disturbance (%)
C_v = flow coefficient for solenoid or booster plus positioner (scfm per psi$^{0.5}$)
C_{v1} = flow coefficient for positioner (scfm per psi$^{0.5}$)
C_{v2} = flow coefficient for solenoid or booster (scfm per psi$^{0.5}$)
R_v = ratio of inherent time constant to time for full scale response (dimensionless)
T_v = time for a full scale response of control valve or VFD (sec)
X_v = actuator factor for pre-stroke dead time (seconds * scfm per psi$^{0.5}$)
Y_v = actuator factor for stroking time (seconds * scfm per psi$^{0.5}$)
τ_i = valve inherent constant (sec)
τ_{v1} = valve or drive time constant 1 (sec)

The pre-stroke dead time and stroking time factors for actuator exhaust are greater than the factors for actuator fill in Table 7.1a because a greater exhaust rate is inexpensive and useful on a diaphragm actuator to go to an air fail position faster.

Quick release valves that provide a large flow coefficient when triggered have been removed from Table 2.1 because of the discontinuity created in the valve response.

Table 7.1a. Diaphragm actuator pre-stroke dead time and stroking time factors

Actuator (sq. in.)	Travel (inches)	Pressure (psig)	Spring (lb/in)	Fill factors X_v	Fill factors Y_v	Exhaust factors X_v	Exhaust factors Y_v
66	21/8	4.5–16	275	0.015	0.338	0.225	0.610
100	31/2	6–22	335	0.031	0.861	0.404	1.446
215	41/8	6–21	610	0.105	2.276	1.200	3.902
46	3/4	3–15	735	0.012	0.190	0.045	0.256
46	3/4	6–30	1,470	0.016	0.226	0.031	0.290
69	3/4	3–15	1,100	0.020	0.297	0.071	0.401
69	3/4	6–30	2,210	0.028	0.355	0.048	0.457
105	3/4	3–15	1,670	0.033	0.466	0.115	0.630
105	3/4	6–30	3,320	0.046	0.574	0.078	0.727
156	3/4	3–15	2,500	0.046	0.676	0.161	0.913
156	3/4	6–30	5,000	0.065	0.811	0.110	1.046
220	2	3–15	1,260	0.074	2.004	0.500	2.810
220	2	6–30	2,520	0.104	2.243	0.390	3.038
310	2	3–15	1,650	0.103	2.790	0.569	3.724
310	2	6–30	3,100	0.143	3.379	0.388	4.265
450	2	6–26	4,500	0.323	4.386	1.260	6.552
450	2	6–26	4,500	0.380	4.586	1.353	6.870

Table 7.1b. Piston actuator pre-stroke dead time and stroking time factors

Actuator (sq. in.)	Travel (inches)	Pressure (psig)	Spring (lb/in)	Fill factors X_v	Fill factors Y_v	Exhaust factors X_v	Exhaust factors Y_v
17	3/4	60	–	0.085	0.050	0.024	0.050
28	3/4	60	–	0.165	0.086	0.035	0.086
56	3/4	60	–	0.296	0.169	0.050	0.169
89	2	60	–	0.715	0.719	0.196	0.719
131	2	60	–	0.995	1.060	0.272	1.060
222	2	60	–	1.730	1.800	0.738	1.800
17	4	60	–	0.020	0.278	0.024	0.278
28	4	60	–	0.051	0.460	0.035	0.460
56	4	60	–	0.099	0.901	0.050	0.901
89	4	60	–	0.181	1.453	0.196	1.453
131	4	60	–	0.227	2.144	0.272	2.144
222	4	60	–	0.603	3.600	0.738	3.600

Table 7.2. Flow coefficients of accessories

Accessory type	Connection sizes (inches)	Supply C_v	Exhaust C_v
Positioner	–	0.37	0.31
I/P Transducer	–	0.39	0.36
Volume booster	3/8 and 3/8 ports	3.74	2.29
Volume booster	3/8 and 1/2 ports	3.74	2.52
Volume booster	1/2 and 3/8 ports	5.32	2.30
Volume booster	1/2 and 1/2 ports	5.32	2.53

7.5.2 VFD RESPONSE

For a VFD the time constant simplifies to Equation 7.3a because the ratio (R_v) is essentially zero. The time for a full scale speed change as detailed by Equation 7.2b is 100 percent divided by the speed setpoint maximum rate of change $(\Delta\%SP / \Delta t)_{\max}$ determined in the VFD setup. The substitution of Equation 7.3b into 7.3a yields Equation 7.3c revealing the VFD time constant is simply the percent change in controller output divided by the speed setpoint rate limit.

$$\tau_{v1} = T_v * \frac{\Delta\%CO}{100\%} \tag{7.3a}$$

$$T_v = \frac{100\%}{(\Delta\%SP / \Delta t)_{\max}} \tag{7.3b}$$

$$\tau_{v1} = \frac{\Delta\%CO}{(\Delta\%SP / \Delta t)_{\max}} \tag{7.3c}$$

7.5.3 DEAD TIME APPROXIMATION

The total dead time for a valve or VFD as expressed by Equation 7.4 is the summation of any preemptive dead time (θ_{vz}) plus the fraction Y of the small time constants converted to an equivalent dead time and the dead time associated with deadband and threshold sensitivity or resolution limit (θ_{vx}). The Y faction increases as the ratio of the small time constant to largest time constant in the loop decreases as defined by Equations 4.1a through 4.1c in Chapter 4. The VFD has no preemptive dead time in the speed response ($\theta_{vz} = 0$).

The first time constant (τ_{v1}) is defined by Equation 7.1c for control valves and by Equation 7. 3c for VFD. There is also a second time constant (τ_{v2}) associated with the valve positioner (e.g., digital valve controller) and VFD controls (e.g., speed, torque, and flux vector). Normally this second time constant is so small (e.g., 0.1 second) that it only comes into play for small changes in stroke or speed.

$$\theta_v = \theta_{vz} + Y * \tau_{v1} + Y * \tau_{v1} + \theta_{vx} \tag{7.4}$$

where

$\Delta\%CO$ = change in PID controller output to correct for load disturbance (%)
$\Delta\%SP / \Delta t$ = rate of change of speed setpoint (%/sec)

R_v = ratio of inherent time constant to time for full scale response (dimensionless)
T_v = time for a full scale response of control valve or VFD (sec)
Y = fraction of small time constant converted to dead time (Equation 4.1a)
θ_v = total valve or VFD dead time (sec)
θ_{vz} = valve pre-stroke dead time (sec)
θ_{vx} = valve or VFD delay from deadband and sensitivity or resolution limit (sec)
τ_{v1} = valve or drive time constant 1 (sec)
τ_{v2} = valve or drive time constant 2 (sec)

7.5.4 DEADBAND AND RESOLUTION

Figure 7.4 shows the effect of backlash and stick-slip on the response of a control valve. When the signal reverses direction, the stroke does not change direction until the change in signal exceeds the valve dead band from backlash. While technically valve dead band is defined for a full scale up and down stroke, in actuality deadband occurs whenever and wherever there is a change in signal.

After the valve works through the deadband, there is still no response until an additional change in signal exceeds the resolution or threshold sensitivity limit from stick-slip. Stick-slip is better described by a threshold sensitivity limit where the stroke will match or exceed the additional step change in signal beyond backlash in a single step rather than be limited to a series of fixed sized steps in the valve response. Common practice is to quantify stick-slip as a resolution limit. Stick-slip is most severe near the closed position due to the friction of seating and sealing surfaces. In general a lower leakage specification (tighter shutoff) translates to larger friction and consequently greater stick-slip near the closed position. Figure 7.4 shows the steps in the valve stroke from seating and sealing friction as the valve reopens.

The term, “hysteresis”, is frequently mistakenly used for deadband. In hysteresis, the output immediately responds to a change in direction of input, but the path bows due to lost energy.

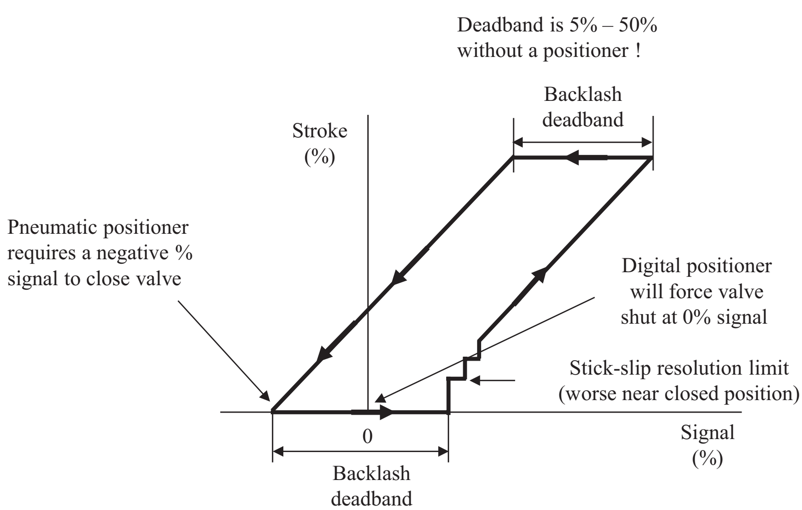


Figure 7.4. Backlash dead band and stick-slip resolution limit.

Figures in Appendix A show hysteresis by itself and combined with deadband. Unlike deadband, there is no delay in the response from hysteresis and the slope (the amount of change in the stroke for a change in signal) is not constant. A small but negligible amount of hysteresis can exist in an actuator's response from the elasticity of the diaphragm. Pinch valves may exhibit a significant amount of hysteresis from the elasticity of the boot. In general, deadband is the more dominant effect.

The official definition of deadband in any device is consistent with Figure 7.4 where the effect is only seen on a reversal of direction, the term deadband is used for an adjustable parameter in the VFD and PID that is often really a dead zone. The distinction is that for a dead zone there is no response regardless of direction for a change in signal that is within the dead zone centered on the setpoint.

As previously mentioned, the stick-slip including shaft windup is an order of magnitude larger (e.g. 4 percent) from sealing and seating friction for tight shutoff isolation valves being used as throttling valves. The type of piston actuators supplied with these on-off valves also introduce a large amount of backlash (e.g., 8 percent) because of links and pinned shaft to stem connections. Rack and pinion actuators introduce a significant resolution limit (e.g., 5 percent) from the rack teeth. Wear of the teeth can increase the slip.

Well-designed control valves for throttling service do not use on-off types of actuators or connections. Diaphragm actuators and double acting piston actuators are used on these valves with direct connections and digital positioners. The principle types of well-designed valves are linear motion sliding stem (globe) and rotary motion v-ball and contoured butterfly valves. A sliding stem valve has the best resolution (e.g., 0.1 percent) and deadband (e.g., 0.2 percent). A rotary motion v-ball and contoured butterfly valves with direct splined shaft connections have the next best resolution limit (e.g., 0.2 percent) and dead band (e.g., 0.4 percent). A diaphragm actuator with a threshold sensitivity of 0.1 percent enables these valves to achieve their best precision. In comparison, the traditional design of double acting piston actuator with a direct connection introduces a threshold sensitivity of about 1 percent. The performance cited is for a digital positioner and an actuator with more than adequate thrust or torque. If a marginally sized actuator is used or a pneumatic positioner is used, the precision deteriorates dramatically. The use of high volume spool instead of a high gain relay type of positioner can introduce a threshold sensitivity of 5 percent or more.

While a VFD has no stick-slip or threshold sensitivity limit, a resolution limit is appreciable for an 8 bit analog to digital (A/D) speed input card and a dead band on speed setpoint changes may be entered by the supplier or user in the VFD setup to reduce changes in speed. This dead band may be beneficial in preventing the VFD chasing noise, but often the dead band setting is indiscriminately set too large because there is no understanding of the dead time created and the consequential detrimental effect.

The additional delay (θ_{vx}) from the valve or VFD deadband ($\%DB_v$) and resolution or threshold sensitivity limit ($\%S_v$) increases as the controller is detuned expressed as a fraction (K_x) of the PID gain for a quarter amplitude response. The fraction is 0.75 ($K_x = 0.75$) for the most aggressive Lambda tuning for an integrating process gained by a Lambda equal to the dead time. Equation 7.5a and first half of Equation 7.5b were developed in the Theory section of Chapter 2 of *Advanced Control Unleashed* for the effect of dead band on the rate of change of the controller output, open loop gain, and total loop dead time. The percent change in actual valve position ($\Delta\%AVP$) correction required for a load disturbance is determined from trend charts.

$$\theta_{vx} = \frac{\%DB_v + \%S_v}{\Delta\%CO / \Delta t} \tag{7.5a}$$

$$\Delta\%CO / \Delta t = K_c * \Delta\%PV / \Delta t \tag{7.5b}$$

$$\Delta\%CO / \Delta t = \frac{K_x}{(K_i * \theta_o)} * K_i * \Delta\%AVP \tag{7.5c}$$

$$\Delta\%AVP = \Delta\%CO - \%DB / 2 \tag{7.5d}$$

$$\Delta\%CO / \Delta t = \frac{K_x}{\theta_o} * (\Delta\%CO - \%DB / 2) \tag{7.5e}$$

$$\theta_{vx} = \left[\frac{\%DB_v}{K_x * (\Delta\%CO - \%DB_v / 2)} + \frac{\%S_v}{K_x * \Delta\%CO} \right] * \theta_o \tag{7.5f}$$

where

K_x = fraction of maximum gain for quarter amplitude load response (dimensionless)
K_i = open loop integrating process gain (1/sec)
$\%AVP$ = step change in actual valve position (%)
$\Delta\%CO$ = change in PID controller output to correct for load disturbance (%)
$\Delta\%CO / \Delta t$ = rate of change of PID controller output (%/sec)
$\%DB_v$ = control valve or VFD deadband (%)
$\%S_v$ = control valve or VFD resolution or sensitivity limit (%)
θ_{vx} = valve or VFD delay from deadband and sensitivity or resolution limit (sec)

A compensator can be developed for valve deadband. When the PID output changes direction by more than the noise band, the PID output is incremented or decremented by a bias for a positive or negative change, respectively. The magnitude of the bias is the deadband minus the noise band. A bias magnitude larger than this value will cause valve slip, which is generally worse than deadband. Since the actual valve deadband varies with time and position, deadband compensation must be conservative and incomplete.

7.5.5 WHEN IS A VALVE OR VFD TOO SLOW?

We can reuse Equation 6.2 to estimate the amount of valve or VFD delay that causes no appreciable need to retune the controller for self-regulating processes.

$$\theta_v < 0.5 * (\lambda + \theta_o) - \theta_o \tag{7.6a}$$

Similarly, we can reuse Equation 6.3 to estimate the amount of valve and VFD delay that causes no appreciable need to retune the controller for integrating processes.

$$\theta_v < 0.25 * (\lambda + \theta_o) - \theta_o \tag{7.6b}$$

Since a major portion of the valve dead time depends upon the size of the controller output change ($\%\Delta CO$), the smallest change for the effect of dead band and resolution limits and the

largest change for the effect of rate limiting should be used in determining if the valve or VFD dead time should be reduced.
where

λ = Lambda (closed loop self-regulating time constant or integrating arrest time) (sec)
θ_v = total valve or VFD dead time (sec)
θ_i = implied loop dead time from PID tuning (sec)
θ_o = original loop dead time (sec)

A less exact way of looking at the effect of rate limiting is to compare the maximum correction time for a load disturbance with the reset time of the PID. The correction time is the change in output required divided by the stroke or speed rate limit. The correction time should be less than the reset time. For surge control where the surge valve needs to go fully open, this requirement simplifies to the stroking time being less than the reset time. Boosters are added on surge valve positioner outputs to satisfy this requirement.

7.5.6 LIMIT CYCLES

Resolution limits and stick-slip cause continual hunting of a controller with integral action on any process because the measured process variable never exactly matches the set point. Integral action will ramp the controller output for an offset of the process variable from the set point. Figures in the test results show the limit cycles in the process variable and controller output for flow and level loops.

For a fast process variable, such as flow, the controller output ramps back and forth (sawtooth wave), and the process variable steps up and down (square wave) from stick-slip. If there are no disturbances, the result is a perpetual constant amplitude oscillation called a limit cycle. The limit cycle amplitude is the stick-slip multiplied by the open loop gain (Equation 7.7a). The open loop self-regulating process gain is the change in PV in percent of scale divided by the change in controller output in percent. For a fast self-regulating process (e.g., flow) the limit cycle period is proportional to the integral time and is inversely related to the controller gain (Equation 7.7b). Detuning the controller (decreasing the controller gain and increasing the controller integral time) increases the limit cycle period. The limit cycle amplitude is independent of tuning.

$$A_o = S_v * K_o \tag{7.7a}$$

$$T_o = 4 * T_i * [1/(K_o * K_c) - 1] \tag{7.7b}$$

where

A_o = amplitude of limit cycle (%)
K_c =controller gain (dimensionless)
K_o =open loop self-regulating process gain (dimensionless)
S_v =valve stick-slip (%)
T_i = controller integral time (sec/repeat)
T_o = period of limit cycle (sec)

If there are two or more integrators in the process and control system that affect the final element, deadband will cause a limit cycle. For an integrating process, such as level, the controller output is a sinusoidal oscillation and the process variable ramps with some rounding of the peaks (smoothed sawtooth) from backlash. The flow trend chart shows a clipped oscillation. If there are no continual disturbances, the result is a limit cycle. The limit cycle amplitude is the deadband divided by the controller gain (Equation 7.8a). For a true integrating process (e.g., level) the limit cycle period is proportional to integral time and is inversely related to controller gain (Equation 7.8b). Detuning the controller (decreasing the controller gain and increasing the controller integral time) increases both limit cycle amplitude and period.

$$A_o = B_v \ / \ K_c \tag{7.8a}$$

$$T_o = 5 * T_i * [1 + 2 / (K_c^{0.5})] \tag{7.8b}$$

where

A_o = amplitude of limit cycle (%)
B_v = valve backlash (%)
K_c =controller gain (dimensionless)
T_i = controller integral time (sec/repeat)
T_o = period of limit cycle (sec)

The limit cycle in process variables for a valve with minimal stick-slip and backlash (e.g., 0.2 percent) is often not noticeable because the oscillation is filtered (smoothed or washed out) by volumes. Sensor and signal filter time constants have a similar effect.

In the literature, the source of limit cycles is usually stated to originate from stick-slip. When limit cycles from backlash are addressed, the case is stated to be a level loop. The use of integral action is not stated as a necessary condition. The rules on the source of limit cycles that lead to a deeper understanding are:

1. A limit cycle from stick-slip or resolution limits develops when there is at least one integrator in the process or control system.
2. A limit cycle from backlash or deadband develops when there are at least two integrators in the process or control system.

The integrators can be from an integrating process response (e.g., level and batch temperature and composition) or from integral action in the feedback controllers in DCS or in the positioner. Two or more controllers with integral action in the quadruple cascade control system of continuous reactor temperature to jacket temperature to coolant flow to valve position control, will result in a limit cycle from backlash even if the temperature has no integrating process response.

The limit cycle from stick-slip for self-regulating processes and from backlash for all processes can be stopped until the next disturbance by turning off the integral action when the controlled variable is close to set point. The offset from set point that triggers the suspension of integral action should be just greater than the limit cycle amplitude. This integral deadband feature exists in DCS controllers and digital positioners. The enhanced PID developed for wireless and the fast external feedback of actual stroke or speed can also stop limit cycles without the need to set an offset.

7.6 INSTALLED FLOW CHARACTERISTICS AND RANGEABILITY

In linear systems the dynamics (gain, dead time, and time constant) are constant. We have already seen that the dead time and time constant for valves and VFD is variable and difficult to quantify. We will focus here on the valve gain and VFD gain, which is what most practitioners consider when judging the linearity of a control system. For small changes in signal the valve and VFD gain is the slope at the operating point on a plot of flow versus signal taking into account the effect of the pump curve, static head, and the system frictional loss. The frictional loss is the pressure drop from the resistance to flow of everything in the flow path (e.g., piping, manual valves, and process equipment) except for the control valve. Some publications include the control valve pressure drop in the system loss. The system frictional loss is proportional to the flow squared. This plot termed the installed flow characteristic takes into account the effect of nonideal operating conditions including changes with time and of mechanical and piping design. Software programs from the supplier can compute these installed flow characteristics but often the time is not taken to get the input information needed. The main reason for not making the extra effort is the lack of recognition that the open loop gain is proportional to the slope with implications not only as to the increase nonlinearity but also the increase in limit cycle amplitude and decrease in rangeability from amplification of stick-slip.

What is seen in the literature is the inherent flow characteristic. For a VFD, the inherent flow characteristic assumes that there is no static head so that the flow changes linearly with speed. For a control valve, the inherent characteristic assumes the pressure drop across the control valve is constant (frictional loss in rest of system is zero). There are many types of valve inherent flow characteristics (e.g., trim types for sliding stem valves). The inherent flow characteristic is the starting point for getting the installed flow characteristic and rangeability of the control valve.

7.6.1 VALVE FLOW CHARACTERISTICS

Figure 7.5 shows the theoretical inherent flow characteristic for the three major types of trim. The distinguishing aspect of these flow characteristics for process control is the slope (change in flow/change in stroke), because this slope is the valve gain. The plot is commonly labeled as percent flow versus percent stroke. In reality it is really the flow coefficient expressed as a percent of the maximum flow coefficient versus percent stroke. The inherent flow characteristic does not take into account various application effects (e.g., pipe reducers, pump curve, system frictional loss, high viscosity, and flashing).

The inherent flow characteristic for quick opening trim has a constant slope from 0 to 40 percent. At 40 percent stroke about 70 percent of the flow capacity is attained. Above the 40 percent position, the slope drastically decreases. Above the 60 percent position, there is very little additional flow. Quick open characteristics are used for pressure relief and surge control where a rapid change in flow is needed. Very few applications benefit from a quick open characteristic.

The inherent flow characteristic for linear trim has a constant slope throughout its stroke range. The fractional flow coefficient is the fractional stroke (Equation 7.9). The linear flow characteristic is useful for vessel pH and header pressure control systems where the valve pressure drop is relatively constant. The next section on installed flow valve characteristics shows how a variable pressure drop causes a linear inherent flow characteristic to distort to a quick opening installed characteristic.

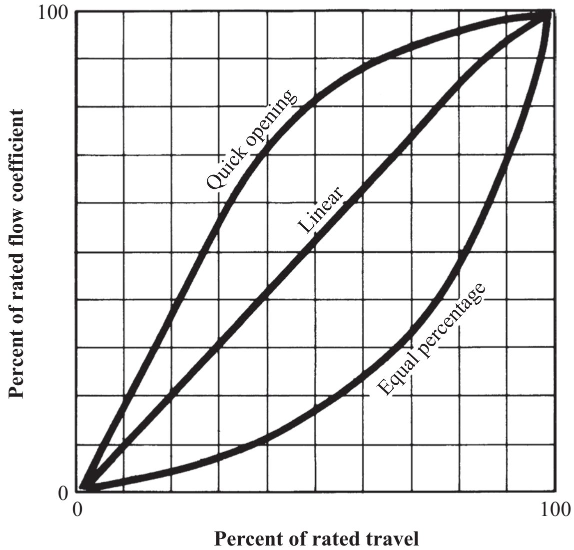


Figure 7.5. Theoretical inherent valve characteristics.

The inherent characteristic for equal percentage trim has a slope that is proportional to flow. The change in flow is equal to the flow on a percentage basis. The slope is almost flat near the closed position and exponentially increases with stroke. The fractional flow coefficient is an exponential function of the fractional stroke (Equation 7.10). The range over which the equal percentage is valid is normally taken as 50 ($R = 50$). Equal percentage trim is the most popular choice because of many practical advantages.

The equal percentage trim minimizes the effect of stick-slip near the closed position where seating and sealing friction is greatest. This decrease in stick-slip increases the practical rangeability of the control valve.

The control valve pressure drop at rated capacity normally only needs to be about 20 percent of the system friction loss for equal percentage trim, which reduces the pump head and hence the pump size and energy requirement. The equal percentage characteristic is more forgiving of under-sizing because of higher slope near the wide open position and of over-sizing because of the lower slope near the closed position.

The next section on installed valve characteristics shows how a variable pressure drop causes an equal percentage inherent characteristic to approach a linear installed characteristic. A constant valve gain is desirable for flow and liquid pressure loops and for temperature and pH loops on well mixed vessels.

For cases where the valve pressure drop is constant, the equal percentage trim compensates for a process gain that is inversely proportional to feed flow. Examples are pipeline concentration control by the direct manipulation of reagent and additive flows, and for heat exchanger control by the direct throttling of coolant valves. If a secondary flow loop is installed, there is no compensation of the process gain by the valve gain. For cascade control systems, where the secondary loop throttling the valve is a flow loop, the primary loop (e.g., pH, concentration, or temperature) does not see the nonlinearity of the installed characteristic since it is manipulating flow rather than valve position. While cascade control is generally beneficial, there are cases where the cascade rule that requires the secondary loop to be 5 × faster than the primary loop may necessitate detuning a fast primary loop to slow it down. For inline systems, if feedforward or ratio control is not needed, the faster response and nonlinear compensation from direct manipulation of a sliding stem (globe) valve with a digital positioner may provide better rejection of fast disturbances.

For linear trim:

$$C_x = \frac{X_v}{X_{v\max}} \tag{7.9}$$

For equal percentage trim:

$$C_x = R^{\left[\frac{X_v}{X_{v\max}} - 1\right]} \tag{7.10}$$

where

C_x = flow coefficient expressed as a fraction of maximum (dimensionless)
R = range of the equal percentage characteristic (e.g., 50)
X_v = actual valve stroke (%)
$X_{v\max}$ = maximum valve stroke (%)

The equal percentage inherent flow characteristics offers these practical advantages:

- Minimum effect of stick-slip near the closed position
- Maximum rangeability
- Minimum pressure drop and hence pump size and energy use
- Maximum forgiveness of under-sizing and over-sizing
- Maximum linearity for valve pressure drop that significantly decreases with flow
- Beneficial process gain compensation for some inline control systems

Everything presented here and in the literature is based on theoretical flow characteristics, pump curves, and system curves. The actual inherent characteristics of control valves deviate from the theoretical ones particularly at the low and high ends of the stroke. Also, many rotary valves do not follow the equal percentage characteristic. In general, the inherent characteristics of rotary valves fall somewhere between the theoretical equal percentage and linear characteristics.

The installed characteristic of control valves in Figures 7.7a and 7.7b takes into account the effect of a variable valve pressure drop. The pressure drop available to a control valve is seen in Figure 7.6 as the difference between pump head (upstream pressure) per pump curve and the sum of the static head and system frictional loss (downstream pressure) per system curve. The variation in pressure drop due to a frictional loss in the piping system makes an equal percentage trim much more linear.

Operation within the proper throttle range of a properly selected trim (equal percentage for most applications), valve nonlinearity does not cause a noticeable deterioration in performance because process controllers are usually detuned by a factor of 4 or more. However, to develop a deeper understanding of the consequences of improper design and the potential to do better for difficult and demanding loops, a detailed analysis is offered.

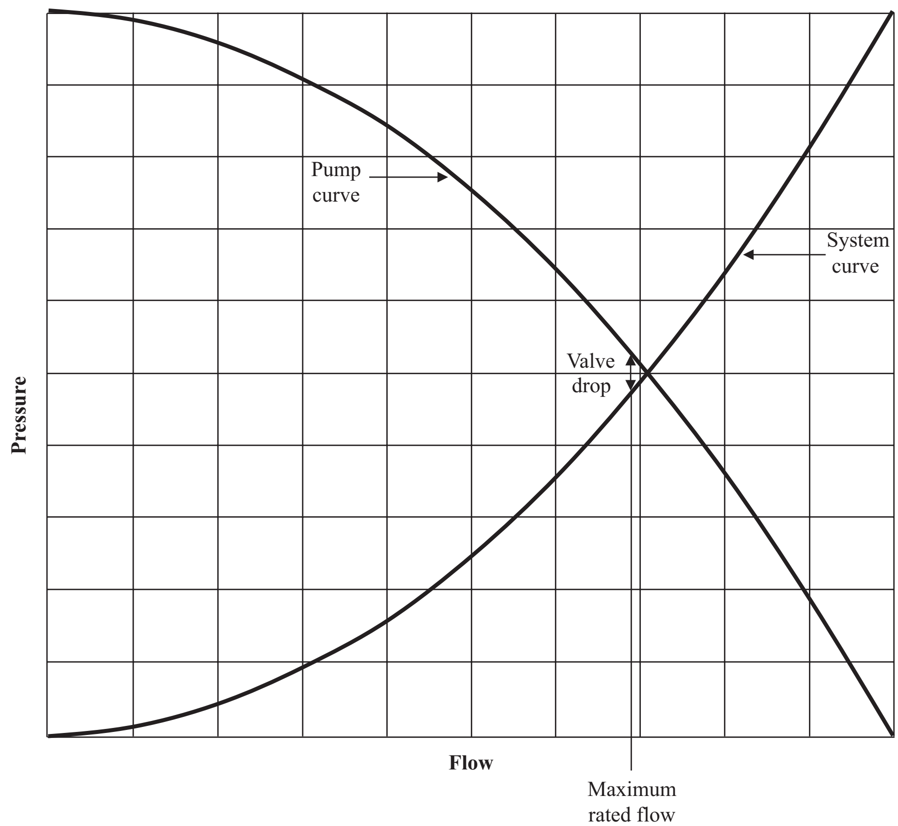


Figure 7.6. Available valve pressure drop.

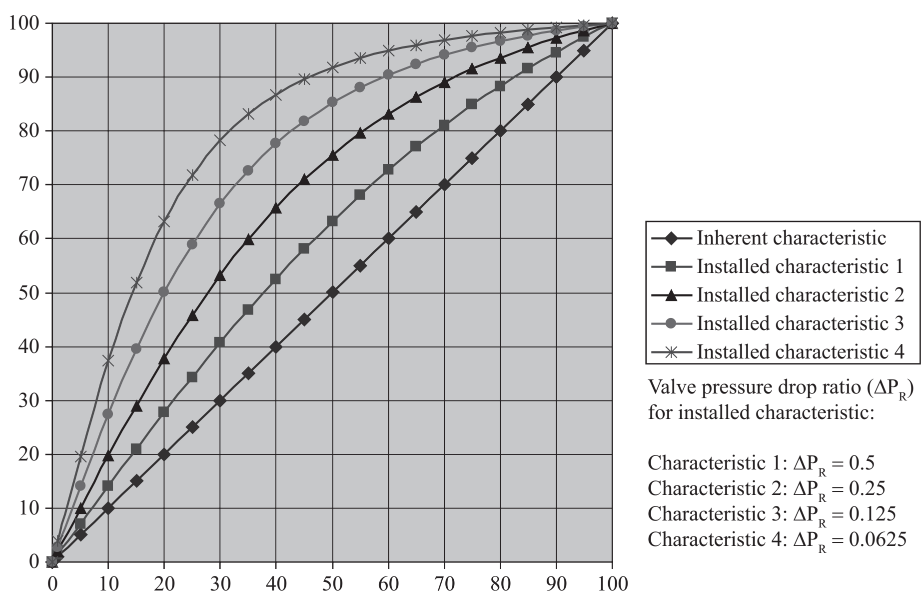


Figure 7.7a. Effect of valve to system pressure drop ratio on installed flow characteristic of *linear trim.*

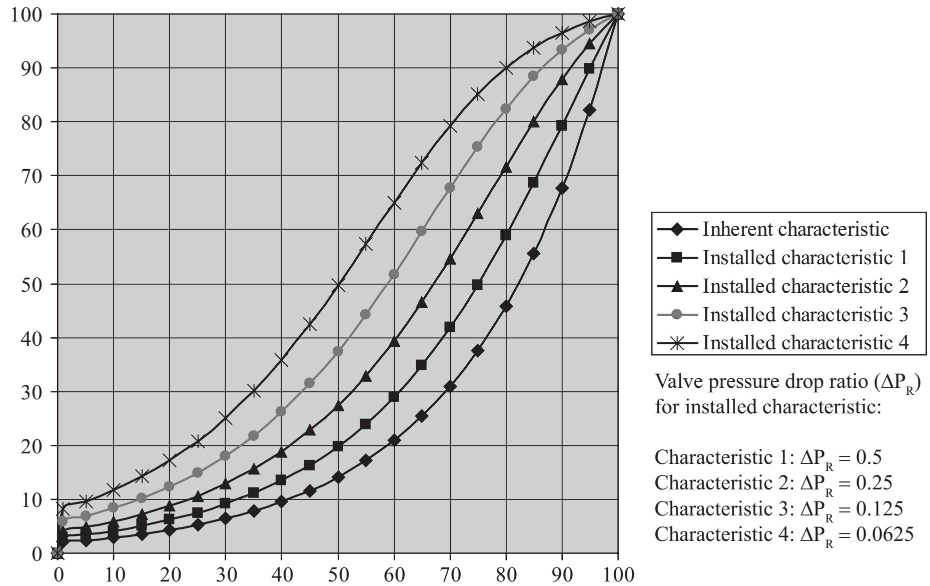


Figure 7.7b. Effect of valve to system pressure drop ratio on installed flow characteristic of *equal percentage trim.*

At the maximum rated flow, a pressure drop that is 10 percent of the system frictional loss may be used to conserve energy. This corresponds to a pressure ratio of 0.1 as defined by Equation 7.11 as the ratio of the valve pressure drop for a fully open valve to the valve pressure drop for a completely closed valve. The maximum *rated* flow corresponds to a 100 percent position. In the valve sizing, the maximum *required* flow may be chosen to be at a position in the more linear throttle range of the control valve (e.g., <75 percent for a sliding stem (globe) valve, < 60° for a ball valve, and <45° for a conventional disc butterfly). Sizing programs that allow you to specify the pump and systems curves and static head, will give the pressure drop and flow at other valve positions.

$$\Delta P_R = \frac{\Delta P_{100\%}}{\Delta P_{0\%}} \tag{7.11}$$

where

ΔP_R = valve pressure drop ratio (dimensionless)
$\Delta P_{0\%}$ = valve pressure drop at completely closed position (kPa)
$\Delta P_{100\%}$ = valve pressure drop at fully open position (kPa)

The installed characteristic (Equation 7.12) is a function of the pressure drop ratio and the fractional flow coefficient (C_x) that depends upon the inherent flow characteristic. A plot of Equation 7.12 for a linear and equal percentage trim shows how the installed characteristic becomes more “quick opening” and linear, respectively, as the pressure drop ratio decreases (Figures 7.6a and 7.6b). For both trims, the flow characteristic bends over more at the top end of the stroke and the flow at the low end of the stroke is higher as the pressure drop ratio decreases.

The result is a decrease in rangeability. Thus, the desire to save energy by decreasing the valve pressure drop at maximum flow must be tempered by the need for rangeability.

$$Q_x = \frac{C_x}{\sqrt{\Delta P_R + (1 - \Delta P_R) * C_x^{\,2}}} \tag{7.12}$$

where

C_x = flow coefficient expressed as a fraction of maximum (dimensionless)
ΔP_R = valve pressure drop ratio (dimensionless)
Q_x = flow expressed as a fraction of the maximum rated flow (dimensionless)

The loss in sensitivity at high flow can be so great (particularly for rotary valves) that the controller output wanders since the controller sees no real change in flow. The controller output limit should be set to prevent the controller output from wandering on the flat part of the curve, but some plants may be reluctant to give up that last 0.5 percent of capacity. In these applications, signal characterization may be a temporary solution. A better solution would be to increase the pump head or decrease the system frictional loss.

The valve flow characteristic is best captured online with a high rangeability low noise flow meter (e.g., magmeter or Coriolis meter) and by a ramp in the valve signal at normal operating conditions. A digital positioner is essential for this test. For small sliding stem (globe) valves, the whole ramp may be done in less than 20 seconds (20 points for a one second update time). The valve gain can be computed online by a simple calculation in the DCS to divide the filtered change in flow by the change in stroke. Too slow of a ramp can make the calculation too noisy by making the true change in flow small relative to the size of the noise. Alternately, an adaptive controller can be used to identify the process gain at various operating points for gain scheduling

7.6.2 VALVE RANGEABILITY

The rangeability of a control valve is the ratio of the maximum to minimum controllable flow. The maximum flow is near the intersection of the system frictional loss curve with the pump curve for maximum speed. The definition of the minimum controllable flow depends upon whether the final element is a valve or VFD and the viewpoint of the supplier, user, or contractor.

Some valve manufacturers define the minimum flow as the point where the inherent flow characteristic exceeds some specified tolerance. Process control consultants tend to specify the minimum flow to be where the valve gain becomes unacceptably small (e.g., valve gain <0.5). The author prefers to base the minimum flow on the valve stick-slip (S_v) near the closed position. In this case the minimum flow is the flow on the installed characteristic at a valve position equal to the stick-slip. The use of stick-slip provides a rangeability that is much larger than what control consultants might suggest but still much smaller than what piping valve manufacturers might state.

The minimum practical factional flow can be estimated by substituting the slip-stick near the closed position into the equations for the linear trim (Equation 7.10) or equal percentage trim (Equation 7.11). The resulting minimum fractional flow is then substituted into the equation for the installed characteristic to get the minimum fractional flow (Equation 7.15a). The rangeability of the control valve is simply the inverse of the minimum controllable fractional flow (Equation 7.15b).

Minimum fractional flow coefficient for a linear trim and stick-slip:

$$C_{x\min} = \frac{S_v}{X_{v\max}} \tag{7.13}$$

Minimum fractional flow coefficient for an equal percentage trim and stick-slip:

$$C_{x\min} = R^{\left[\frac{S_v}{X_{v\max}} - 1\right]} \tag{7.14}$$

Minimum controllable fractional flow for installed characteristic and stick-slip:

$$Q_{x\min} = \frac{C_{x\min}}{\sqrt{\Delta P_R + (1 - \Delta P_R) * C_{x\min}{}^2}} \tag{7.15a}$$

$$R_v = \frac{1}{Q_{x\min}} \tag{7.15b}$$

where

C_x = minimum flow coefficient expressed as a fraction of maximum (dimensionless)
ΔP_R = valve pressure drop ratio (dimensionless)
$Q_{x\min}$ = minimum flow expressed as a fraction of the maximum (dimensionless)
R_v = rangeability of control valve (dimensionless)
R = range of the equal percentage characteristic (e.g., 50)
$X_{v\max}$ = maximum valve stroke (%)
S_v = stick-slip near closed position (%)

7.6.3 VFD FLOW CHARACTERISTICS

In a VFD for liquid flow, the pump characteristic curve shifts with pump speed. Since there is no control valve, there is no valve drop and the flow is at the intersection of the pump curve and the system frictional loss curve (Figure 7.8a).

For zero static head and an idealized pump, motor, and VFD, the change in flow with speed is linear. If the static head is negligible, the loss in pump efficiency and the increase in slip at low speed cause a dip in flow at low speed (Figure 7.8b). The installed characteristic is the inherent characteristic.

For a significant static head, the pump speed must always be large enough to provide a pump head greater than the static head to ensure a positive flow. In Figure 7.8c, a pump speed less than 48 percent can result in a negative flow back through the pump. Check valves can be installed but in slurry and polymer service these check valves can fail to completely close. For example, the speed of a VFD on a reactant feed to a catalyst bed reactor dropped the pump head below the static head in the reactor. Reverse flow of reactants and catalyst in the reactor went through the VFD pump into the reactant tank creating a severe safety hazard. The check valve stuck open due to catalyst and polymer product. The plant consequently decided not to use VFD pumps.

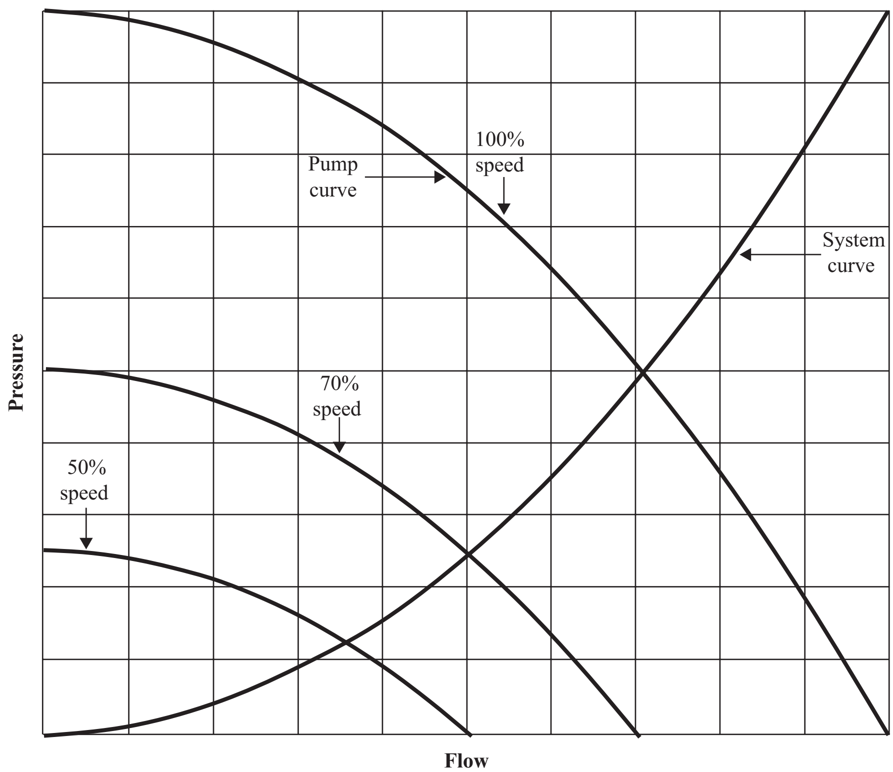


Figure 7.8a. Variable speed pump head and system pressure drop curves for *zero static head*.

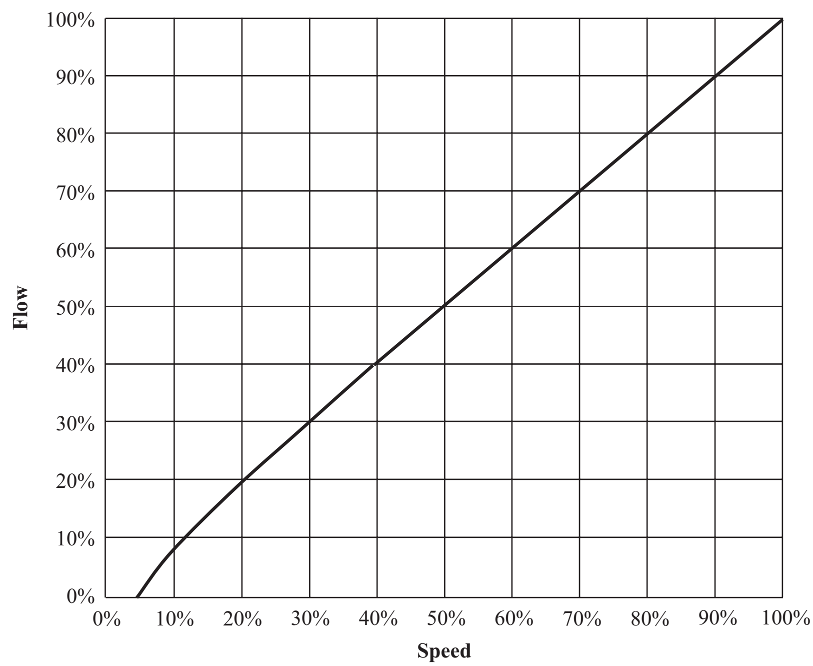


Figure 7.8b. Variable speed pump installed flow characteristic for *zero static head*.

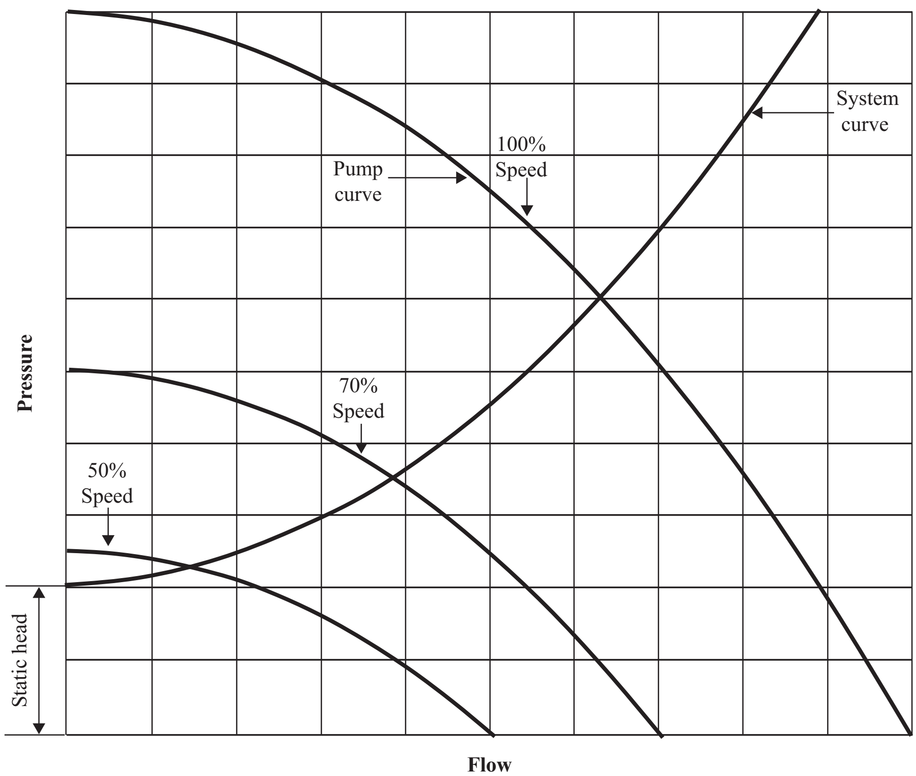


Figure 7.8c. Variable speed pump head and system pressure drop curves for *high static head*.

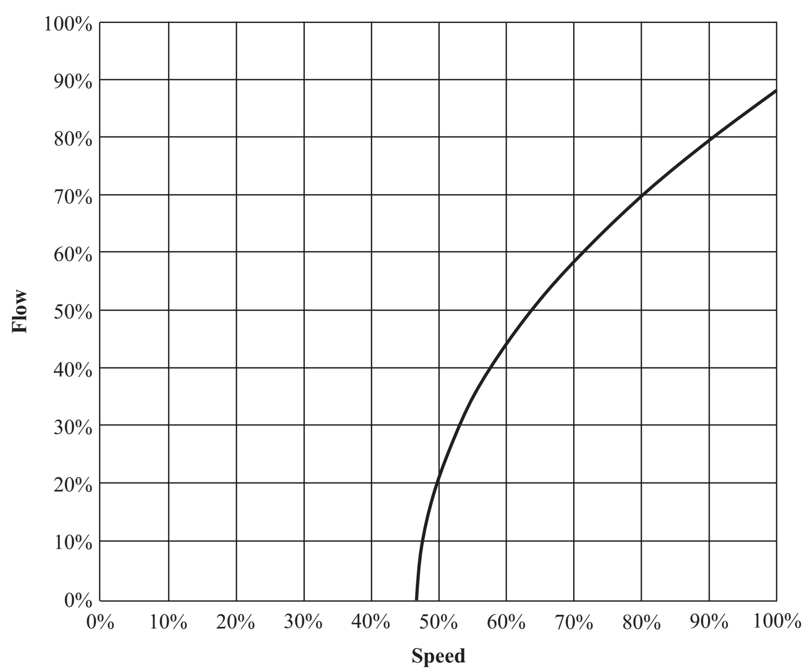


Figure 7.8d. Variable speed pump installed flow characteristic for *high static head*.

The static head from a downstream pressure can be difficult to predict. For a runaway exothermic reaction, the pressure can increase dramatically creating an abnormal situation not envisioned in the selection of the minimum speed. High static heads create hazards and severely limit the rangeability of the VFD.

If we ignore the loss in pump efficiency and increase in slip, a pump curve that approaches the static head will show a sharp bend downward to zero flow at low speed. The plummet of the speed depicted in Figure 7.8d at low speed causes a significant increase in gain and a nosier flow at low speeds. A flatter pump curve accentuates the nonlinearity.

7.6.4 VFD RANGEABILITY

If there was no static head and no slip and the motor and frame is properly designed to prevent overheating at low flows, the rangeability of a VFD would be impressive. A drive with closed loop slip control by the cascade of speed to torque control can achieve a rangeability of 80:1, which is comparable to the rangeability of a magnetic flow meter.

When the pump head is operating near the static head, the minimum controllable flow is set by rapid changes in the static head and frictional loss. These rapid changes could be due to noise and sudden or large disturbances. The speed cannot be turned down below the amplitude of these fast fluctuations. The rangeability for the case shown in Figure 7.8c and 7.8d is only 2:1 regardless of drive technology.

7.7 BEST PRACTICES

Nearly all of the literature on valve and VFD selection focuses on sizing and for valves leakage using data on process operating conditions and requirements. Unfortunately, there is no guidance on selection to improve the precision and response of valves and VFD. This neglect is mostly due to a lack of understanding that these final control elements are the means for affecting the production unit and responsible for the manipulation of nearly every process and utility stream. In this chapter we have seen the major problems and solutions. Towards the goal of achieving the best control loop performance, users should consider the following during the specification of control valves and VFDs.

7.7.1 CONTROL VALVE DESIGN SPECIFICATIONS

- Actuator, valve, and positioner package from a control valve manufacturer.
- Digital positioner tuned for valve package and application.
- Actuator that provides at least 150 percent of the maximum thrust or torque required.
- Diaphragm actuators where size permits (large valves and high pressure drops may require piston actuators).
- Sliding stem (globe) valves where size and fluid permits (large flows and slurries may require rotary valves).
- If rotary valves are used, verify from sizing tables the valve has a good inherent flow characteristic (e.g., v-ball or contoured butterfly).
- Low stem packing friction.

- Low sealing and seating friction of the internal closure element (e.g., plug, ball, disc).
- Booster(s) on positioner output(s) for large valves on fast loops (e.g., compressor anti-surge control).
- Valve Sized for a throttle range that provides good linearity:
    - 5 to 75 percent (sliding stem globe).
    - 10 to 60 percent (v-ball).
    - 25 to 45 percent (conventional butterfly).
    - 5 to 65 percent (contoured and teethed butterfly).
- Online diagnostics and step response tests for small changes in signal.
- External reset feedback (dynamic reset limit) using fast digital positioner feedback for slow valves to prevent PID output from changing faster than valve can respond.

7.7.2 VFD DESIGN SPECIFICATIONS

- High resolution input cards.
- Pump head well above static head.
- On-off valves for isolation.
- Design B TEFC motors with class F insulation and 1.15 service factor.
- Larger motor frame size.
- XPLE jacketed foil/braided or armored shielded cables.
- Separate trays for instrumentation and VFD cables.
- Inverter chokes and isolation transformers.
- Ceramic bearing insulation.
- Pulse width modulated inverters.
- Properly set deadband and rate limiting in the drive electronics.
- Drive control strategy to meet rangeability and speed regulation requirements.
- If tachometer feedback control is used, speed control should be in drive not DCS.
- External reset feedback (dynamic reset limit) using tachometer or inferential speed feedback to prevent PID output from changing faster than drive can respond.

7.8 TEST RESULTS

Test results were generated using a MiMiC virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in load flow at the process input (ΔF_L). Table 7.3 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The test cases use aggressive tuning settings computed with the short cut method to maximize disturbance rejection. Since these settings are right on the edge of causing an oscillation, the implied dead time is close to the original loop dead time.

Figures 7.9a, b, c, and d show the effect of 4 percent control valve deadband in the manual and automatic response of two common loops. The common loops chosen for the tests are flow and level to show the differences in the responses between a fast self-regulating and slower integrating process.

Table 7.3. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
7.9a, b	Flow	1 dimensionless	1	1	$\Delta F_L = 10\%$	Valve deadband
7.9c, d	Level	0.01 1/sec	10	—	$\Delta F_L = 10\%$	Valve deadband
7.10a, b	Flow	1 dimensionless	1	1	$\Delta F_L = 10\%$	Valve resolution
7.10c, d	Level	0.01 1/sec	10	—	$\Delta F_L = 10\%$	Valve resolution

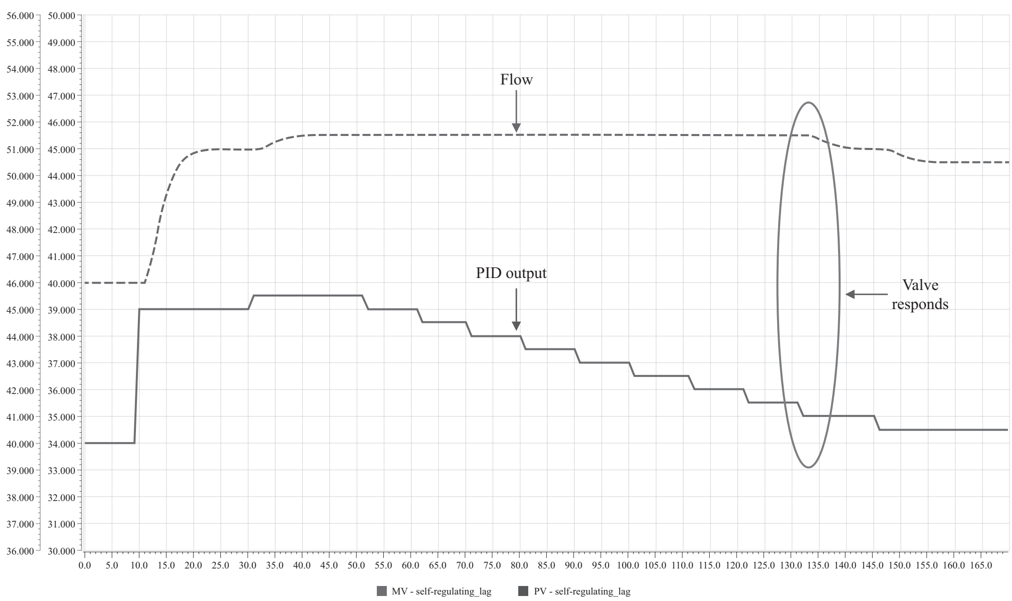


Figure 7.9a. *Flow* control *open loop* response for *4% deadband* (PID in manual).

For the manual test to the flow loop seen in Figure 7.9a, the first step up of 5 percent caused an immediate and full response in the flow measurement since the step was greater than the 4 percent deadband. A subsequent step up of 0.5 percent resulted in a good response as well since the step was in the same direction. When the direction is reversed, there is no flow response until the succession of steps down total more than 4 percent indicating the PID output has worked its way through the valve dead band from backlash. Small steps and a flow measurement with low noise and good precision offer an excellent way of quantifying the actual deadband and resolution limit. A fast readback of actual valve position works as well for well-designed control valves but not for on-off valves posing as throttling valves because movement of the actuator shaft does not correspond to movement of the internal closure element (e.g., plug, ball, and disc).

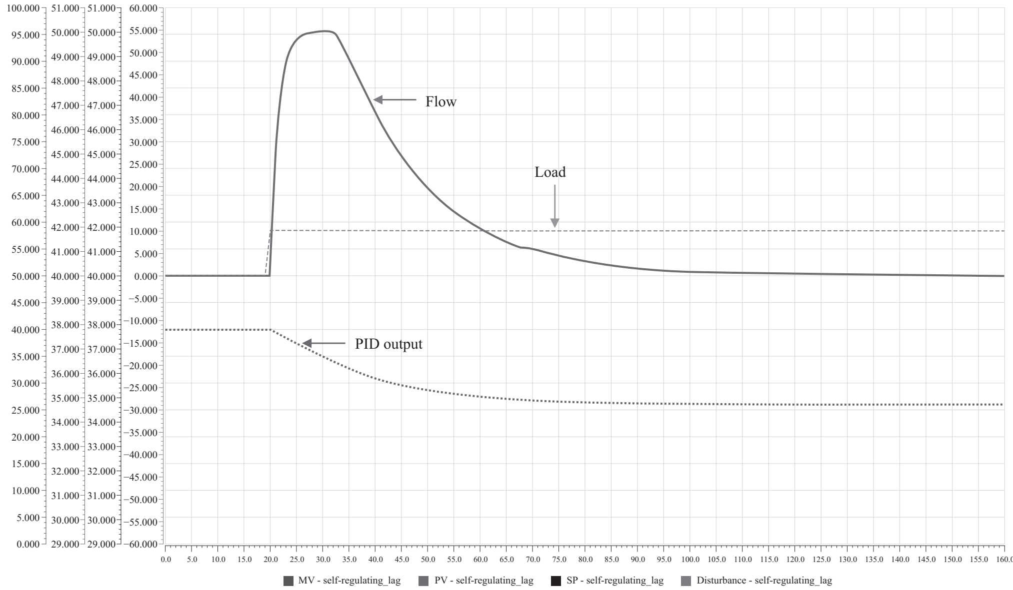


Figure 7.9b. *Flow* control *closed loop* response for *4% deadband* (PID in auto).

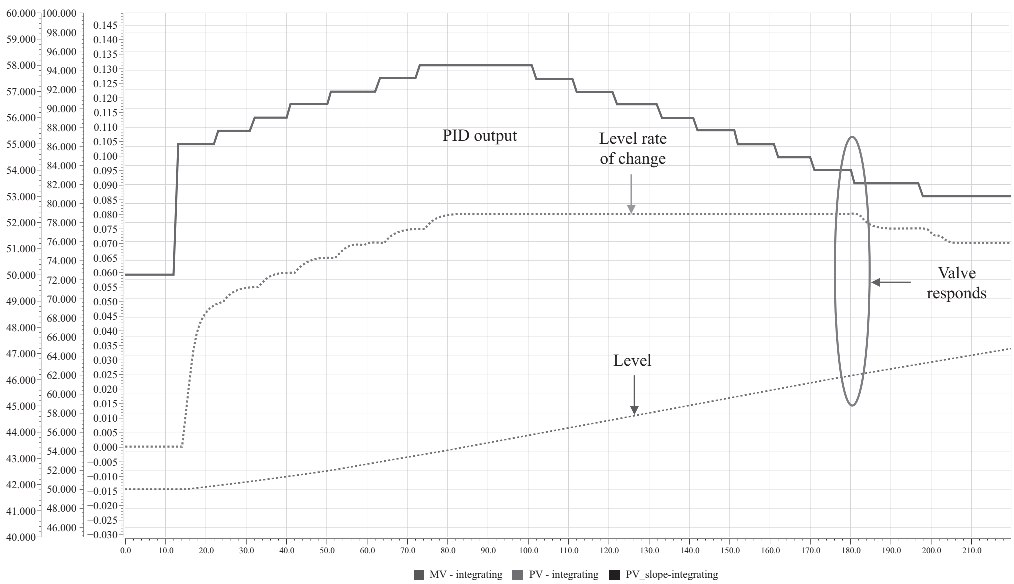


Figure 7.9c. *Level* control *open loop* response for *4% deadband* (PID in manual).

The automatic flow loop response to a 10 percent load disturbance in Figure 7.9b does not develop a limit cycle because there is only one integrator in the system (PID integral action). The effect of deadband is seen in the deterioration of the ability to handle load disturbances. The dead time has doubled from the time required for the PID output change from integral action to exceed the deadband. Since the aggressive tuning makes the implied dead time close to the

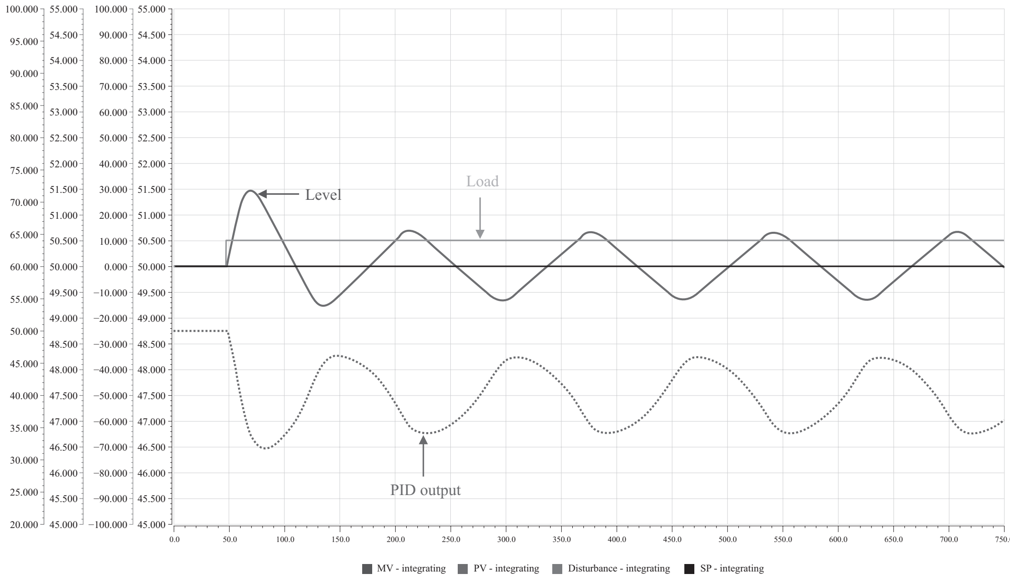


Figure 7.9d. *Level* control *closed loop* response for *4% deadband* (PID in auto).

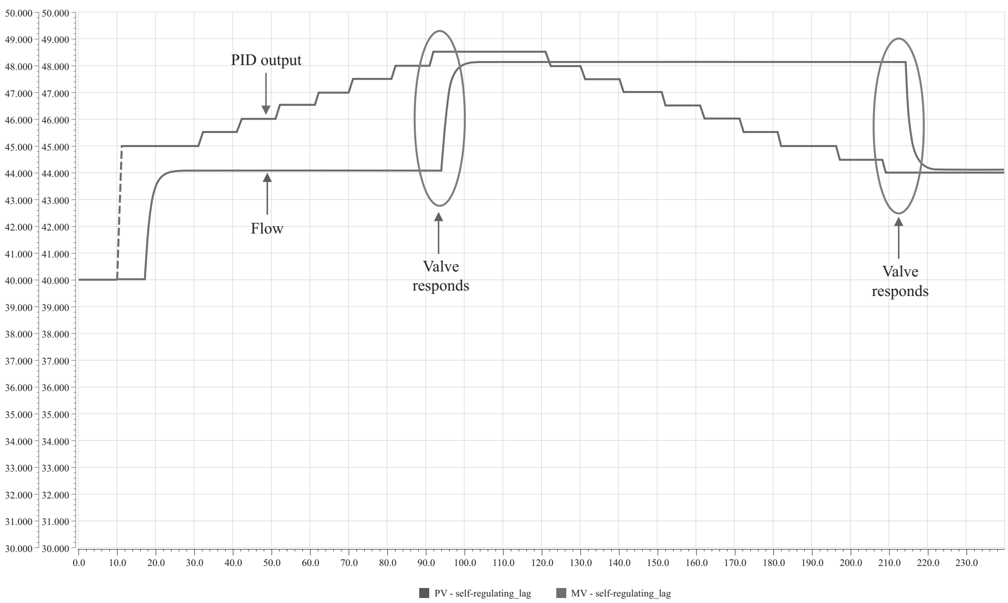


Figure 7.10a. *Flow* control *open loop* response for *4% resolution* limit (PID in manual).

original dead time, the doubling of the dead time results in nearly a doubling of the peak error and a quadrupling of the integrated error.

For manual tests of the level loop, the effect of deadband is difficult to see in Figure 7.9c even for this fast level response because the process is always gradually ramping in one direction or another due to the lack of process self-regulation. For real level processes, the ramp rate

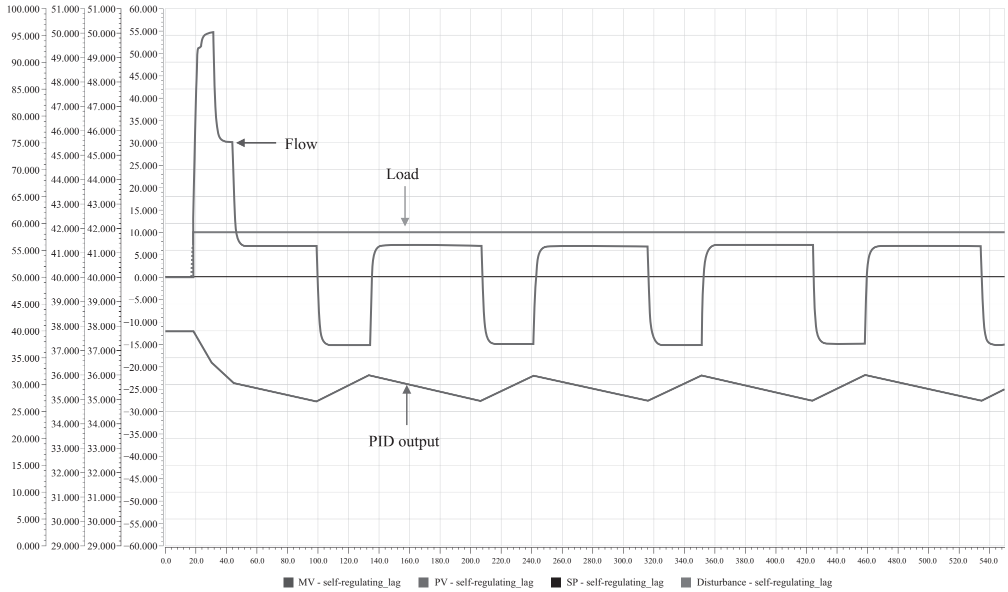


Figure 7.10b. *Flow* control *closed loop* response for *4% resolution* limit (PID in auto).

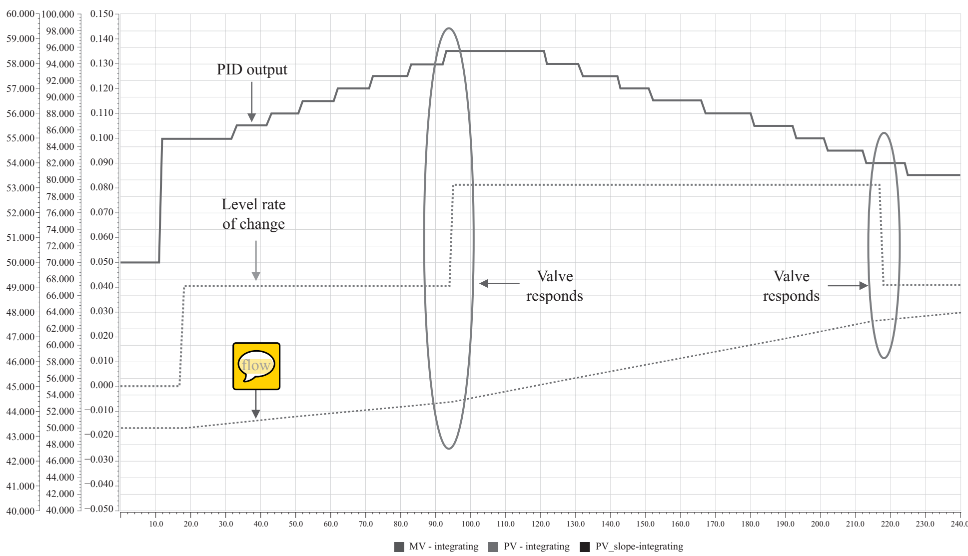


Figure 7.10c. *Level* control *open loop* response for *4% resolution* limit (PID in manual).

is several orders of magnitude slower making distinction of level changes from valve response impossible to detect unless a level rate of change is computed.

To compute the level rate of change with a fast update rate and good signal to noise ratio, the level is passed through a filter to attenuate noise and then into a dead time block. The output of the block (old level) is subtracted from the input to the block (new level) and divided by the

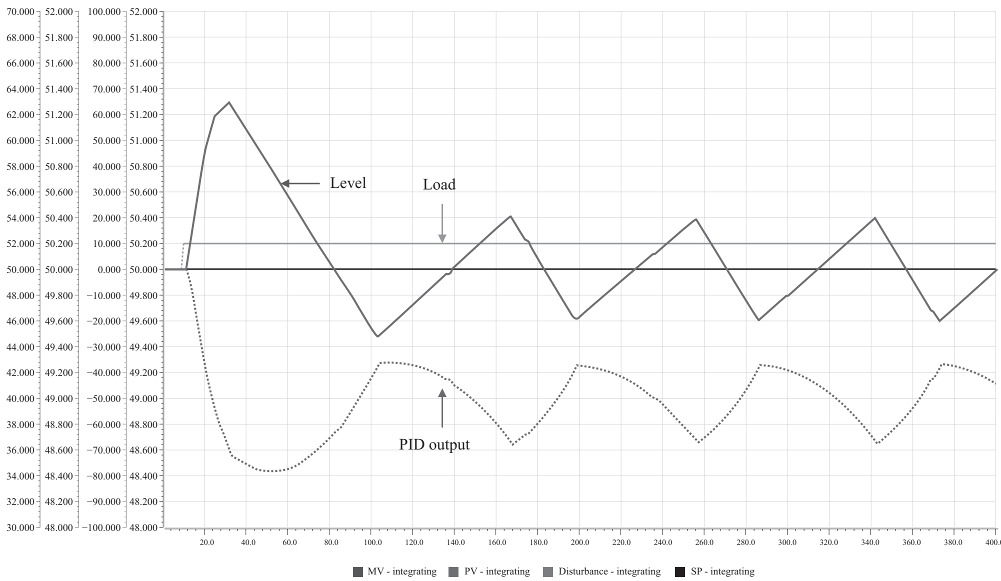


Figure 7.10d. *Level* control *closed loop* response for *4% resolution* limit (PID in auto).

block dead time to get the rate of change. The dead time interval is chosen to be large enough (e.g., 10 seconds for this test but 1,000 seconds for plant level) to show a change in level that is much larger than any noise at dead time block. The rate of change is updated every execution of the dead time block (e.g., one second) once the dead time block queue has filled up after a download. The level rate of change is an inferential flow measurement. In fact this method is used to compute discharge flow rates from level and weight measurements on feed tanks. The use of a dead time block creating a train of updates eliminates the delay associated with the time interval in a periodic calculation.

Figure 7.9c shows how the level rate of change response is similar to flow. The valve responded to the first 5 percent step since it is larger than the deadband and then responded to subsequent 0.5 percent steps since these were in the same direction. When the steps reversed direction, there is no response in the level rate of change until the steps total more than the dead band.

The automatic level response seen in Figure 7.9d develops a limit cycle from backlash since there are now two integrators in the system (level process and PID integral mode). The waveform seen is typical for integrating processes. The level cycle is a sawtooth since the level response from integrating process action is a ramp up and down. The PID output cycle has rounding of the peaks and valleys making the cycle look more like a sine wave. An integrating process with a significant secondary time constant such as a thermal lag in a batch temperature would also show some rounding of the PV peaks and valleys. If integral action is suspended in the PID by the use of integral deadband, a PID with external reset feedback of a fast valve position readback, or an enhanced PID developed for wireless, the limit cycle will stop since there is only one integrator left in the system.

Figures 7.10a, b, c, and d show the effect of 4 percent control valve resolution in the manual and automatic response of two common loops. For the manual tests in Figures 7.10a and 7.10c, the resolution limit prevents the response to a small step even when in the same direction until the net total of the steps in one direction exceeds the resolution limit. For the level process, a level rate of change must be computed as before to show the effect of the valve response on the level process.

The flow loop in automatic develops a limit cycle in Figure 7.10b because only one integrator in the system is needed to trigger the limit cycle. The flow cycle is a square wave since the time constant in the flow process is only about one second and there is no PV filter resulting in little smoothing of the valve action. The PID output ramps from the integral mode reaction to the steps in the PV PID creating a sawtooth. These waveforms are typical for any loop where the primary time constant is small and the valve is fast. If integral action is suspended in the PID, the limit cycle will stop since the only integrator in the system is stopped.

The limit cycle for the automatic operation of the level loop seen in Figure 7.10d shows a similar ramping action and sawtooth in the level response except the corners are sharper for the resolution limit. The PID output cycle shows less rounding and some scalloping at the peaks and valleys as a result of the higher PID gain associated with a level loop PID tuning acting on the sharper reversals of the direction of the level.

KEY POINTS

1. The principle problem in the dynamic response of valves is a deadband from backlash and a resolution limit from actuators and valve friction.
2. Control valves should not be used for isolation and on-off valves should not be used for throttling.
3. Positioner diagnostics from on-off piping valves used as control valves may be meaningless when the positioner feedback is on actuator shaft movement that does not match the response of the internal closure element due to backlash, shaft-windup, and stick-slip.
4. A careful examination of the entire actuator-shaft-stem-positioner-packing-seating-sealing design is needed to avoid being fooled by an on-off valve posing as a control valve. Clues as to the deception are the original intent of the valve (on-off or throttling) and whether the manufacturer's products were originally intended to be piping or control valves. While on-off valves are generally cheaper with higher capacity and lower leakage, the life cycle cost in terms of process variability and analysis is excessive.
5. To keep the correction time of large control valves less than the PID reset time, a volume booster may need to be added on each positioner output with the booster bypass opened enough to ensure stability.
6. Boosters should never be used in place of positioners on rotary valves with diaphragm actuators because a positive feedback loop is created by the high outlet port sensitivity of the booster and an excessive resolution limit is created by the poor inlet pressure sensitivity of the booster.
7. The installed flow characteristic of a control valve and VFD determines the linearity and rangeability of these final control elements.
8. A large deadband or resolution limit will cause a loss of rangeability in both a valve and a VFD.
9. A well-tuned digital positioner (digital valve controller) can increase the rangeability of a control valve.

10. A low valve pressure drop to system pressure drop ratio will cause a loss in rangeability in control valve.
11. A high static head will cause a loss of VFD rangeability.
12. Closed loop slip control (cascade control of speed to torque in the drive) can double the VFD rangeability from 40:1 to 80:1 for zero static head applications.

CHAPTER 8

Effect of Disturbances

8.1 INTRODUCTION

Much of the differences in approaches to controller algorithms and tuning can be traced back to assumptions made about the type and importance of disturbances. Each method has merits based on the disturbance frequency, location, and time lag. Here we gain an understanding of how to reduce process variability from upsets originating from changes in raw materials, production rates, weather, operating conditions, or other loops. The knowledge provides the guidance on how to track down the origin and eliminate or at least reduce the size or speed of the disturbance.

8.1.1 PERSPECTIVE

The emphasis in the control literature is on setpoint response. When the ability to handle disturbances is studied, a step disturbance is typically shown as entering the loop at the process output at the point of the measurement. Often sensor and measurement delays and lags, filter time, and proportional-integral-derivative (PID) module execution time are not included. Consequently, the disturbance appears immediately at the PID input. This approach simplifies the mathematical analysis and the dynamic compensation of feedforward signals and shows the advantage of model based algorithms, such as Internal Model Control (IMC). The tuning to prevent overshoot for a step disturbance in the process output is the same as for a setpoint change for a proportional-integral (PI) on error structure. In contrast, the tuning for a step disturbance in the process input (load disturbances) can use more aggressive tuning to minimize peak and integrated errors. We can also use this aggressive tuning for load disturbances for setpoint changes by the use of options, such as two degrees of freedom (2DOF) structure or the introduction of a setpoint filter or lead-lag to reduce overshoot without excessive increase in the time to reach setpoint. For disturbances on the process output these options just mentioned for setpoint changes are not applicable. Consequently, tuning rules were developed that focused on a setpoint response without these options because this tuning also reduced overshoot and oscillation for disturbances on the process output. As a result, the IMC and Lambda tuning methods use a tuning parameter that is the closed loop time constant (time to reach 63 percent of a setpoint change after dead time).

The rules work reasonably well for moderate self-regulating processes where the primary process time constant is about the same size as the dead time by setting the tuning parameter to be a multiple of the dead time. Rules were expanded to deal with dead time dominant and near or true integrating responses. For integrating processes, Lambda tuning switches to an arrest time (time to stop an excursion after the dead time) that addresses load disturbances. As one can imagine this sets us up for a remarkable spectrum of proposed solutions and a considerable difference of opinions based on a specific disturbance location and type of open loop response. If all of the processes were self-regulating with a process time constant about the same size as the total loop dead time, whether the disturbance is on the process input or output would not matter. Since this is not the case, a switch in the Lambda tuning parameter and rules by considering processes with relatively large process time constants to be near-integrating with low limits on the reset and rate time enables minimization of variability from process input disturbances.

The disturbance most commonly encountered that is of greatest interest enters into the process typically upstream of the process dynamics and is a change in flow, typically a feed flow, made by a level controller or an operator. This disturbance entering the process about the same point as the manipulated flow is termed a load disturbance. Changes in composition or temperature of the feed or manipulated flow are also considered load disturbance but these are usually much slower.

Liquid processes with a large volume with mixing due to agitation, boiling, or recirculation have a large process time constant. Step disturbances at the process input are slowed down and in the case of oscillatory disturbances attenuated by the filtering action of the process time constant that can be approximated as the well-mixed volume residence time (volume/flow). For inline liquid and gas flow unit operations, the residence time is short and ends up being mostly a transportation delay due to the lack of back mixing. For these processes with a small process time constant, the location of the disturbance does not significantly impact the errors or tuning strategy. In fact, these loops can be tuned more in terms of achieving the best setpoint response.

Surge tank level controllers should be tuned to fully utilize the available volume to reduce the size and speed of flow changes to downstream equipment. The same criterion applies to column sump level controllers that manipulate bottoms flow and overhead receiver level controllers that manipulate distillate flow. Equations 1.21a through 1.22k show how to set the Lambda arrest time to maximize the absorption of flow variability.

Operators tend to make abrupt feed changes overreacting to situations and seeking perceived sweet spots. Humans are not good at anticipating the effect especially when the dead time is significant. Most of the feed changes by operators are due to production rate changes or equipment problems or simply a change in shift. A plantwide feedforward scheme can make the change in production rate slower, smoother, and more consistent with the process requirements. A valve position control (VPC) or model predictive control (MPC) system can make gradual and timely corrections to keep the process close to a constraint without violation of the constraint. If a flow controller is in cascade or remote cascade mode rather than auto mode, there is less of a need or temptation for the operator to change the flow.

For a given size disturbance, the impact increases with the rate of change of the disturbance. The worst case is the step disturbance seen throughout the literature. Step disturbances result from compressors, fans, or pumps starting or stopping and from relief valves or on-off valves opening or closing. These actions are typically initiated by manual actions, sequences (e.g., batch operations and automated startups and transitions), and safety instrumentation systems. Snubbers (restrictors in the air lines) can be used to slow down the stroke of on-off valves but the adjustment is not as accessible or flexible as the tuning settings and analog output (AO) block setpoint rate limits in a PID.

Most disturbances are not a step change because flows are typically manipulated by PID with reset action. If a flow loop is used, the PID tuning uses more integral action rather than proportional action (e.g., PID gain = 0.2 and reset time = 2 seconds) to deal with the valve nonlinearities. The flow control closed loop time constant (Lambda) and thus the disturbance time constant for the process loops affected by the flow change is about 10 seconds. If there is no secondary flow loop, the feedback action of primary process composition and temperature loops has even larger closed loop time constants. However, in the case of setpoint changes rather than load disturbances, there is a large initial step from proportional action and a kick from derivative action for a structure with PID on error. If a secondary flow loop is not used, the primary PID output changes are immediately passed on as abrupt valve position or speed changes and hence flow changes to affected loops. Thus, the advantage of a secondary flow loop extends beyond isolating the primary process loop from valve and speed nonlinearities to slowing down the most prevalent fast disturbance being flow and enabling flow feedforward control (e.g., flow ratio control) to deal with the disturbance, directly compensating for most of disturbance before it affects other primary process loops.

For continuous operations there are not many setpoint changes. The disruptive nature of setpoint changes is more an issue for batch operations and automated startups and transitions in product grade or type. In Chapter 2 on Unified Methodology, the relative merits of different PID structures and set-point lead-lag are discussed in terms of reducing the step and kick without overly slowing down the setpoint response. Note that for many batch pressure and temperature loops in the chemical industry, the time to reach setpoint is more important for reducing batch cycle time than minimizing steps in utility flows. An AO block setpoint rate of change limit can be used with external reset feedback to slow down the action of the control valve. The resulting move suppression can be set to be slow enough to prevent a large disruption to the manipulated flow source (e.g., utility header) without appreciably affecting the time the controlled variable (e.g., reactor temperature) gets to setpoint.

The equations in Chapter 3 for peak and integrated error are for step changes in load. A correction factor can be applied to these equations that show the beneficial effect of slowing down the disturbance. For a disturbance with a first order time constant, the factor is simply the fractional exponential response in the critical time frame of the PID correction that is estimated as the dead time plus Lambda. The fastest reasonable response is a Lambda equal to one dead time. Due to unknowns, a Lambda equal to two dead times is used. For a disturbance that ramps, the open loop error (process variable error if PID is in manual) is replaced by the open loop ramp rate (process variable ramp rate if PID is in manual) multiplied by the dead time plus Lambda. For a disturbance that ramps due to a near or true integrating process, the open loop error (process variable error if PID is in manual) is replaced by the open loop ramp rate (process variable ramp rate if PID is in manual). The time units of the open loop error are cancelled out by the time units in the integrating process gain in the equations in Chapter 3 for the integrated error and peak error.

Slow load disturbances will exhibit longer recovery times (slow protracted approach to go back to setpoint). An increase in integral action (decrease in reset time) can help the PID deal with the continual increase in the load with time. The load disturbance will appear to accelerate until the inflection point is reached in the exponential response of a load upset characterized by multiple disturbance time constants.

Oscillatory disturbances are particularly problematic because a perpetual state of upset is created and the possibility of resonance exists. If the period of the disturbance is near the ultimate period of a loop, closed loop control will increase the amplitude (resonance). The best solution is of course to eliminate the oscillatory disturbance. Most often these oscillatory disturbances are caused by inappropriate tuning, valves with excessive backlash or stiction, batch operations, and on-off control. PID control should replace on-off control (e.g., level measurement and PID control instead of level switches). In terms of tuning, the most common mistake is a reset time that is too small particularly for level loops. For surge tank level control, the transfer of a change in inlet flow from batch operations to manipulated outlet flow can be smoothed to take advantage of available inventory. For valves, the use of rotary valves designed for tight shutoff with piston actuators is the most frequent culprit. If the disturbance period is significantly less than twice the ultimate period, the amplification can be reduced by tuning the affected PID slower (smaller gain and rate time and greater reset time). Feedforward can provide a preemptive action reducing the need for feedback control and the consequences of slowing down the PID tuning. If the disturbance period is much larger than twice the ultimate period, the tuning solution is to make the PID faster (larger gain and rate time and smaller reset time). The application of special notch filters is highly dependent upon accurate knowledge of the source of the disturbance. This book advocates the careful judicious use of the standard Distributed Control Systems (DCS) first order filter shown to be of greatest general utility in industrial processes where nonlinearities and unknowns rule.

8.1.2 OVERVIEW

If there were no unmeasured disturbances, feedback control would not be necessary. Process engineers and operators could home in on the best PID output and just leave it at this value. In fact many process engineers are much more comfortable with setting a stream flow as per a process flow diagram than relinquishing the manipulation of the flow to a PID controller that they don't quite understand. In batch operations, often flows are sequenced based on process design and empirical knowledge rather than released to a PID loop for fed-batch control. Also algorithms could be designed to focus on providing the best setpoint response and compensating for known disturbances.

The reality is all loops are subjected to disturbances that are not measured and in many cases from an unknown source. Furthermore, these are primarily load disturbances that enter as inputs into the process, complicating the analysis of the loop response to the disturbance making dynamic compensation of feedforward signals more complicated. The prevalence of the PID is due to the ability to deal with unmeasured load disturbances.

For the unusual case where the lag in the path of the disturbance to the process output is much less than the lag in the path of the manipulated variable to the process output, the PID reset time must be increased. The reset time of a controller tuned for maximum load disturbance rejection should be increased by a factor of four for a PID and two for a PI controller to minimize overshoot for the extreme application where the disturbance path lag is less than one tenth the manipulated variable path lag per test data by Shinskey. This is in addition to any increases in the reset time due to unknowns or nonlinearities.

If the load disturbance path has a lag that is more than four times the lag in the manipulated variable path to the process output, the PID gain can be increased and reset time decreased per test results in this chapter. However, these aggressive settings make the PID more vulnerable to fast disturbances and to overshoot for setpoint changes. The load disturbances must always be slow and the factors for a setpoint lead-lag or the PID structure (e.g., setpoint weights beta and gamma) must be adjusted to prevent the increase in overshoot in the setpoint response from the smaller reset time.

Disturbances should be tracked down and eliminated or at least reduced. The best point for mitigation can be found by examining the pathways of variability. For recycle streams and heat integration, the source can be downstream besides upstream. For interactions, operator and control system actions on the common equipment need to be examined and tuning or decoupling be used. The use of trend charts with intelligent process variable and time scaling is essential. Tools such as power spectrum analyzers and data analytics (e.g., principal component analysis [PCA]) are very helpful to identify candidates for root causes.

Disturbances that upset important loops and cannot be eliminated should be slowed down as much as possible by the use of secondary flow loops and the use of a PID structure that reduces the proportional and derivative reaction to setpoint changes. Feedforward of measured load disturbances should be used to provide preemptive correction eliminating as much as possible the need for feedback correction.

Resonance and interaction should be reduced by tuning to provide significant separation of dynamics in terms of periods of oscillations. The feedforward of offending PID outputs to the most critical loop should be used to decouple the loops if tuning does not solve the problem or the increased variability in the detuned loop is unacceptable. When multiple interactions exist and the dynamic compensation in the feedforward signals is critical, MPC is a more sure proof solution.

8.1.3 RECOMMENDATIONS

1. Eliminate all manual and as much as possible on-off actions.
2. Slow down disturbances by replacing on-off control with PID control and tuning the PID with an appropriate closed loop time constant or arrest time or add AO block setpoint rate limits with external reset feedback.
3. Tune PID controllers to have a nonoscillatory response for worst case conditions.
4. Tune surge tank level controllers to maximize the absorption of flow variability.
5. Take advantage of liquid volumes to attenuate oscillations (e.g., use volumes downstream of inline pH control system to filter oscillations).
6. Use intelligent control with embedded process knowledge (e.g., plantwide feedforward control, VPC, and MPC) and move suppression to provide an automatic gradual optimization, eliminating the need for setpoint changes or intervention by operations.
7. Do not introduce excessive deadband in the variable speed drive setup.
8. Use valves with the least backlash and stiction and actuators and positioners with the best threshold sensitivity and resolution.
9. Measure all potential disturbances and use feedforward control for load upsets that are too fast or large to be adequately corrected by feedback control.
10. For step disturbances on the output of the process, reduce the PID gain and increase the reset time to reduce overshoot.
11. For step disturbances on the output of a process with significant process or final control element dead time, realize feedforward corrections will arrive too late and the best bet is to slow down the disturbance to enable feedback correction.

12. For step disturbances on the output of a process with a significant process or final control element lag but negligible dead time, add a lead to the feedforward signal to compensate for the lag in the feedback path.
13. Use secondary flow loops to slow down flow changes and enable flow feedforward (e.g., flow ratio control).
14. For setpoint changes to primary loops that upset other important loops, use the 2DOF PID structure or a setpoint lead-lag to keep the setpoint response fast while slowing down the disturbance to other loops (see Chapter 2).
15. Tune the PID controllers to prevent amplification of disturbance oscillations near the ultimate period (increase closed loop time constant to decrease PID gain to reduce resonance).
16. To reduce interaction, make fast loops faster or slow loops slower so that the closed loop time constants of the loops are dramatically different.
17. Use a feedforward of offending PID output to decouple affected PID.
18. If load disturbances at the process input are always so slow that the approach back to setpoint is much slower than the initial excursion in an over-damped (nonoscillatory) response, cautiously increase the PID gain and decrease the reset time to help the PID deal with a growing disturbance and adjust the setpoint lead-lag and structure to prevent excessive overshoot in the setpoint response. For this case, the disturbance reaches a peak in the departure from setpoint in about two to four dead times but takes more than 40 dead times after the peak to return to setpoint. There is no undershoot, just a long protracted approach back to setpoint.

8.2 DISTURBANCE DYNAMICS

In the control literature, the disturbance is generally taken as a step change. In industrial processes, there are few true step changes even for sequenced flows because of the valve slewing rate and time constant of intervening volumes. The exceptions are compressor surge and water hammer and other disturbances involving momentum balances instead of just material and energy balances. Small volumes (e.g., utility headers) and fast valves can create nearly step disturbances. Most often the disturbance can be characterized by a time constant, rate limit, a dead time, or an oscillation period and amplitude.

8.2.1 LOAD TIME CONSTANTS

The most common process disturbance is a change in process or utility flow. The change in flow setpoint may be a step but the change in PID output is gradual because integral action is used more than proportional action in flow loops. Loops where proportional action dominates have a high controller gain because of the existence of a large process time constant that in turn provides a gradual process response. Consequently, the response of loops on fast and slow processes can be characterized by a closed loop time constant (τ_c) that is set as the Lambda parameter in the Lambda tuning method for self-regulating processes. The load disturbance time constant can be taken as being the Lambda (λ_L) of the loop that is the source of the load disturbance. If a different tuning method is used, the tuning rule defined by Equation 1.6c can be solved for Lambda. This equation can be used for disturbances originating from integrating processes by finding the equivalent self-regulating process open loop gain and time constant by the use of Equation 1.5a for the conversion of dynamics between true and near-integrating processes.

Composition, pH, and temperature disturbances as they pass through an intervening volume have a time constant as indicated by Equation 8.1b that is the residence time of the portion of the total volume that is perfectly mixed. Volumes with non-ideal mixing are split into plug flow and perfectly mixed volumes. The plug flow volumes acts as a pure transportation delay and the perfectly mixed volume provides a process time constant. The agitated vessel entry in Table 4.2 shows that the total residence time minus the turnover time from agitation and recirculation is approximately the perfectly mixed residence time and hence the primary process time constant.

Equations 3.11a and 3.11b show the effect of a load disturbance time constant on the peak error for given settings. Since the effect of the primary time constant in the affected process is also a factor, Equation C.1 in Appendix C is multiplied by the load time constant factors for the originating loop and intervening volume. The result is Equation 8.1a for the peak error that includes the effect of both time constants and the tuning of the loops that are the source and destination of the disturbance.

The Lambda (λ_L) for the loop that is the origin of the load disturbance (e.g., flow loop) is a closed loop time constant. The Lambda (λ_o) for the loop that is the destination of the disturbance is an arrest time since the primary loop considered (e.g., composition, pH, or temperature) has a near-integrating, true integrating or runaway response.

$$E_x = [1 - e^{-(\theta_o + \lambda_o)/\tau_o}] * [1 - e^{-(\theta_o + \lambda_o)/\tau_L}] * [1 - e^{-(\theta_o + \lambda_o)/\lambda_L}] * E_o \tag{8.1a}$$

For intervening back mixed liquid volumes between the origin and destination for composition, pH, and temperature control the load time constant is the residence time:

$$\tau_L = \frac{V_b}{F_f} \tag{8.1b}$$

where

E_o = open loop error (error with controller in manual) (e.u.)
E_x = peak error from the load disturbance (e.u.)
F_f = total feed rate to intervening liquid volume (m^3/sec)
V_b = back mixed intervening liquid volume (m^3)
λ_o = Lambda (arrest time) of the primary loop affected by the load disturbance (sec)
λ_L = Lambda (closed loop time constant) of the loop (e.g., flow loop) that is the source of the load disturbance (sec)
θ_o = total loop dead time (sec)
τ_o = open loop time constant of loop that is affected by disturbance (sec)
τ_L = load disturbance time constant of intervening volume (sec)

The effect of load time constants on integrated error is not as dramatic because the loop has difficulty catching up with an exponentially growing disturbance (first order exponential response of load disturbance). The result is a long protracted approach to setpoint after the peak error. In the test results Figure 8.2 shows the return to setpoint becomes painfully slow for a load time constant that is 16 times the dead time. Since the ultimate period is close to 40 seconds, this case corresponds to a load time constant that is about four times the ultimate period. Figure 8.3 reveals a slight increase in integral action can help the PID catch up to the disturbance.

8.2.2 LOAD RATE LIMIT

An on-off valve being stroked will create a rate of change disturbance in flow based on the slewing rate and installed flow characteristic of the valve. Many on-off valves have nearly a quick opening characteristic. As a result for a closing valve, the initial rate of change of flow is nearly zero until the valve is less than 50 percent open. As the stroke gets within 25 percent of the closed position, the rate of change dramatically accelerates. To provide a more uniform rate of change, a flow control loop is preferred. On-off valves should be used in series that are coordinated with the throttle valve to stop dribbling and leakage at the end of batch charges. Setpoint rate limits on the flow PID can provide a smooth and consistent rate of change.

Move suppression also creates a load rate limit. Move suppression is the most frequently used tuning adjustment in MPC. Directional move suppression has been found to be a powerful option in VPC by the use of setpoint rate limits on an AO or flow loop. The VPC PID uses external reset feedback of actual valve position or manipulated flow to provide a smooth response.

The peak error from a rate limited load disturbance can be approximated by Equation 8.2a by multiplying the rate of change by the sum of the destination loop dead time and Lambda. The rate of change of flow is converted in Equation 8.2b to a rate of change of the open loop error by the multiplication of the rate of change of a flow by a process gain.

When the disturbance stops there is a second upset from the sudden stoppage of the change in flow. The time constant of an intervening or destination volume with some mixing will smooth out the sharp corner seen in the flow trend chart so that the process can make a smooth transition. Note that the time constants of these volumes will appear as a delay between ramps of the open loop error and the disturbance flow. If such a volume doesn't exist, a small filter time constant should be added to the AO or flow setpoint. The use of Lambda tuning will automatically provide the effect of the filter via a closed loop time constant.

$$E_x = \left(\Delta E_o / \Delta t\right) * \left(\theta_o + \lambda_o\right) \tag{8.2a}$$

$$\Delta E_o / \Delta t = K_p * \left(\Delta F_f / \Delta t\right) \tag{8.2b}$$

where

$\Delta E_o / \Delta t$ = rate of change of open loop error (error with controller in manual) (e.u./sec)
E_x = peak error from the load disturbance (e.u.)
$\Delta F_f / \Delta t$ = rate of change of feed flow (m^3/sec/sec)
F_f = total feed rate to intervening liquid volume (m^3/sec)
λ_o = Lambda of the loop affected by the load disturbance (sec)
θ_o = total loop dead time (sec)

8.2.3 DISTURBANCE DEAD TIME

A disturbance dead time has no direct effect on the minimum peak error. A disturbance dead time will affect the dynamic compensation of feedforward signals if the dead time occurs after the feedforward measurement. In this case, the feedforward signal must be delayed to prevent a feedforward correction arriving before the disturbance and causing inverse response. Conversely, if the feedforward measurement is delayed (e.g., sample transportation delay and analyzer cycle time) resulting in a feedforward correction dead time that is larger than the disturbance dead time to the same point in the process, the feedforward correction will be late. While a lead time can cancel out a lag time, there is nothing that can cancel out dead time. Consequently there is no feedforward dynamic compensation for this situation and the best thing that can be done is to simply decrease the feedforward gain. If the excessive dead time in the feedforward correction is greater than the total loop dead time, feedforward may do more harm than good due to the size of the secondary oscillation created.

8.2.4 DISTURBANCE OSCILLATIONS

The minimum peak error from an oscillation if there is no resonance can be estimated by Equation 8.3a which is a reuse of Equation 6.7a developed to show the attenuating effect of a time constant on an oscillation amplitude. Equation 8.3c is a translation to the time domain by the use of Equation 8.3b for the relationship between oscillation frequency and period.

For oscillation periods between half and two times the ultimate period (e.g., two to eight dead times), resonance can be occurring. Figure 8.1 shows the general effect of resonance where the action of feedback control becomes in phase with the disturbance, causing an amplification of the oscillation amplitude. The amplification becomes greater as the PID is more

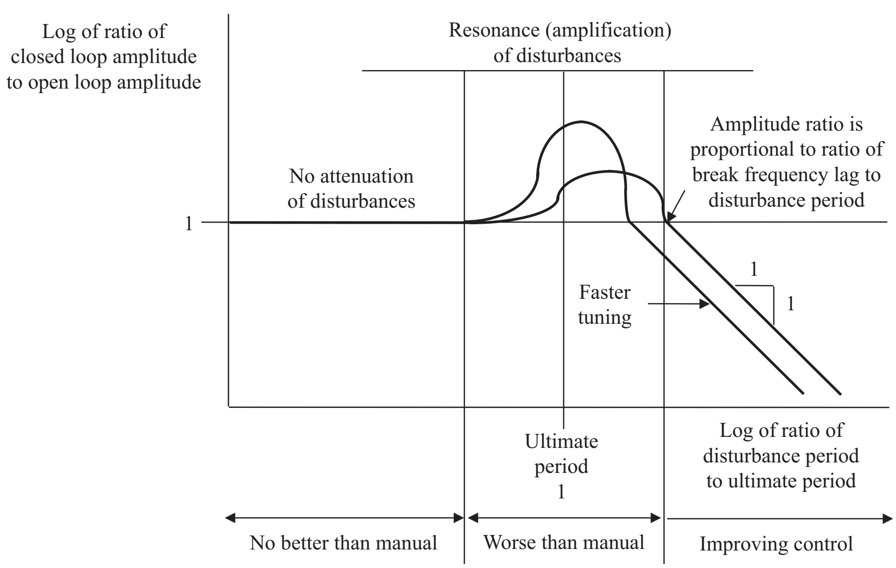


Figure 8.1. Attenuation of oscillation amplitude by primary process time constant and resonance from feedback control.

aggressively tuned. Note that abscissa (X axis) of this plot is the log of the ratio of oscillation period to ultimate period whereas the literature uses an abscissa that is the log of the ratio of the oscillation frequency to the natural frequency. An abscissa in the time domain enables a better visualization from trend chart oscillation periods and estimating the ultimate period as simply four times the dead time but the result is horizontal flip of what is normally seen in the literature.

$$A_x = A_o * \frac{1}{\sqrt{1+(\tau_o * \omega_o)^2}} \tag{8.3a}$$

Converting the oscillation from the frequency domain to the time domain:

$$\omega_o = \frac{2*\pi}{t_o} \tag{8.3b}$$

The resulting equation shows attenuation increases as the time constant to period ratio increases:

$$A_x = A_o * \frac{1}{\sqrt{1+(\tau_o * \frac{2*\pi}{t_o})^2}} \tag{8.3c}$$

where

A_o = original oscillation amplitude (sec)
A_x = attenuated oscillation amplitude (sec)
t_o = period of oscillation (sec)
τ_o = open loop time constant of process affected by oscillation (sec)
w_o = frequency of oscillation (radians/sec)

For oscillation periods much less than the ultimate period, the process provides significant attenuation per Equation 8.3c but the PID provides no feedback correction as seen in the Figure 8.4a test results for a fast load oscillation. These oscillations can be effectively considered to be noise. PID tuning and feedback action does not affect the process oscillation. Figure 8.4b shows more aggressive tuning causes unnecessary extra movement of the valve from reaction to what is effectively noise.

For oscillation periods much greater than the ultimate period, the process provides no attenuation. Here PID feedback correction can be significant. More aggressive tuning provides a greater reduction in the oscillation amplitude as seen in the Figure 8.5a test results for a slow load oscillation. The benefit from aggressive tuning is not as noticeable as the load oscillation period increases toward the point where PID control can make the oscillation in the process variable disappear to the point of being lost in noise and measurement or final control element precision limit cycles. The oscillation will be totally visible in the PID output. Tight PID control will almost completely transfer variability from the process variable to the manipulated variable. This transfer has many implications in terms of tracking the path of variability propagation and what variables are chosen for developing neural network models and projection to latent structures (partial least squares) models.

8.3 DISTURBANCE LOCATION

Whether a disturbance passes through a large primary process time constant is the primary source of disagreement in control. Actually disagreement is a very polite term for the extremely negative published and vocalized criticism of the opposing view and associated methodology including tuning rules. The objective here and in this book is to provide more comprehensive understanding that reveals the basis of the opposing views, how each camp is right in their own way, and how the techniques converge. The use of principles on the effect of process dynamics and tuning can expand our minds and learn from rather than discredit an opposing view.

The author comes from a background of practitioners and consultants in the chemical industry and the original suppliers control systems in the Northeast (e.g., Foxboro, Fischer & Porter, Leeds & Northrup, and Taylor) dating from the 1950s to the 1980s. The most notable consultant from this era is Greg Shinskey whose accomplishments, articles, and books have provided knowledge on the effect of chemical engineering and mechanical engineering design that is orders of magnitude deeper and more extensive than all other authors combined. The author shared the same view as Shinskey for decades and related to the solutions for dealing with disturbances at the input of processes with a significant process time constant. The solutions were aggressive action to provide tight control of a key process variable such as composition, pH, and temperature in columns, crystallizers, evaporators, neutralizers, and reactors. Level control was tight where residence time and material balance control was needed but loose where the objective was minimizing the manipulated changes in flow to downstream unit operations keeping the vessel from overflowing or running dry. Surge tank level control was always loose. Distillation receiver level control was loose or tight depending upon whether the level controller manipulated distillate flow or reflux flow, respectively. Tight distillate level control was needed for the changes in distillate flow by the column temperature controller to result in a corresponding reflux flow to affect the column. Tight control here also provided some internal reflux control that inherently compensated for upsets in overhead column temperature.

The basic concept is that variability does not disappear but is targeted to a part of the process or a frequency that least affects process efficiency and capacity. Process engineers sometimes want variability to disappear in both the process variability and manipulated variable and are often reluctant to turn over this responsibility to a control system particularly in a batch operation. Instead of fed batch control for composition or temperature, process engineers want to schedule flow changes.

Tight control means maximizing the transfer of variability from the process variable to the manipulated variable. For large industrial chemical plants, the raw material and utility systems (e.g., steam, cooling water, nitrogen) typically have large volumes and tight pressure control so that large and fast movement of a raw material or utility flow does not upset other users. This may not be the case as these chemical plants are pushed way beyond nameplate capacity or for pharmaceutical and other high value-added products with production rates of hundreds of kilograms instead of millions of kilograms per day.

More aggressive PID action by overshooting the final resting value is needed for tight concentration, pH, and temperature control of continuous and batch unit operations with large liquid volumes, tight pressure control of a single large gas volume with a large residence time, and level control.

Loose control means minimizing the transfer of variability from the process variable to the manipulated variability. This goal is stated in a more positive manner as maximizing the absorption of variability.

Moderated control means making the approach of the process variable to setpoint and the manipulated variable to the final resting value gradual with no overshoot. Moderated control is needed in the composition and temperature control of unit operations with recycle streams, heat integration, and no back mixing (e.g., hydrocarbon, gas, and oil operations), inline systems, extruders, and sheets. These processes typically have a moderate self-regulating response with a time constant to dead time ratio between 0.25 and 4 except for the cases of significant heat transfer lags.

In the last 20 years, IMC and Lambda control have become increasingly popular. Both IMC and Lambda control are based on disturbances arriving literally or effectively at the process output by either bypassing a large process time constant or passing through a small process time constant. Both of these approaches focus on pole-zero cancellation that has roots in servo-mechanism and aerospace control theory. This approach provides the opportunity for moderated control in moderate self-regulating processes. The tuning parameter is a closed loop time constant (time for a process variable to reach 63 percent of a setpoint after an initial delay in the setpoint response) called gamma and Lambda. For integrating processes, the Lambda parameter becomes an arrest time (time for a process variable to reach the peak error after the initial delay in the load response). The switch of the tuning parameter to an arrest time opened the door for a convergence of views.

Books and papers particularly those originating from universities largely shows disturbances entering the loop downstream of the process. The disturbances that enter directly into the measurement bypassing the process are in most cases extraneous. The best action is no action by the PID and in this respect these disturbances are effectively noise even though some can be quite slow (e.g., day to night temperature changes). Most of these disturbances arise from changes in ambient conditions (e.g., temperature, sun, and wind affecting filled system capillaries and load cells), installation (e.g., phase changes in impulse lines and pipeline movement on load cells), and operating conditions (e.g., absolute pressure change in differential pressure measurement and liquid velocity causing electrostatic potentials in high purity water streams or gas velocity kinetic energy impact on furnace probe total pressure measurement). Better sensor and transmitter technology and installation practices have eliminated many of these disturbances.

Changes in ambient conditions can affect process inputs. For example, a cold blue northerner rain storm can cause a sudden drop in column overhead temperature and a sudden drop in the suction temperature and rise in suction humidity for air compressors.

There are undoubtedly disturbances that enter into the process output that can be corrected by a change PID output and consequently a change in process input. For these cases, the correction is going to arrive late because of the dead times and time constants in the process. By only considering these types of disturbances the focus can be on setpoint response because a step change in the process variable is effectively the same as step change in setpoint for a PID on error structure most often shown or assumed in the control literature. The tuning must be moderated to prevent overshoot in the return to the existing setpoint in the disturbance response or overshoot in the approach to new setpoint in the setpoint response. There may also be an objective of minimizing abrupt movements in the PID output and overshoot of the final resting value of the PID output. The focus on a gradual return or approach to setpoint and final resting value is achieved by tuning methods such as IMC and Lambda tuning for self-regulating processes based on pole-zero cancellation in the frequency domain. These methods can also work well for disturbances in the process input besides the process output for control loops on unit operations where the primary time constant is small. These unit operations dominate in pulp and paper, the back end of food and drug, and the front end of oil, gas, and hydrocarbon processes. The key characteristic in these unit operations is the lack of a significant back mixed liquid volume or gas volume. The result is a primary time constant much less than the process dead time (small time constant to dead time ratio).

The dynamic compensation of a feedforward signal may not be needed or simplifies to a small lead time that cancels out the effect of the small primary time constant (lag time) in the process. If there is an appreciable lag in the disturbance path or feedforward measurement, that does not exist in the path of the feedback correction into the process, a feedforward lag time is set equal to the disturbance lag or feedforward measurement lag.

If Lambda is set equal to half the dead time and a switch to integrating process tuning rules is done for a time constant to dead time ratio of three, tight control with tuning settings similar to those originally developed by a large number of PID control consultants for the chemical industry. The methods converge for tight control. In practice, the minimum Lambda is one dead time and the switch is done for a ratio of four to ensure an over damped response even with gain scheduling or adaptive control to account for nonlinearities.

The author has turned to Lambda tuning seeing the value of objectives other than minimum peak or integrated error and a slow over damped response instead of a fast critically damped response that is perilously close to a underdamped response, the value of setting an arrest time for maximizing the absorption of variability, and the multitude of benefits in terms of faster test time, better conceptual understanding, and more effective tuning rules offered by a near integrator approximation.

The near-integrating methodology is consistent with what a PID sees and needs to do. In an integrating process, the response of process variable is a ramp after the dead time. For a near-integrating process, the process variable also ramps in the critical time frame of the PID response that is two dead times. In an integrating process the PID output must be driven past the final resting value for the process variable to approach the setpoint. If the PID output does not overshoot the final resting value, the best the PID can do is stop the excursion. The PID output has to temporarily go past the final resting value to change the sign of the ramp to bring the process variable back to setpoint. For near-integrating processes, PID output overshoot of the final resting value is almost as important. While the process variable will eventually reach setpoint without this overshoot, the approach will be incredibly slow.

8.4 DISTURBANCE TROUBLESHOOTING

The first flow that has significantly changed or started to oscillate is normally directly or indirectly associated with the root cause of a disturbance. If the original disturbance is temperature or concentration, the correction in the manipulated flow of a pressure, temperature, or concentration loop is the key to focusing on the right unit operation and loop. The most frequent unidentified disturbance is a change in raw material composition. Often this changes the utility or vent flow for temperature or pressure control, a reagent flow for pH control, and a recycle, reflux, or reactant flow for composition control. The deployment of wireless flow measurements can help track down disturbances, since the predominant inputs to processes are manipulated flows.

The most common readily preventable disturbances are feed rate changes by the operator. Often these are seen at the start of a shift as operators go for their “sweet spot”. Undersized surge tanks, equipment performance problems, equipment constraints, manual startups, and manual actions are conducive to setting up a continual transient in plant operation. The automation of all manual and corrective actions to make them repeatable and predictable coupled with plantwide flow feedforward control can move process diagnostics from the basic to the advanced level by looking at changes in the flow ratios as indicative of changes in raw materials, operating condition, utilities, and ambient conditions. These disturbances and front end production rate changes typically propagate downstream, so the furthest upstream loop with an oscillation is normally the culprit. However, when the flow is set to centrifuges and dryers, level control on slurry tanks or evaporators may manipulate a flow into the vessel. In this case, a change in back end production rate propagates upstream. Changes in recycle streams also end up going back to the origin upstream creating an integrating effect. However, since the disturbance is a readily observable flow change, the upset is easy to spot. In fact, the best way of finding the source of any disturbance is the first flow with a significant change on a trend chart of all of the PFD flows, achievable with wireless measurements and an expansive trend or trajectory visibility on an improved operator interface. The time range of the trends of the flows must be intelligently set to spot the initiating event.

The most troublesome disturbances are oscillations. The amplitude and period of the oscillations in the process variables and manipulated variables are the keys to classifying and tracking down the source of fast and slow oscillations. Whether an oscillation is considered to be fast or slow depends upon the oscillation period relative to the ultimate period. The following discussion of the sources of fast and slow oscillations is based on an ultimate period of about 40 seconds, which was the test case at the end of the chapter.

8.4.1 SOURCES OF FAST OSCILLATIONS

Fast oscillations are particularly insidious because the best thing a PID controller can do is ignore them. Action taken by PID controller can do more harm than good in terms of resonance, amplification, and perpetuation leading to increased process variability and premature valve failure. Here are some root causes, diagnostic methods, and fixes.

An oscillation is considered fast if it is less than twice the ultimate period. The ultimate period can be considered to be about four times the total loop dead time. An exception is the severely dead time dominant process where the ultimate period is about twice the dead time. Some sources of fast oscillations and examples of causes are listed.

In all cases, it is assumed the subject loop is stable.

1. Pressure fluctuations (e.g., oversized pressure regulation)
2. Bubble formation (e.g., flashing and heat increase to boiling mixture)
3. Bubble collapse (e.g., cavitation and cold liquid increase to boiling mixture)
4. Dissolved gases (e.g., spargers and reactions)
5. Poor mixing (e.g., insufficient back mixing—axial agitation)
6. Sloshing and vortexing (e.g., vessel feed entry and exit)

7. Flame instability (e.g., oversized burner and pressure fluctuations)
8. Hammer (e.g., sudden large movement of liquid valves)
9. Surge (e.g., sudden closure of gas valves)
10. Interactions (e.g., poor tuning of fast loops that affect subject loop)
11. Valve oscillations (e.g., poor valve positioner-booster tuning and too small an actuator size)
12. Burst of oscillations (e.g., slow secondary loop)
13. Decaying oscillations with period = ultimate period (e.g., subject loop PID gain too large)
14. Decaying oscillations with period = 0.6 × ultimate period (e.g., subject loop PID rate time too large)
15. Decaying oscillations with period = 1.6 × ultimate period (e.g., subject loop PID reset time too small)

The first 10 sources persist when the subject loop is put in manual.

If the oscillations are close to the ultimate period of the subject loop, the oscillations will grow from resonance when the subject loop is in automatic. If the oscillations are less than the subject loop dead time, resonance should not be an issue but amplification and perpetuation can occur. A high controller gain or rate time setting makes resonance, amplification, and perpetuation worse. A signal filter set just large enough to keep PID output fluctuations within the valve deadband and resolution limit may help. In general, signal filters should be minimized because they add dead time to the loop or even worse hide process excursions if they become the largest time constant in the loop. If the filter is large enough to become a secondary time constant, the performance of integrating and runaway processes are seriously degraded.

The correction for sources 1 through 7 involves improving equipment, control valve, and pressure regulator design and installation. Keeping the liquid pressure above the vapor pressure at the sensors is critical. The correction for sources 8 and 9 involve slowing down the closure of the offending valves taking into account the installed flow characteristic of these valves and improving pressure and surge control. The correction for source 10 is to slow down the tuning of the least important loop or add decoupling or go to MPC.

The correction for source 11 is to tune the positioner for a stable and smooth response with minimal overshoot for actuator and valve combination. If a booster is used on the positioner output, a bypass valve around the booster must be opened to prevent instability (e.g., very fast limit cycle). If fast oscillations occur near the closed position, make sure the actuator is sized to handle at least 150 percent of the anticipated largest pressure drop and largest friction stem and seating-sealing friction. The correction for solids causing an alternating loss and burst of flow is a valve design where solids cannot accumulate. The correction for source 12 is a faster secondary loop and external-reset feedback to prevent the primary loop output from changing faster than the secondary loop can respond. The correction for sources 13 through 15 is less proportional, integral, and derivative action in the subject loop.

8.4.2 SOURCES OF SLOW OSCILLATIONS

Slow oscillations can be difficult to recognize especially when the period is beyond the typical time frame of the trend chart or there are intervening disturbances or recycle. Slow oscillations can be more detrimental to product quality because the large period means the amplitude is less attenuated by intervening volumes.

An oscillation is considered slow if it is greater than twice the ultimate period. The ultimate period can be considered to be about four times the total loop dead time. An exception is the severely dead time dominant process where the ultimate period is about twice the dead time. Some sources of slow oscillations and examples of root causes:

1. Batch operations (e.g., batch reactor and centrifuge)
2. Defrosting (e.g., crystallization)
3. Regeneration (e.g., demineralized water units and catalyst beds)
4. Ambient conditions (e.g., day—night, sun—shade, rain—snow)
5. Dead band (e.g., valve backlash)
6. Resolution limit (e.g., valve stick-slip)
7. Threshold sensitivity limit (e.g., valve stick-slip)
8. Split range discontinuity (e.g., heating and cooling)
9. Low valve gain (flat installed characteristic)
10. Low PID gain or reset time (e.g., integrating or runaway process)

The oscillations for items 1 through 4 persist when the subject loop is put in manual.

A particularly disruptive batch operation is the centrifuge because the cycle times are much more frequent than many other batch operations, such as batch reactors and crystallizers. Surge tank level controllers that have batch feeds should have the level controller arrest time set to spread the upset over the maximum possible range of level. See Equations 1.21a through 1.22k in Chapter 1 to compute the arrest time that maximizes the absorption of variability. Note that simply decreasing the PID gain will cause the slow rolling oscillations noted in item 10.

The period of oscillation for items 1 through 3 is the time between successive end points of batch operation and start points of defrost and regeneration. The start points of defrost are rather unpredictable if done only when crystal growth excessively limits heat transfer. In any case, the oscillations should be recognizable when included on a trend chart for the subject loop with a suitable time scale. Often the effect of rail or truck shipments is overlooked. If there is no agitator or recirculation stream in the raw material storage tank, concentration differences exist as layers with only a slight degree of mixing by mass transfer from areas of high to low concentration or settling by gravity. The layering makes the upset more discontinuous.

The effects for item 4 can be quite varied. For previous generations of transmitters, you could see the effect of weather and day to night in the change in controller output. Most smart transmitters have significantly reduced the effect of ambient temperature changes. Load cells and capillary systems may still be affected by sun and wind. Air compressor suction flow density and moisture will change with temperature and humidity. A rain storm can drastically lower the temperature and vapor flow in the upper part of a column and reflux condenser for a reactor or column. Adding ambient conditions to the trend chart can greatly help in spotting affected components realizing that variability in the measurement is transferred to the PID output.

Items 5, 6, 7, and 8 can cause a limit cycle (equal amplitude oscillation). The amplitude may not appear to be equal due to disturbances, noise, and changes in the backlash and stick-slip with operating point. A sure sign of these items is if the period increases as the PID gain is decreased. The oscillations associated with item 8 will traverse the split range point where the discontinuity is the greatest. The discontinuity and oscillations are larger if there is a change in phase (e.g., water to steam) or if tight shut-off control valves are used (e.g., high seating or sealing friction). Whereas a limit cycle will occur for items 6 through 8 if there is just one or more integrator, item 5 requires two or more integrators somewhere in the process (e.g., level or batch operation), controller (e.g., PI or PID), or positioner. Integral deadband in the PID or positioner can turn off integral action when the controlled variable within the deadband at setpoint to kill the limit cycle if the deadband is set to be greater than the limit cycle amplitude. External reset feedback with a fast readback of actual valve position can inherently stop the limit cycle with no adjustment required. Similarly, the enhanced PID for wireless will kill a limit cycle if a threshold sensitivity limit screens out unnecessary updates from noise.

Unnecessary excursions across the split range point that create item 8 limit cycles and decrease process efficiency (energy use and yield), can be reduced by the addition of a setpoint rate limit in the direction of the split range point. External reset feedback is turned on to make sure the PID has the necessary patience and does not try to out run the set point. Safety factors must be considered. For example, the transition from cooling to heating might be slowed down but not vice versa to prevent the start of a runaway response (acceleration in increase of temperature) for an exothermic reactor.

Item 9 causes a wandering of the PID output from a flat characteristic most often associated with rotary valve openings larger than 50 degrees and low valve to system drop pressure ratios causing a running out of valve capacity.

Item 10 may be difficult to detect because the period of oscillation is so great. For large distillation columns, the trend chart needs to cover several shifts of operation to see the period. Feed disturbances make will change the oscillation to the point it is no longer recognizable. The column may appear to be oscillating with no particular pattern. The product of the PID gain and reset time should be greater than twice the inverse of the integrating process gain. A generally conservative fix is to simply increase the reset time by one or two orders of magnitude.

To see the oscillations and recognize the period and source, it is important to set the trend chart PV and PID output scale small enough to see small oscillations but large enough so that noise does not hide the pattern. Compression must be removed for small oscillations. Intelligent filtering and velocity limiting can help screen out noise for PID control but the raw fast measurement should be retained and made available for trending and analysis.

The time scale must be large enough to cover several periods. The oscillations are usually more apparent in the PID output. Loops affected by the culprit loop for items 5 through 10 will show an oscillation of the same period. Simply inverting the measurement of the culprit on the same chart as the affected measurement can provide the visual recognition of the obvious time correlation.

Culprit loops are typically on the same unit operation or upstream. Recycle and heat integration can result in downstream loops either initiating or aggravating the problem. Regardless, the oscillations should stop when the culprit loop is put in manual. The most frequent source of oscillations is a level loop on a surge tank with too low a PID gain or reset time. The solution is to compute the arrest time as previously mentioned to spread the change in flows into the surge tank over the entire level range.

A power spectrum analyzer can rapidly point you to the culprit by indicating which loops are experiencing significant peaks in the power at the same frequency. The first step is to get the data for loops in automatic into the power spectrum analyzer. The data gathering must be done quickly enough to prevent aliasing. For most chemical processes, it is sufficient to use data from a historian with no compression and an update time of one second. For liquid/polymer pressure and surge control and sheets (webs), you may need to store the data using a device with an exceptionally fast scan time (50 milliseconds or less) that is directly connected to the process variable input or PID output terminals.

Data storage is so cheap, old rules about compression, and exception and periodic reporting should be going by the way side. Unfortunately, old practices may be hard to undo. It is important to realize that the role of Information Technology (IT) is to meet the needs of the user and not vice versa.

Statistical tools with PCA can reveal correlations. These tools are particularly valuable as the number of variables increase. A drill down into the principal components can show what inputs are most correlated with the oscillation observed. Process and automation system knowledge is then used to verify if the input output correlation is a cause and effect or a coincidence or result of output correlations.

8.5 DISTURBANCE MITIGATION

The consequences of disturbances can be moderated and almost totally alleviated by reducing the size or speed of the disturbance. Here are some suggestions:

1. Tune utility pressure and temperature loops for tight control.
2. Minimize the dead band and resolution limits in final control elements.
3. Replace sequenced or manual actions with PID control loops.
4. Tune loops for an over damped response.
5. Tune the composition, pH, and temperature loops that most directly affect process efficiency and capacity for tight control.
6. Maximize variability absorption on surge tank distillate receiver level control where discharge flow is manipulated to minimize disturbance to downstream units.
7. Treat processes with a time constant to dead time ratio greater than four as near integrators.
8. Use a setpoint lead-lag or a 2DOF structure to prevent abrupt changes to manipulated variable.
9. Use setpoint rate limits on AO blocks and secondary flow loop PID blocks whose output is setting a flow. Use external reset feedback of actual valve position, drive speed, or flow to prevent the output of the PID manipulating the setpoint from changing faster than the block setpoint rate limits will allow.
10. Use external reset feedback to the upper loop of a lower loop in cascade control.
11. Use an enhanced PID for wireless transmitters and at-line or offline analyzers.
12. Increase the closed loop time constant of least important and the slowest self-regulating loops and decrease the time constant of the most important and the fastest self-regulating loop to minimize interaction.
13. Set the closed loop time constant of self-regulating loops to maximize coordination so that corrective actions arrive at the same point in the process at the same time (e.g., blend composition and reactant stoichiometry control).
14. Add feedforward control with dynamic compensation for load disturbances.

8.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in load flow at the process input (ΔF_L). Table 8.1 summarizes the test conditions.

The test cases use aggressive tuning settings computed with the shortcut method to maximize disturbance rejection. Since these settings are right on the edge of causing an oscillation, the implied dead time is close to the original loop dead time. These settings are modified in some tests to show the effect of changes in the PID gain and reset time. The process used in the tests had a moderate self-regulating response to avoid the attenuating effect of a large primary process time constant from dominating the view.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

Figure 8.2 shows the effect of an increasing load disturbance time constant (disturbance lag) on the closed loop response. As the disturbance lag increases the peak error is reduced but the integrated error (area between plot of the process variable and setpoint) is about the same because the return to setpoint is slower. The PID output has to continue to change after the peak error to catch up with the exponentially growing disturbance (first order exponential response of open loop error). When the time after the peak equals the disturbance lag, the exponential response reaches an inflection point and slows down but the error is so small there that integral mode action is small and the approach stays slow.

Figure 8.3 shows the effect of tuning on the largest lag from the previous test case. The process variable scale range is smaller to show the PID response better. An increase in PID gain reduces the peak error which makes the starting point of the approach closer to the setpoint. A decrease in reset time helps integral action catch up with the increase load d isturbance. The best test case is a 50 percent increase in gain and 25 percent decrease in reset time. The reset time could have been decreased by 50 percent to give the fastest return to setpoint without causing an oscillation. The danger here is that the loop will oscillate severely for a fast disturbance. Unless all fast disturbances are small and all nonlinearities are addressed by the scheduling of tuning settings, more aggressive tuning is theoretical but not a practical solution. For setpoint changes a setpoint lead-lag or 2DOF structure must be used to prevent overshoot and oscillations for the aggressive settings.

Table 8.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
8.2	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 20\%$	Slow load lags
8.3	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 20\%$	Slow load tuning
8.4a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 0\%$	Fast load oscillations
8.5a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 0\%$	Slow load oscillations

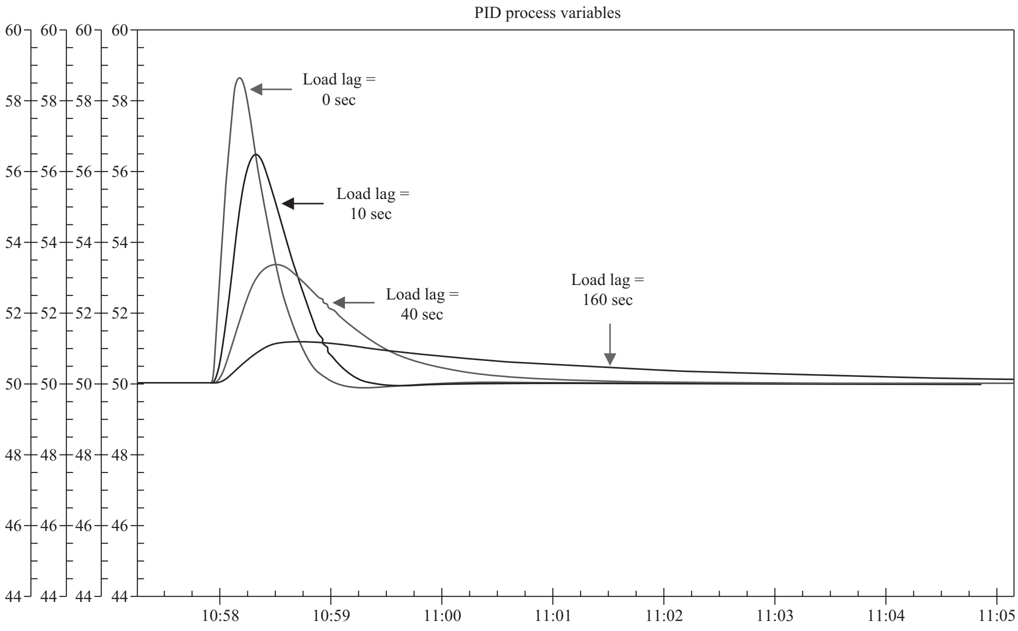


Figure 8.2. Effect of load *lags* on PV for moderate self-regulating process.

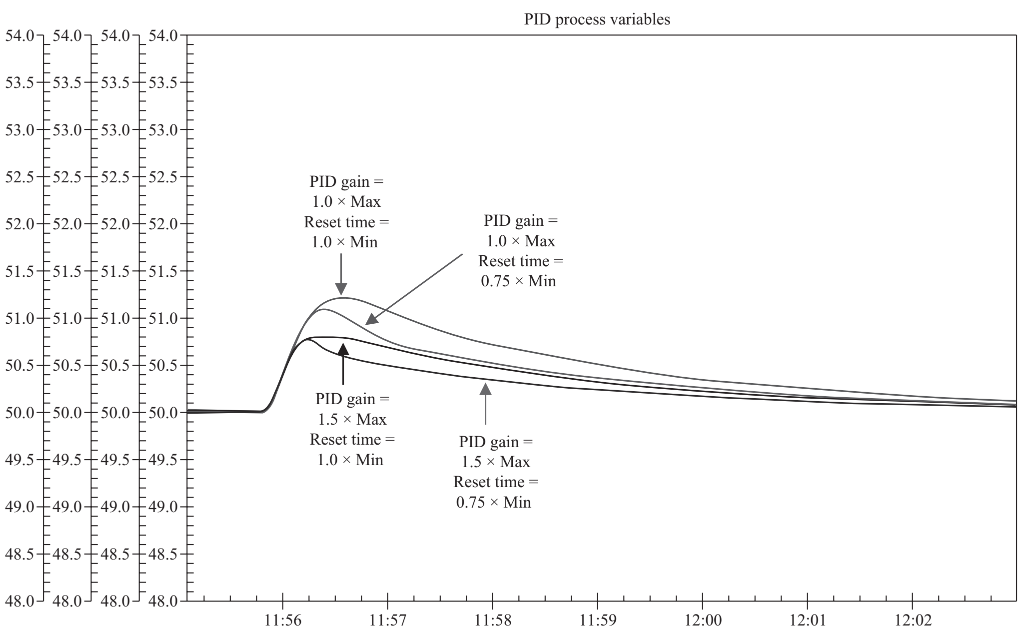


Figure 8.3. Effect of *faster tuning* for slow load lag on PV for moderate self-regulating process.

Figures 8.4a and 8.4b show the effect of tuning settings on the process variable and PID output, respectively, for a fast load oscillations. A PID in manual plot eliminates the feedback correction so that only the effect of the attenuation by the process time constant is seen, which is still significant for this moderate self-regulating process. For oscillation periods less than half the ultimate period, the oscillations in the process variable for different tuning are the same as the oscillation for the loop in manual. The most aggressive tuning cause oscillations in the PID output that will wear out the valve packing without any noticeable effect on reducing the process variable amplitude. For oscillation periods between half and twice the ultimate period, the aggressive tuning causes amplification from resonance where the feedback correction oscillation gets in phase with the load oscillation. When the load oscillation period becomes greater than twice the ultimate period the most aggressive tuning settings noticeably decreases the amplitude of the process variable oscillation compared to the PID in manual oscillation. The oscillation in the PID output is larger for the most aggressive settings.

Figures 8.5a and 8.5b show the effect of tuning settings on the process variable and PID output, respectively, for a slow load oscillations. Here we clearly see the benefit of the most aggressive tuning settings. The benefit will not be so clear for oscillations with a period that is incredibly long (e.g., 1,000 times the ultimate period), which might be the case for day to night effects on operating conditions. Equation 8.1a can be used to show the effect of a very slow load oscillation approximated as a load time constant equal to about one-eighth of the load oscillation period.

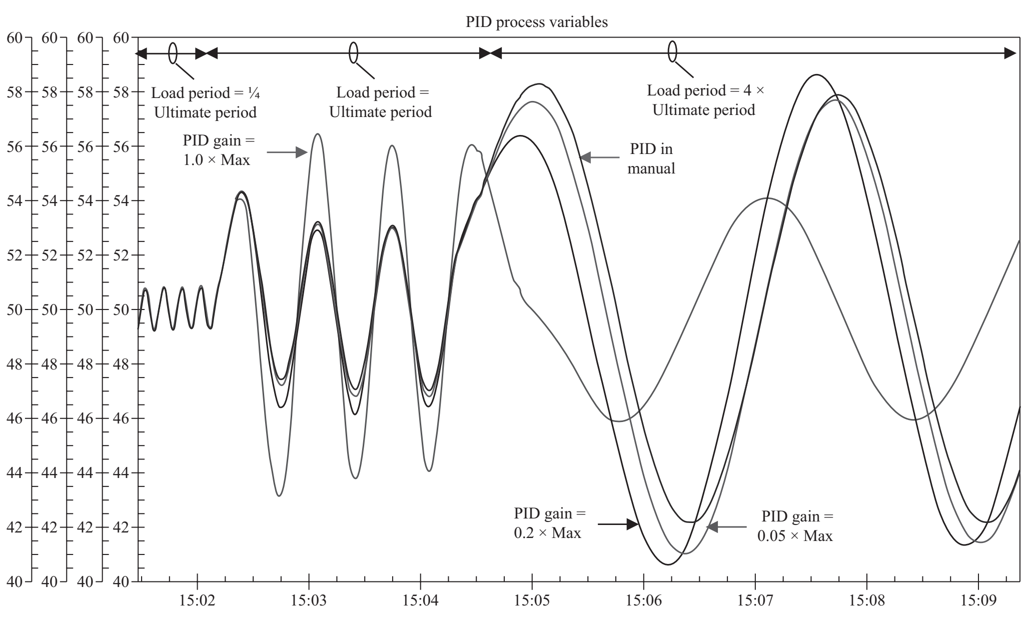


Figure 8.4a. Effect of *fast load oscillation* period and PID gain on *PV* for moderate self-regulating process.

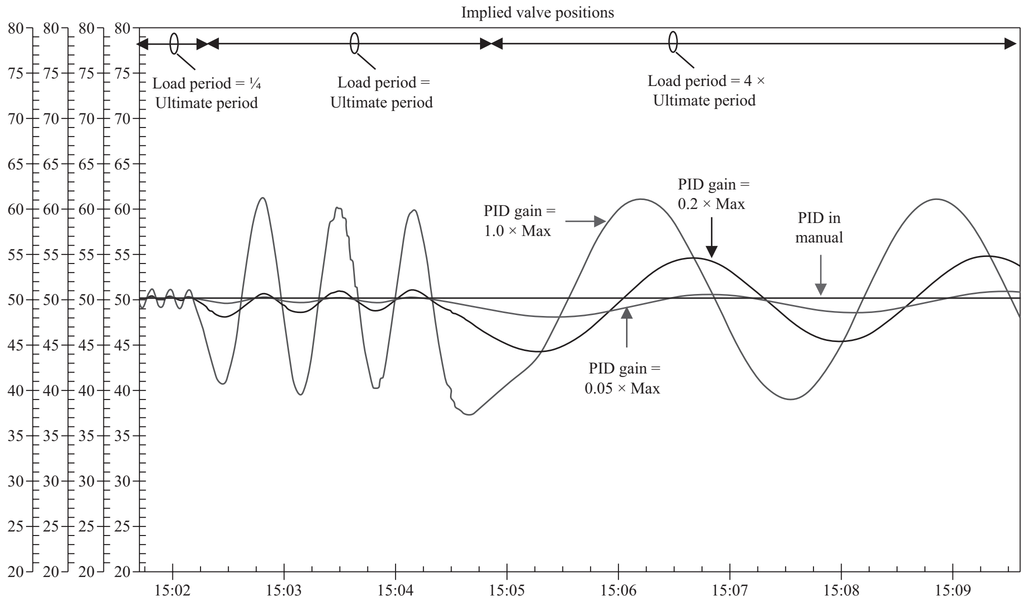


Figure 8.4b. Effect of *fast load oscillation* period and PID Gain on *valve* for moderate self-regulating process.

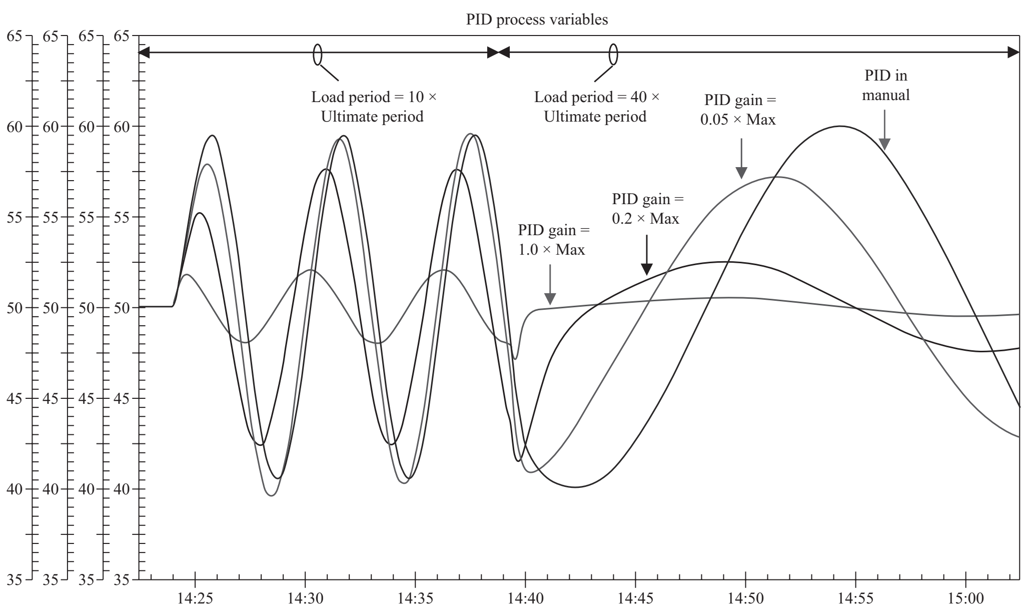


Figure 8.5a. Effect of *slow load oscillation* period and PID gain on *PV* for moderate self-regulating process.

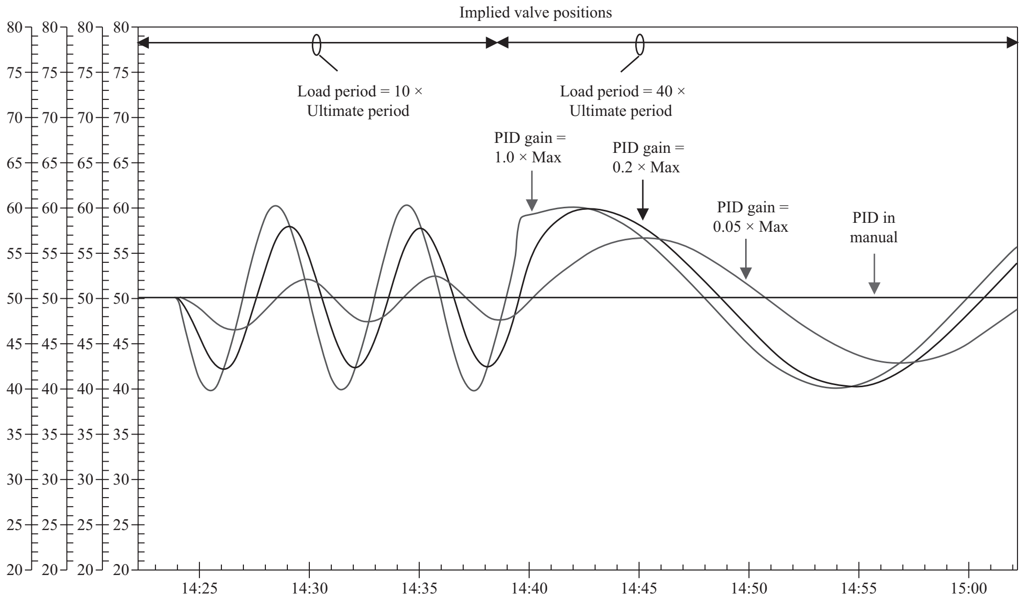


Figure 8.5b. Effect of *slow load oscillation* period and PID gain on *valve* for moderate self-regulating process.

KEY POINTS

1. A large primary time constant is beneficial and detrimental for disturbances at the process input and output, respectively.
2. A large disturbance time constant or slow rate of change minimizes peak error.
3. A disturbance dead time only affects feedforward dynamic compensation.
4. The use of a near integrator approximation eliminates most of the disagreement and ineffectiveness of tuning methods and provides much faster test times for loops with large process time constants.
5. For oscillation periods less than half the ultimate period feedback control does nothing except wear out valves and upset utility and raw material systems.
6. For oscillation periods between half and twice the ultimate period, PID control does more harm than good from resonance.
7. For oscillation periods much greater than the ultimate period, PID control can totally transfer the oscillation amplitude in the process variable to the PID output.
8. Large disturbance oscillations should be mitigated even if they are slow.

CHAPTER 9

Effect of Nonlinearities

9.1 INTRODUCTION

Nonlinearity is not just a change in the open loop gain. Any change in dynamics or deviation from a response describable by a fixed open loop gain, dead time, and time constants is considered to make the loop nonlinear.

The dynamics in 99 percent of the control loops in the process industry are not constant. We have seen from previous chapters how equipment, process, valve, variable speed drive (VSD), measurement design, installation, and operating conditions affect the overall loop dynamics. There is always a tradeoff between control loop performance and robustness, which is the ability to tolerate changes in dynamics without causing excessive oscillations. The tuning that gives the minimum peak error and integrated error and the fastest setpoint response is extremely sensitive to nonlinearities. The focus in the literature has been on linear systems, exact identification of dynamics, and tuning that makes the test case trend chart and performance metrics appear to be the best. Considerable gamesmanship and pride is at play creating at times bitter disagreement on the tuning methods that achieve these results. Missing is the bigger picture of being able to meet other objectives such as coordination of loop responses and to deal with unknowns and non-ideal behavior, such as nonlinearities. The automation profession could benefit from seeing beyond the personal investment in a tuning technique and seeing how to improve the loop dynamics and the use of the proportional-integral-derivative (PID) as knowledge and objectives change. Thus, while this book sees the Lambda tuning method as being able to address the issues in industrial processes, this is not to say other tuning methods are wrong or a roadblock. If the view is not closed to possibilities and different tuning methods, the PID features and tuning software can be better utilized, loop analysis made easier, system dynamics improved, and tuning settings modified to better meet application requirements.

9.1.1 PERSPECTIVE

A linear control loop has a constant dead time, constant primary and secondary time constants, a constant open loop steady state gain for self-regulating processes, and a constant open loop integrating process gain for near or true integrating and runaway processes. The common control loop that satisfies these criteria is a mass flow loop with a Coriolis flow meter manipulating a VSD where the static head, signal input card resolution, and setup dead band and the setpoint rate limit are negligible.

The open loop gain is the product of the manipulated variable gain, process gain, and measurement gain. For nonlinear components, the plot of the output versus the input to the component is not a straight line but a curve. The slope of the line connecting the old and new operating point on a curve changes with operating point creating what is termed *operating point nonlinearity*. For a disturbance, the new operating point for the controlled variable is temporary but for a manipulated variable the new point is the new norm. For a setpoint change, the new operating point becomes the norm for the process variable (PV) but may be just temporary for the manipulated variable in an integrating or runaway process. To understand this, you need to remember what drives the response of the different types of processes. For an integrating or runaway process, the PV is driven to a new setpoint by a temporary unbalance between the load and the manipulated variable. For a self-regulating process, the PV is driven to a new setpoint by a new steady state value of the PID output (termed final resting value).

For small changes, the gain is the slope of the curve. Not commonly understood or discussed in the literature is that gain identified by manual or automatic tests is not the local slope of the curve but is as shown in Figure 9.1, the slope of the line connecting the new and old points on the curve for the test step size. For some large disturbances or large setpoint changes, the large distance between the old and the new operating points moderates the gain nonlinearity. Consequently the nonlinearity is not as bad as depicted by local changes in slope of the curve. Additionally, the use of even imperfect signal characterization offers a greater than realized advantage in reducing the dependence of gain on step size, freeing an auto tuner or adaptive tuner to focus on other nonlinearities.

For a primary loop manipulating the setpoint of a secondary loop, the manipulated variable gain is simply the secondary setpoint scale span divided by 100 percent and thus linear. For a control loop manipulating a flow, the manipulated variable gain is the slope of a line connecting the previous operating flow when at setpoint and the current flow on a plot of the installed characteristic of the control valve or VSD. The plot is a straight line rather than curve giving a constant slope for the installed characteristic of valve with an inherent linear characteristic and a constant pressure drop and for a VSD where the destination static head is negligible (static head is much less than the pump or fan discharge head at the lowest speed).

The steady state process gain for liquid composition, pH, and temperature is identified on a plot of the PV versus the ratio of the manipulated variable to the feed variable. The X axis indicates there is a flow ratio gain (gain component included in the process gain) that is inversely proportional to feed flow.

The open loop integrating process gain (product of the valve or VSD gain, process gain, and measurement gain) is the rate of change (%/sec) in the PV divided by the change in the controller output (%). The percent units cancel out leading to the commonly stated units of 1/sec. The time units in the open loop integrating process gain occurs from the time units of the flow change not being canceled out by time units, in the process gain, as was the case for self-regulating processes. For liquid level, the process gain is inversely proportional to the product of the cross sectional area and density. The process gain for gas pressure is proportional to the absolute temperature and inversely proportional to the volume. The process gain for batch temperature and composition is inversely proportional to the liquid mass in the batch. When batch temperature is manipulating coil or jacket temperature, the process gain is also proportional to the heat transfer area covered by liquid and hence is a function of batch level, which all or partially cancels out the effect of change in mass.

The measurement gain is simply 100 percent divided by the PV measurement span since modern measurements have linearization integrated into the transmitter (e.g., square root extraction for differential head meters and sensor matching for thermocouples and resistance temperature detectors).

Control valve or VSD time constants and dead time are inversely proportional to the rate of change of the PID output for scan rate, deadband, threshold sensitivity, resolution, setpoint rate, and slew rate limits. A change in the PID output less than these limits will not introduce the dead time. A change in structure to proportional or derivative action on PV rather than error slows down the rate of change of the controller output that in turn increases valve and VSD time constants and dead time. Detuning the PID also slows down valve and VSD dynamics. The result is a larger response time and peak error from disturbances.

The process time constants for composition, pH, and temperature control for well-mixed liquid volumes is basically the residence time (volume/flow) and is thus inversely proportional to feed rate. The process dead time for these operations from mixing is relatively fixed and small and can be estimated as half the turnover time for a vessel with baffles and good geometry for mixing. The turnover time is the time to completely turnover the vessel contents and is estimated as the volume divided by the sum of agitation and recirculation flow rates. The process dead time from injection of reactants for composition control is usually negligible. Unfortunately, the process dead time from injection of reagents for pH control is large because the reagent flow is small. The dead time is a transportation delay that is the injection piping and dip tube volume divided by the reagent flow. This dead time is consequently inversely proportional to feed flow.

The process time constant or integrating process gain of gas pressure, composition, and temperature response depends upon direction particularly in batch processes. The difference in time constant can be an order of magnitude for bioreactors. The time constant for an increase in bioreactor pressure by air addition is often smaller than for a decrease in pressure by gas venting. The time constant for an increase in glucose concentration from glucose addition is much smaller than for a decrease in glucose concentration by glucose consumption by cells. The time constant for an increase in bioreactor temperature by heat addition is much smaller than the time constant for a decrease in bioreactor temperature by ambient cooling and evaporation. Often these loops are considered as one sided processes (processes with a response in one direction) necessitating the use of a PID structure without integral action to prevent overshoot.

The process time constant is less than 20 percent of the residence time of unit operations that are essentially plug flow. Radial mixing may exist but there is negligible back mixing. Nearly all of the residence time is a transportation delay and hence a dead time. Examples of plug flow volumes are furnaces, catalyst bed gas reactors, kilns, static mixers, extruders, sheet lines, spin lines, and web lines.

The measurement time constants for dew point, pH and temperature sensors depend upon operating conditions and in particular, process velocity. Measurement dead time from a transportation delay is inversely proportional to recirculation and sample flow rates.

One of the most difficult nonlinearities to deal with is inverse response where the initial reaction of the PV is in the opposite direction of the final response. Most notably, this inverse response occurs from shrink when cool feed water is increased and swell when cool water is decreased to boiler drums due to bubble collapse and formation. Shrink and swell also occurs for a decrease and increase in firing rate, respectively. In distillation columns, shrink and swell occurs when the bottom level controller decreases and increases reboiler steam flow, respectively. Inverse response can also occur when cool air entering the furnace is increased before the fuel is increased for pressure or temperature control. Preheating of feedwater and air flow decreases the inverse response in boiler drums and furnaces. An inverse response is often the result of primary and secondary processes in parallel where the secondary process is faster with an opposite sign. The secondary process causing the inverse response normally has a nonlinear gain and the combination of the two processes does not lend itself to a linear first or second order plus dead time approximation as a single process.

9.1.2 OVERVIEW

For composition and temperature control of continuous operations on well-mixed liquid volumes, PID tuning does not change much with feed rate except for the effect of a nonlinear installed flow characteristic. For a well-mixed (completely back mixed) volume, nearly all of the residence time is a process time constant (mixing and transportation delay is negligible). The process dead time and hence the reset time and rate time is relatively constant. The effect on controller gain from an increase in process gain is canceled out by the increase in process time constant as the feed rate decreases. For these well-mixed volumes, the PID gain does not need to be decreased at low production rates if the nonlinear flow characteristic is excluded by a PID manipulating a secondary flow controller or a linear trim control valve with a high valve to system pressure drop ratio or a VSD with low static head.

For pH control, the increase in reagent injection delay at low production rates results in a need to decrease the PID gain and increase the PID reset time at low production rates. The effect can be minimized by reducing the piping volume between the control valve and the injection point and eliminating dip tube volume by injecting the reagent into high flow feed and recirculation streams.

For composition, pH, and temperature control of continuous operations on plug flow volumes, such as the tube side of a heat exchanger, the process gain and dead time both increase as the feed is decreased. The PID gain must be decreased and the reset time increased at low production rates. A nonlinear installed flow characteristic will enhance this effect of production rate.

For batch operations, the integrating process gain is largest for composition control at the start of the batch when liquid level (liquid volume) is the lowest. For temperature control, the integrating process gain is also proportional to the heat transfer coefficient area. Since the coil or jacket area covered by process fluid is proportional to level, the integrating process gain stays about the same when temperature is controlled by manipulating the coil or jacket temperature. If the jacket also covers the bottom of the vessel, for low liquid levels the integrating process gain changes with level because the relationship between level and volume in the heel is nonlinear.

The fact that the process gain for operating point process nonlinearities, such as pH, varies with the slope of the line connecting the new and old operating points makes gain identification problematic. The size of the process gain continually changes with size of the step change in the manipulated flow. The step size for tests should be the size of the PID output change for the load and setpoint change of greatest interest.

Signal characterization where the PV is translated from pH to the abscissa of the titration curve and scales 0 to 100 percent reagent demand inherently eliminates the nonlinearity. Even if characterization is not accurate, it frees up an adaptive controller to compensate for unknowns or for the effects of changes in production rate.

For inverse response, a smarter feedforward may be warranted that reduces the initial reaction of the process in the wrong direction. This has turned out to be particularly important on boiler drum level where the shrink or swell can cause a low or high drum level trip, respectively. The additional feedforward to decrease the inverse response from transient changes in phase, decays out leaving the traditional feedforward based on material and energy balances. The sum of the two feed forwards compensates for the inverse response that is in itself the effect of two different process responses. The first process response is temporary and has the opposite sign of the final process response.

9.1.3 RECOMMENDATIONS

1. Use a flow loop and cascade control to isolate valve and VSD nonlinearities from a process loop such as composition and temperature.
2. When a flow loop is not feasible possibly due to rangeability limits, schedule tuning settings preferably with an adaptive tuner as function of PID output to account for the changes in the valve or VSD gain from a nonlinear installed flow characteristic. Scheduling is a term used to describe the computation and consequential setting of tuning settings as a function of operating conditions.
3. Schedule tuning settings preferably with an adaptive tuner as function of flow rate to account for the changes in transportation delays.
4. For changes in a primary process time constant with direction in loops tuned with self-regulating process tuning rules, schedule the reset time based on direction.
5. Schedule Lambda based on direction for changes in the ramp rate of the PV with direction for processes tuned with integrating process tuning rules.
6. Use signal characterization to reduce process operating point nonlinearities and to free up an adaptive tuner to handle unknowns and production rate nonlinearities.
7. To reduce the valve and VSD 86 percent response time (time to reach 86 percent of final response) for small changes in PID output, decrease the deadband, resolution limit, and threshold sensitivity limit and speed up valve slewing rate and VSD setpoint rate limits.
8. To decrease the dead time from the valve or VSD for small setpoint changes, use a structure that has some proportional action on error or use a setpoint lead-lag.
9. To reduce the effect of inverse response, increase the Lambda to at least three dead times. If tighter control is needed, add a smart feedforward that provides an initial correction in the opposite direction of the final correction that decays out before the traditional feedforward based on material and energy balances takes effect.
10. For changes in open loop gain, dead time, and time constants, that cannot be compensated by signal characterization, scheduling of tuning settings or adaptive control, estimate the equivalent gain margin required. Decrease the PID gain (increase Lambda) so the gain margin available (ratio of the ultimate gain to the current PID) is greater than the gain margin required.

9.2 VARIABLE GAIN

The gain of a nonlinear loop element (e.g., process or valve) is the slope of the line drawn between the new and old point on a plot of element output versus input. For small excursions the slope of the line is the slope of the curve at the operating point which leads to the general classification as an operating point nonlinearity.

The reality of the process or valve gain being the slope of a line connecting new and old points poses additional difficulties in gain scheduling and adaptive control; the process or valve gain is not just the slope at each point on the curve but the slope of the line connecting new and old points and hence a function of step size. On the plus side, this reality has a beneficial moderating effect on the nonlinearity. The slope on a titration curve for a strong acid and strong base system can change by factor of ten for each pH unit deviation from setpoint. For this extreme case, the slope at pH 7 is 10^6 times larger than the slope at pH 1. Fortunately, the controller does not see the full extent of the nonlinearity unless the changes in the pH setpoint cover the entire range of the titration curve and tight control is needed at each point. While a lower loop in a pH to pH cascade could see nearly the entire range, tight control in the lower loop is not so important and most titration curves are moderated by the presence of a weak acid or weak base.

Figure 9.1 is the *zoomed-in* view of the neutral region of a titration curve for a weak acid and weak base so that effect of step size can be graphically visualized. For a true strong acid and strong base, the line would look vertical despite the factor of ten changes in slope due to graphical resolution limitations. For this extreme case the slope of the line for various pH excursions outside for a fixed setpoint translates to a change in gain that is orders of magnitude smaller than would be seen from plot curvature. Figure 9.1 for a much less difficult titration curve shows

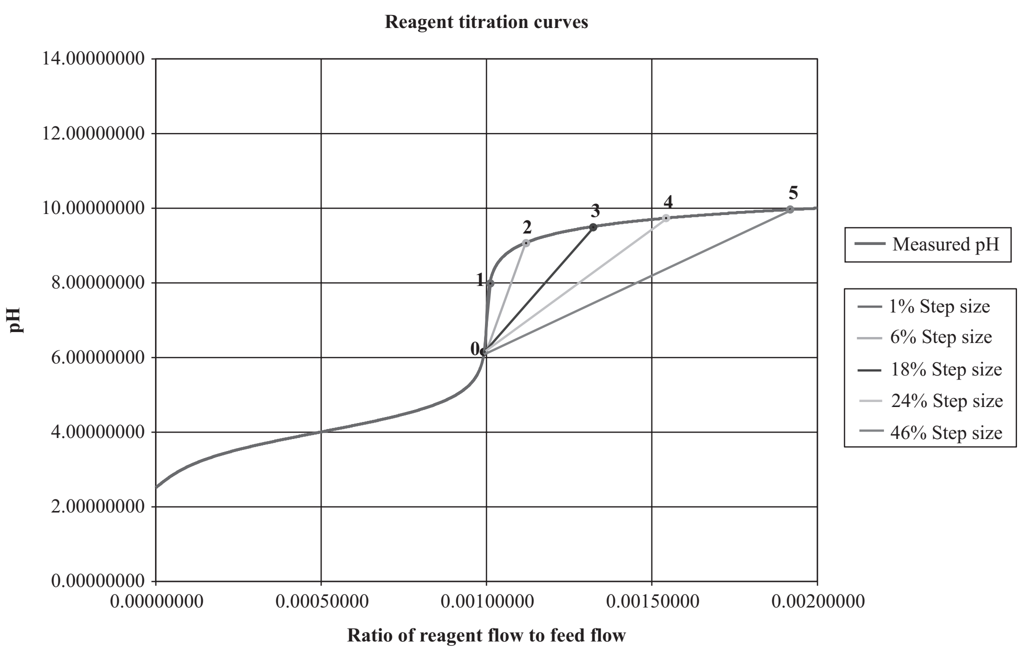


Figure 9.1. Effect of step size on process gain identified (gain is slope of line connecting old and new points).

how the pH process gain identified (the slope of the line connecting original operating point 0 with new points 1, 2, 3, 4, and 5) changes significantly for the five different step sizes and only matches the local slope of the curve for a small step (e.g., <1 percent).

The operating point nonlinearity seen is generally greater for manipulated variables than for PVs (controlled variables). For PVs, the normal operating point (old point for a test or an upset) is the setpoint. For manipulated variables, the normal operating point is determined by the PV setpoint and the load. For loops with significant changes in setpoint or load, the nonlinearity seen is more extensive.

The process gain of a composition, conductivity, pH, and temperature loop is computed on a plot of the PV versus the ratio of manipulated flow to feed flow. This ratio is often not shown and the X axis is manipulated flow for an assumed feed flow. As noted in Chapter 1, the more inclusive labeling of the X axis as a ratio leads to the identification of an inverse relationship of the process gain to feed flow via the flow ratio gain. The consequences in terms of nonlinearity are quite severe in that the process gain approaches infinity as the feed flow approaches zero. For volumes with significant transportation delays such as plug flow reactors, jackets, and coils, the process dead time also greatly increases as the feed flow decreases. The combination of a large increase in process gain and dead time at low production rates leads to instability.

The inherent equal percent flow characteristic creates a valve gain that is proportional to flow. The use of an equal percentage trim on a composition, conductivity, pH, or temperature control loop helps compensate for the process gain. However, the installed flow characteristic deviation from the theoretical equal percentage characteristic from the sizing of the valve, and changes in valve operating conditions introduces errors into the compensation that are unknown.

9.2.1 CASCADE CONTROL

The author prefers the use of a lower (secondary) flow loop and cascade control instead of having the primary PV directly manipulate the valve except for gas pressure control. The secondary flow loop corrects for flow upsets that generally originate as upstream or downstream pressure changes and isolates the nonlinearity of the control valve from the upper loop. The flow loop also enables flow feedforward where the flow ratio that corresponds to the desired operating point (composition, conductivity, pH, or temperature setpoint), is set by the operator and corrected by the upper control loop. Oscillations in the flow loop do not show up as variability in the PV of a well-mixed volume because of the attenuation of the relatively fast oscillations by the large process primary time constant. However, for low flow rates the poor signal to noise ratio may necessitate going to computed flow measurement based on valve position or pump speed and the installed flow characteristic.

If a secondary flow loop is not practical, the installed flow characteristic based on piping design is computed and signal characterization and adaptive control in the distributed control system is used on the primary loop. The valve gain is the slope of the line connecting the new and old operating point on the installed flow characteristic. If a secondary flow loop is used and tight flow control is needed, signal characterization and adaptive control may be used here as well.

The use of a lower (secondary) coil or jacket temperature loop isolates the upper (primary) vessel temperature loop from the nonlinearities of the curve and the inverse relationship to flow. The open loop steady state gain for the vessel temperature is nearly linear and mostly dependent upon the scale ranges of the secondary and primary loops as per Equation 1.3j if the effect of the change in sensible heat input from feeds and if the reaction heat input with temperature is negligible. The open loop integrating process gain for batch processes is also quite linear. Contrary to popular opinion, changes in heat transfer area covered by liquid do not normally affect the PID tuning because the change in volume changes the process time constant by the same amount that the change in heat transfer area changes the process gain (see Appendix F).

9.2.2 REVERSALS OF PROCESS SIGN

For conductivity there is normally at least one peak in the curve of the plot of conductivity versus concentration if the X axis covers the entire 0 to 100 percent concentration range for the salt ion, acid, or base. The process gain changes sign if the operating point crosses a peak. Conductivity control is only feasible if operation is limited to one side of a peak unless special intelligence is added. In theory, a change in process sign could be detected to locate the peak and switch the PID control action gain accordingly. More feasible is the translation of the PV to a rate of change of conductivity with respect to a rate of change of the manipulated variable. This PV would continuously decrease from a maximum positive value to a minimum negative value as the concentration increases. The PV would be positive to left of the peak, zero at the peak, and negative to the right of the peak. The rate of change calculation would use a dead time block on both the conductivity and manipulated flow to provide a good signal to noise ratio. The same interface considerations exist that are discussed for pH signal characterization.

9.2.3 SIGNAL CHARACTERIZATION

Signal characterization is a viable option for a well-defined installed flow characteristic curve and PV versus manipulated flow ratio curve. The use of a signal characterizer does not preclude the use of adaptive control and gain scheduling. In fact the use of all is synergistic. The characterizer provides compensation of a heuristic nonlinearity developed from plant or lab tests. This compensation based on steady state data frees the gain scheduler to deal with first principle nonlinearities and the adaptive controller to focus on the identification of dynamic nonlinearities. Note that the process does not need to be self-regulating to benefit from signal characterization. The use of steady state data simply means that the data points are gathered after transients have dissipated. The compensation of the nonlinearity by signal characterization is not adversely affected by the relationship depicted in Figure 9.1 where the process or valve gain depends upon the slope of a line to reach the end point from the present excursion.

The signal characterization is applied to the PID input for PV nonlinearity and to the PID output for valve nonlinearity. The signal characterizer provides a translation from the Y axis (ordinate) to the X axis (abscissa) of the curve. This concept is essential since the X and Y pair entered in the block is the curve Y and X data points, respectively. The signal characterizer has to do the opposite of what the process or valve is doing to compensate for the gain nonlinearity.

There are two primary examples in the use of signal characterizers. The nonlinearity of a titration curve is compensated by identical signal characterizers on the PID input and setpoint that convert pH (Y axis) to percent reagent demand (X axis). The X axis of a titration curve is actually the rate of reagent flow to feed flow. While this recognition is necessary for flow feedforward and realizing there is a flow ratio gain factor that is inversely proportional to flow, the use of percent reagent demand is a great simplification to the controlled variable that compensates for curvature. The resulting PID is called a linear reagent demand controller with a PV and setpoint scale of 0 to 100 percent. The linear reagent demand controller does not see the acceleration of the PV from a pH going on the steep part of the titration curve (e.g., approach to pH 7). For strong acids and strong bases the acceleration for a pH PID is so severe the PID sees a runaway response. Operators often ask what can be done to slow down the pH as it passes through the neutral region. The linear reagent demand controller eliminates this acceleration and restores effectively the primary process time constant. From this view point the operator will be appreciative when looking at the linear reagent demand PID.

Another common example is the installed flow characteristic of a control valve. The input to the signal characterizer is the PID output or split range (splitter) block. The signal characterizer output is the input to the analog output (AO) block that sends the signal to the valve. The signal characterizer converts the desired percent flow (Y axis) to the percent valve signal (X axis). The X and Y pairs are the Y axis scaled in percent maximum flow and the Y axis scale in percent valve signal, respectively. Signal characterization can be done in the digital positioner. Having the characterization in the Distributed Control Systems (DCS) provides more visibility and accessibility to the X and Y pairs.

Even if the curve is not well known or the curve changes with operating temperature and feed composition, signal characterization reduces the effect of auto or adaptive tuner step size and goes a long way to dealing with gross nonlinearities with exceptional resolution. The X and Y pairs should be more closely packed in the area of greatest nonlinearity in the normal operating range. The characterization must extend to include the abnormal but still possible operating ranges. Normally 21 pairs is enough to provide the coverage and resolution needed but a second signal characterizer can be added to the output of the first characterizer to enable effectively a *zoom-in* on particular regions of interest.

Signal characterization can create confusion in the control room. Operators, process engineers, and maintenance technicians are accustomed to monitoring performance of the pH measurement and loop via a PID whose setpoint and PV is pH. It may suffice to simply provide the ability to enter a setpoint in pH on the operator interface and provide a pH indication and trend. If confusion still reigns, a dummy PID with a pH scale can be created to output track the reagent demand PID when in automatic. When the pH PID is in manual, the reagent demand PID would output track the pH PID. The adjustment and tuning of PID parameters would need to be done on the reagent demand PID unless the parameter settings are copied from the dummy PID.

Operations and maintenance also expect the PID output that goes to a valve is the actual signal to the valve. With output signal characterization, the PID output is now desired flow rather than desired valve position. If operators or technicians want to stroke a valve in manual to check its operation or to set a position, they would assume that this could be done by manually adjusting the PID output and the valve position would match the PID output. It may suffice to simply provide a field on the operator interface to enter a manual output in percent valve position and provide an indication and trend of the actual signal to the valve and readback from the valve. If this is still confusing, when in automatic a dummy PID with an output that is desired valve position would track the actual PID output doing control after passing through a signal characterizer that is identical to the one for linearization. When the dummy PID is in manual, the actual PID would track the dummy PID output after going through a signal characterizer that does the reverse of the one for linearization (converts from X axis signal to Y axis flow) in order to provide the actual PID output in % desired flow.

9.2.4 GAIN SCHEDULING

For composition, pH, and temperature control of volumes where there is little back mixing most of the residence time (volume/flow) becomes dead time instead of a primary time constant. For these unit operations the inverse relationship to flow is not compensated by an inverse relationship of process time constant to flow. Gain scheduling with a factor that is inversely proportional to flow can free up the adaptive controller to account for unknowns, changes in gain nonlinearities, and enable a faster and better identification of a variable dead time and time constant.

9.2.5 ADAPTIVE CONTROL

Our knowledge of changes in gain with process and equipment conditions is far from perfect. Signal characterization and gain scheduling can take the loop a long way to gain linearity but the plots and relationships are approximations built on assumptions of operating conditions that is at best quite old. The titration curve for linearizing pH loops assumes the concentration of acids and bases in the feed and in the reagent are constant and match lab conditions. The installed flow characteristic for linearizing valves assumes constant and known system resistances. Gain scheduling for temperature control for production rate assumes a constant heat transfer coefficient. Adaptive control is based on identified dynamics rather than assumptions. Adaptive control provides an essential correction based on current data that is analogous to the need for feedback control despite the best feedforward control.

9.2.6 GAIN MARGIN

The gain margin, ratio of the ultimate gain to the PID gain, provides a measure of how much uncompensated variation in the open loop gain can be tolerated, assuming there are no other nonlinearities. As the gain margin approaches one, the loop approaches instability. The tuning settings for load disturbance rejection without causing an oscillatory response have a gain margin slightly larger than three, which corresponds to an arrest time equal to the dead time for Lambda tuning of integrating processes. The original settings for the Ziegler–Nichols (ZN) closed loop ultimate oscillation and open loop reaction curve method had a gain margin of about 1.7 that resulted in a quarter amplitude oscillatory response to minimize the peak and integrated error from a load disturbance. To eliminate the oscillation and to provide more robustness, the ZN PID gain setting is cut in half. Chapter 1 shows that Lambda tuning with an arrest time equal to the dead time and the short cut method give about the same PID gain setting as the ZN method with the gain cut in half; thus showing the universality of this approach.

An arrest time of three and four dead times, increases the gain margins to about six and eight, respectively. If the objective is to keep the gain margin above two to prevent excessive oscillations, these arrest times can handle variable gain change factors of three and four, respectively. A similar guideline can be used for setting the Lambda for moderate self-regulating processes. For these processes where resonance can occur from an oscillation period close to the ultimate period, the increase in the disturbance oscillation amplitude for the given Lambda settings is about 26 and 20 percent, respectively as explained by Bill Bialkowski in *Process/Industrial Instruments and Controls Handbook* (McMillan 1999).

For gain nonlinearities that are not addressed by secondary loops, signal characterization, gain scheduling, or adaptive control, the gain margin should be increased accordingly by an increase in Lambda resulting in a decrease in PID gain. If Lambda tuning for integrating processes is not used, the reset time must be proportionally increased for near-integrating, true integrating, and runaway processes to prevent the violation of Equation 1.5c and the trigger of slow rolling oscillations. The following best practices summarizes how and when to use the various solutions and important considerations in the process.

Best Practices for Dealing with Gain Nonlinearities

1. Use lower (secondary) loops to isolate nonlinearities from upper (primary) loops. Use a secondary flow loop to isolate the valve nonlinearity from the primary loop. Enable a bumpless transition to a computed flow from valve position using the installed flow characteristic below the meter low flow rangeability limit. Use a secondary coil or jacket temperature loop to isolate the flow ratio gain nonlinearity from the primary vessel temperature loop.
2. Use signal characterization to compensate for significant plant data based steady state nonlinearities. This compensation is not significantly affected by changes in setpoint or load and helps the open loop gain identified by an auto and adaptive tuner be more independent of test step size. For pH nonlinearities, use a signal characterizer applied to the PID setpoint and PV to provide a piecewise fit of the titration curve. For valve nonlinearities, use a signal characterizer applied to the PID or splitter block output to provide a piecewise fit of the installed flow characteristic curve. Pack the pairs of X and Y data points closer together to provide a finer resolution of the piecewise fit in the region of greatest nonlinearity near the setpoint. Use a cascade of signal characterizers to provide the resolution and range needed. The configuration and interface must be setup so that operators, process engineers, and maintenance technicians can readily change setpoints and manual outputs and analyze system performance.
3. For conductivity control, limit the operating range to one side of the peak in the plot of the conductivity versus concentration curve to prevent a change in process action sign or the need to add special intelligence (e.g., translation of the PV to a rate of change of conductivity with respect to the change in manipulated flow).
4. Use gain scheduling for known nonlinearities based on first principles. For example, schedule the PID gain to be inversely proportional to total feed flow for volumes without significant back-mixing and hence no appreciable primary process time constant from residence time.
5. Use adaptive control to compensate for unknown nonlinearities. Use the maximum allowable number of zones (e.g., 5) and narrow the zone width around each normal operating point. For nonlinearities that originate from an unknown curvature in the plot of the manipulated or controlled variable, the step size for identification of dynamics must be varied to cover the entire zone and the highest gain used. If a normal operating point exists in the zone, the steps should be made around this point in the zone or around the midpoint of the zone if this point in the zone is not known. For normal operating point outside the zone, steps should start at the boundary of the zone closest to a normal operating point.
6. For remaining nonlinearities that cannot be addressed by the above, increase Lambda so that the worst case ratio of the maximum possible open loop gain to the identified open loop gain used in the tuning of the PID is never less than two.

9.3 VARIABLE DEAD TIME

The dead time in the process changes with residence time, mixing, and any process lags (e.g., thermal lags) effectively become dead time. The changes with residence time can be computed and used to schedule the tuning settings. Whereas a change in open loop self-regulating or integrating process gain only affected the PID gain, a change in dead time affects the PID reset and rate time besides the PID gain for the short cut method and most other methods not based on closed loop time constants. For Lambda tuning of self-regulating processes, the dead time only affects the PID gain and the low limit to the reset time (half the dead time) and rate time (half the dead time) subject to the high limit for the ISA Standard structure (one-fourth the reset time) as seen in Equations 1.6b and 1.6f. For Lambda tuning of integrating processes, the dead time affects the reset time besides the low limit of the reset time (four times the dead time) and the low limit for the rate time (half the dead time) subject to the high limit for the ISA Standard structure (one-fourth the reset time) as seen in Equations 1.7d and 1.7f.

The dead time from sensors and valves is too difficult to accurately predict to be used for the scheduling of tuning settings. The dead time from *at-line* analyzers can be computed from the analyzer cycle time and multiplex time per Equation 6.4c plus any sample transportation delay. The dead time from *off-line* analyzers is extremely large and unpredictable. The enhanced PID eliminates the need to retune the PID as the dead time changes if the dead time from analyzers or wireless devices is larger than the 63 percent process response time (process dead time plus process time constant).

Consequently, more reliance is put on the adaptive controller to identify and compensate for changes in loop dead time that do not originate from production rate changes, assuming an enhanced PID is used for control loops with discontinuous updates (e.g., at-line or off-line analyzers and wireless devices).

For self-regulating processes, the gain margin must be larger than twice the ratio of the largest to smallest possible product of the open loop steady state gain and total dead time divided by the primary time constant. For all other processes, the gain margin must be larger than twice the ratio of the largest to smallest possible product of the open loop integrating process gain and total dead time.

9.4 VARIABLE TIME CONSTANT

Most time constants are not constant. The only fixed time constants are those associated with transmitter damping settings or signal filters. The time constants from sensors, valves, and process dynamics change with operating conditions and the direction and size of the change in the input.

For back mixed volumes, the primary process time constant for concentration, pH, and temperature control is proportional to the residence time. Since the process gain is inversely proportional to the residence time for these processes, there is no advantage to be gained here by scheduling tuning settings based on the change in the primary time constant with production rate.

The adaptive controller is the only general solution for dealing with the nonlinearity that stems from changes in time constants that are not based solely on direction of change. For a primary process time constant that changes with direction, an adaptive controller can conceivably be setup to change tuning settings based on a variable that gives a low noise indication of the direction of change, this can be readily done by a simple calculation freeing up the adaptive controller to deal with nonlinearities than cannot be so easily addressed.

To detect the direction of change, the PV is passed through a dead time block whose parameter is equal to or greater than the total loop dead time. The old PV that is the output of block is subtracted from the new PV that is the input to the block. The sign of the difference in PVs can be used to indicate direction and to trigger the switching of tuning settings.

9.5 INVERSE RESPONSE

An inverse response results from a secondary unintended process effect of opposite sign with an arrival time sooner than the primary process effect. The PV starts to go in the opposite direction of the final response. The PID output starts to change in the opposite direction.

The most common sources of inverse response are shrink and swell from the collapse and creation of bubbles, respectively and from feedforward signals that arrive too soon. For boiler drum level control, an addition of feedwater much cooler than the boiling mixture will cause a collapse of bubbles causing more liquid to shrink from the drum into the downcomers before the level rises due to the increase in liquid inventory. An increase in firing rate will increase the bubbles in the downcomers forcing more liquid up into the drum causing a swell in level before the increase in steam generation depletes some of the inventory causing the level to drop. A similar shrink and swell develops for distillation column level control by the manipulation of steam flow that causes bubble collapse and creation in the reboiler tubes.

A transient feedforward signal of opposite sign as per the material balance that decays to zero as the material balance takes effect has been successfully used to deal with severe shrink and swell that would cause boiler shutdown. These severe cases are typically caused by boilers with undersized steam drums or boilers being pushed to produce steam rates much higher than the original nameplate capacity.

To help reduce the inverse response, the steam flow goes through a shrink-swell gain multiplier block. The shrink-swell gain is larger than the three-element feedforward gain. The multiplier block output then goes to a filter block. The filter time constant matches the time to the peak of the uncompensated inverse response. The output from the multiplier block is subtracted from the output of the filter block. The net result is a transient feedforward signal of opposite sign for a steam flow change that decays to 86 percent of its initial value after two filter time constants. The transient shrink-swell feedforward is added to the normal material balance feedforward that makes the change in feedwater flow equal to the change in steam flow as part of the three-element boiler drum level control.

If the inverse response cannot be eliminated by a feedforward signal with the right dynamics, the general solution to minimize the oscillation is to increase the gain margin. Except for severe cases, a gain margin of six often suffices (Lambda equal to three dead times).

9.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and step change in setpoint (ΔSP). Table 9.1 summarizes the test conditions.

Table 9.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
9.2	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	Equal % valve
9.3	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	5:1 delay increase
9.4a, b, c, d	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	10:1 lag decrease
9.5a, b, c	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	Inverse response

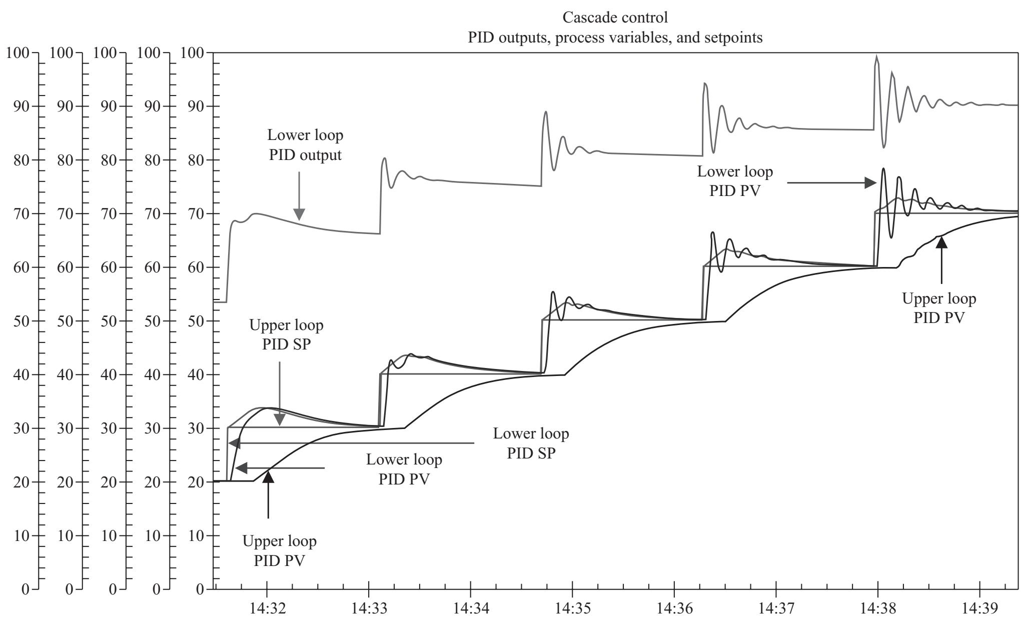


Figure 9.2. Nonlinear control valve *upper and lower Lambdas* = 3 × *dead time* for moderate self-regulating process.

The tests use a cascade control system with a base case where the lower loop dynamics are 10 times faster than the upper loop so that there is no problem with oscillations breaking out from violating the cascade rule. Moderate self-regulating processes are used with both the upper and lower Lambda set equal to the same multiple of loop dead time to help deal with nonlinearities. By setting the Lambda of both loops to be the same multiple, the upper loop stays 10 times faster the lower loop. The exception is when the process time constant is decreased by a factor of 10.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

Figure 9.2 shows that the 5:1 valve gain change for an equal percentage valve going from 50 to 90 percent opening was nicely handled by the lower loop because its dead time was so small. The lower loop breaks out into oscillations but the primary time constant filters them out so that oscillations only start to appear in the upper loop when the valve is 90 percent open. The test case shows that an oscillating lower loop is not detrimental to an upper loop. However, the oscillations will wear out valve packing and possibly upset other users of the same raw material, reagent, or utility flow being manipulated.

Figure 9.3 shows that the 5:1 increase in the upper loop dead time when the upper PID output drops from 50 to 20 percent causes severe slow oscillations in the upper loop that then appear in the lower loop as a result of setpoint changes. Such an effect would be typical for production rate changes if changes in process gain and time constant did not counteract the effect of the change in dead time. The 6:1 gain margin of a Lambda equal to three times the respective loop dead time is not enough. The upper Lambda would need to be increased to about six times the upper dead time to provide the needed gain margin.

Figures 9.4a shows for a step change in lower PID output in an open loop test (PID in manual) the upper loop primary process time constant is 10 times smaller for a decrease compared to an increase in the PV.

Figures 9.4b through 9.4d show that the use self-regulating process tuning rules with an upper and lower PID Lambdas equal to three times their respective loop dead time and decreasing the reset time of the upper PID helped get the upper PID smoothly to setpoint for a decrease in setpoint but caused conservable overshoot for an increase in setpoint. A better solution for self-regulating processes would be to switch the reset time based on whether the PV is increasing or decreasing. For processes using integrating process tuning rules (e.g., batch vessels and bioreactors) where the ramp rate depends upon direction, Lambda (arrest time) should be switched based on direction.

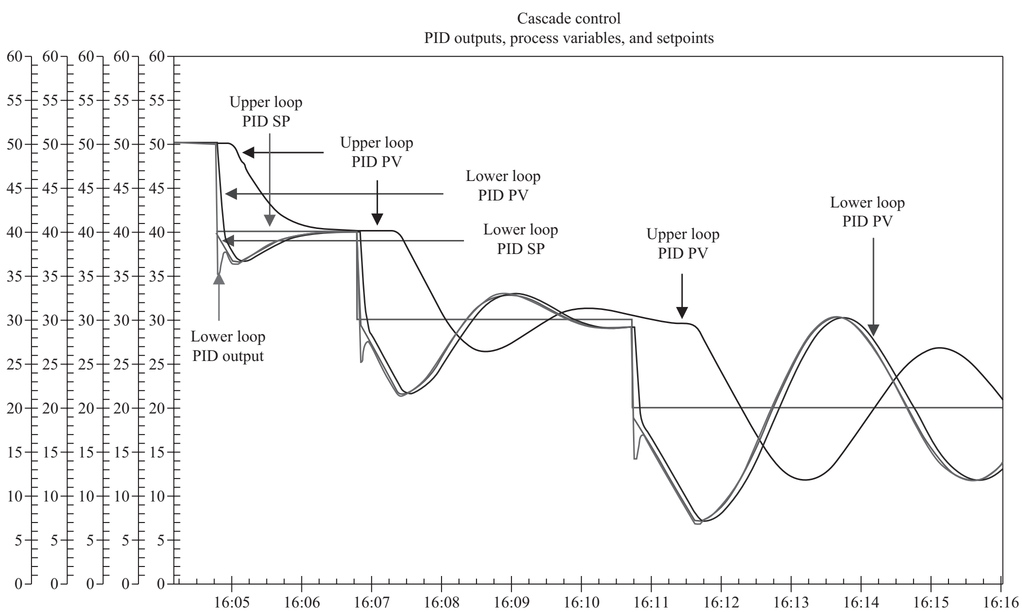


Figure 9.3. Nonlinear dead time *upper and lower Lambdas = 3× dead time* for moderate self-regulating process.

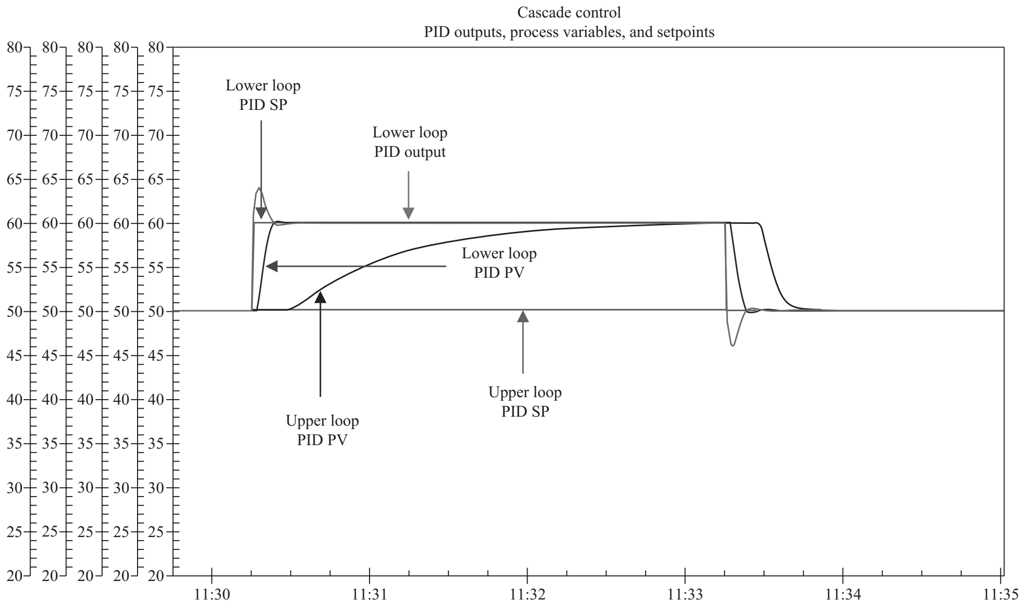


Figure 9.4a. Nonlinear time constant *open loop test* (upper PID in manual) for moderate self-regulating process.

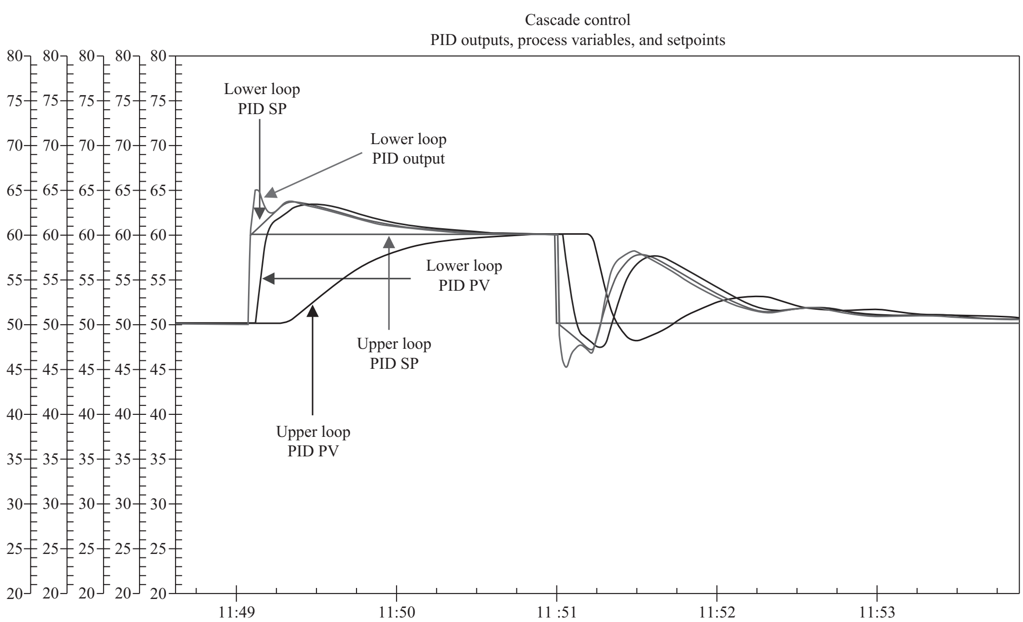


Figure 9.4b. Nonlinear time constant *upper and lower Lambdas* = 3 × *dead time* (upper PID reset time = 1.0 × largest upper loop lag for moderate self-regulating process).

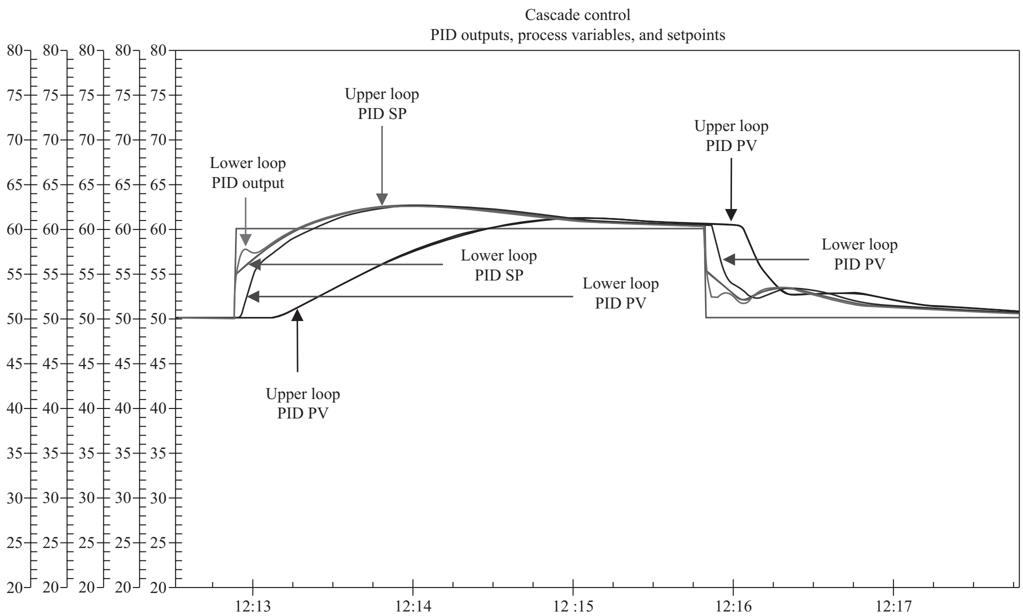


Figure 9.4c. Nonlinear time constant *upper and lower Lambdas* = 3 × *dead time* (upper PID reset time = 0.5 × largest upper loop lag for moderate self-regulating process).

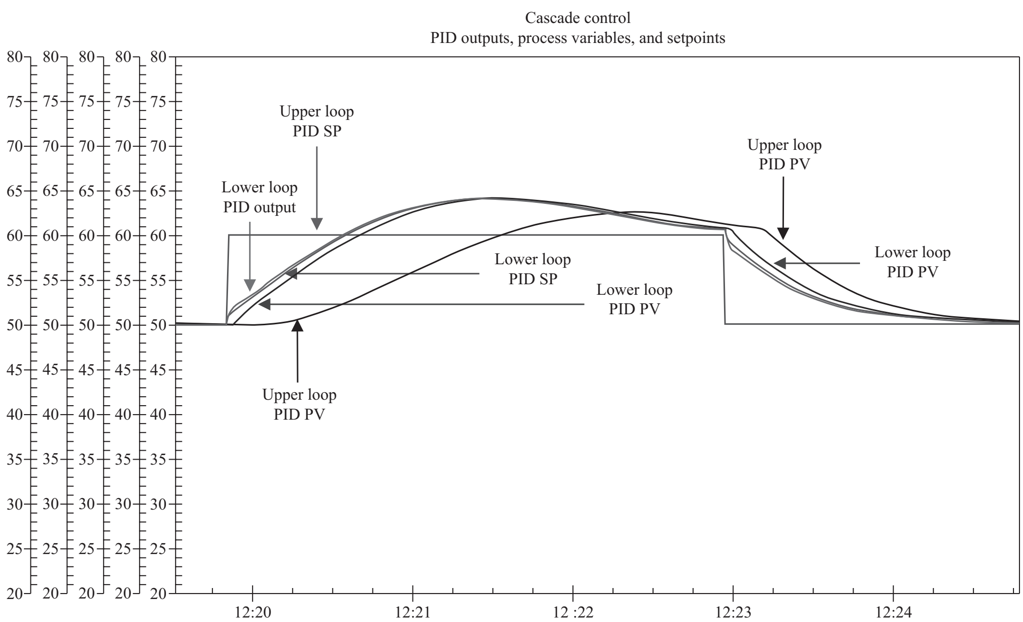


Figure 9.4d. Nonlinear time constant *upper and lower Lambdas* = 3 × *dead time* (upper PID reset time = 0.1 × largest upper loop lag for moderate self-regulating process).

Figures 9.5a shows that for a step change in lower PID output in an open loop test (PID in manual) the distinguishing characteristic of an inverse response where the initial response of the upper PID PV is in the opposite direction of the final response.

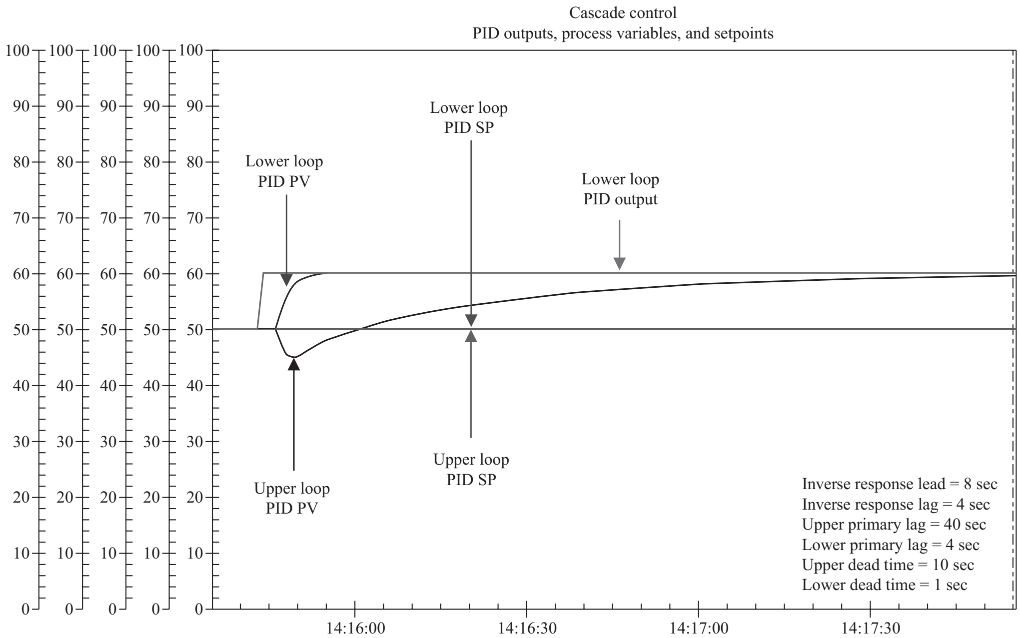


Figure 9.5a. Inverse response *open loop test* (upper PID in manual) for moderate self-regulating process.

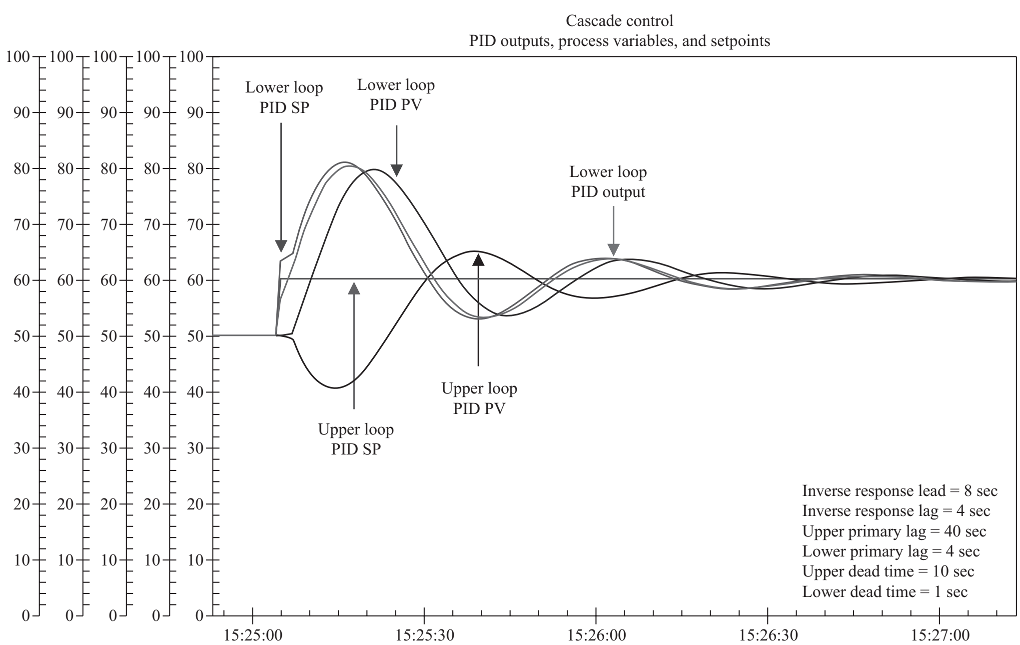


Figure 9.5b. Inverse response *upper and lower Lambdas* = 2 × *dead time* for moderate self-regulating process.

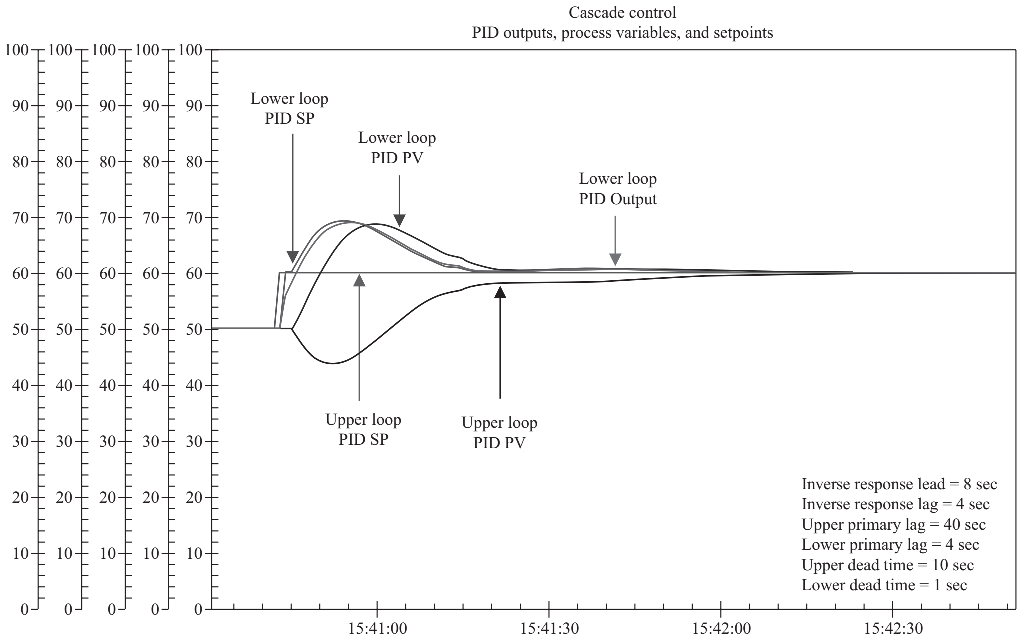


Figure 9.5c. Inverse response *upper and lower Lambdas* = 3 × *dead time* for moderate self-regulating process.

Figures 9.5b and 9.5c show that the use self-regulating process tuning rules with upper and lower PID Lambdas equal to two times their respective loop dead time doubles the amplitude of the inverse response and creates a slow oscillation. The simple increase of the Lambdas to three times their respective dead time reduces the amplitude of the inverse response to almost the open loop test case and prevents the subsequent oscillation. For much more severe cases of inverse response, the upper PID Lambda must be increased to a greater multiple of the upper loop dead time. Note that the initial increase in the lower PID Lambda may help to prevent the sudden changes that aggravate inverse response. However, at some point, only the upper Lambda needs to be increased providing even greater margin from violation of the cascade rule. Even a slight violation of the cascade rule is particularly problematic for inverse response.

KEY POINTS

1. A secondary flow loop can isolate a primary equipment concentration loop from the nonlinearities of a control valve.
2. A secondary coil or jacket temperature loop can isolate a primary equipment temperature loop from process as well as valve nonlinearities.
3. PID input signal characterization for the nonlinearity of a pH titration curve assumes the acid and base concentrations of the feed and reagent are constant.
4. PID output signal characterization for the nonlinearity of a valve's installed flow characteristic assumes the resistance if the piping system is constant.

5. Signal characterization inherently compensates for all steps sizes.
6. Signal characterization is far from perfect but frees up an adaptive controller to deal with unknowns as long as the change in curve slope is in the right direction.
7. Signal characterization requires intelligent operator interfaces to avoid confusion.
8. First principal relationships, such as those developed in Appendix F, can help schedule tuning setting and free up an adaptive controller to deal with unknowns.
9. Changes in the primary process time constant with direction can be detected and used to switch the reset time for loops tuned with self-regulating process methods.
10. Changes in the process ramp rate with direction can be detected and used to switch the arrest time for loops tuned with integrating process methods.

CHAPTER 10

Effect of Interactions

10.1 INTRODUCTION

Since most unit operations have multiple loops and streams, interactions can exist where the proportional-integral-derivative (PID) output of one loop affects the PID process variable (PV) of another loop and vice versa. Interactions can cause loops to confusingly burst into oscillations. Here we look at how to reduce the consequences and prevent the initiation of interactions. In the process we realize some principles in the design of control strategies for plantwide control.

10.1.1 PERSPECTIVE

Often a production rate control loop affects a quality control loop such as a composition, pH or temperature loop. Action of the temperature loop does not affect the feed loops if cooling and heating is manipulated rather than feed for temperature control. Composition and pH control have an effect on production rate by manipulation of a reagent or reactant flow but often this is the smaller feed stream and the effect is minimal and slow in well mixed liquid volumes. Consequently the focus can be on decoupling the effect of production rate changes on the quality control loops. The solution is a feedforward signal of main feed rate setpoint or flow measurement to the quality control loop. If valves are sized properly and external reset feedback is enabled, the flow setpoint rather than the flow measurement is preferred to eliminate measurement noise and to enhance the coordination and timing of ratioed flows. The composition or temperature PID output corrects the manipulated flow to feed flow ratio. This feedforward signal is a half decoupler where the changes in feed flow loop are passed on as changes to composition, pH, or temperature control loop output to preempt the upset.

In the literature, the decoupling signal is a PID output. If the PID output goes directly to a control valve, the nonlinearity and uncertainty of a control valve characteristic creates a large feedforward error. For a relative gain greater than 1.0, the error can cause a nearly full scale divergence. For these and many other reasons, the decoupling signal should be a flow setpoint or measurement in a secondary loop.

The dynamic compensation required is the same as for feedforward loops. The decoupling signal must cause a correction to arrive at the same place at the same time as the upset but with an opposite sign. Technically, this case is not true interaction because the effect is one way and not mutual (composition or temperature loop does not affect the feed loop). The addition of a valve position controller (VPC) to optimize production rate would create full interaction by increasing the feed rate to force the composition or temperature control valve to the furthest controllable position. In general, the use of VPC creates interactions that in the literature has been typically minimized by using slow integral only control in the VPC. However, this puts the composition or temperature loop at risk of running out control valve for large or fast disturbances because the VPC is too slow to react. An integral dead band can be enabled in the VPC to stop integral action for small offsets of the small valve position from VPC setpoint of the desired small valve position. An enhanced PID developed for wireless can be used with a threshold sensitivity limit and this can accomplish the same result as integral deadband. Setpoint rate limits of the loop optimized by the VPC (e.g., production rate setpoint) with external reset feedback of the PV can provide directional move suppression for the VPC.

The solutions in preferential order to reducing interaction is the best control strategy in terms of pairing of manipulated and control, decoupling, move suppression, and detuning. Here we will focus on interactions inherent in the basic control loops.

Consider an upstream pressure and downstream flow controller in series in the same pipeline. There is a small valve upstream and a large valve downstream of the pressure and flow measurements. What valve should the flow loop manipulate realizing the main objective is to control the flow downstream? The *off the cuff* answer is the flow loop should manipulate the large valve. Actually the flow controller should manipulate the valve with the largest pressure drop so that changes in pressure have least effect. Given the valves need to pass the same flow, the smaller valve upstream has the larger pressure drop and should be manipulated for flow control. Process dead time and time constant are not a consideration because they are very fast and the same for each valve for liquid flow.

A classic example often cited is inline composition and total flow control by the blending of two pure component feed streams, 1 and 2, fed to the inlet of a static mixer as shown in Figure 10.1. The total flow and composition is measured downstream of the static mixer. If the composition loop is controlling the concentration of component 1 from feed stream 1 and the desired concentration of component 1 is less than the desired concentration of component 2 at the outlet of the static mixer, then the concentration controller should manipulate flow 1 which is the smaller stream flow. This analysis also works for static mixer pH control. Consider stream 1 to be a reagent flow with a relatively high reagent concentration (e.g., 98 percent sulfuric acid). Stream 2 is a high flow dilute effluent stream (e.g., 0.1 percent ammonia). The historical choice of the pH controller manipulating the reagent stream and the total flow controller manipulating effluent will minimize interaction. Consider the unusual but possible case of an extremely dilute waste reagent stream that is larger than the effluent. Here the pH controller should manipulate the effluent stream and the total flow controller should manipulate the waste reagent flow. In both cases, a feedforward of flow controller PV should be added to the pH controller output to further reduce interaction and improve control for production rate (total flow setpoint) changes.

Another example occurs in large recirculation loops where flow and backpressure are to be controlled. Often the flow controller mistakenly manipulates the Variable Speed Drive (VSD) and the backpressure controller manipulates the control valve. The relative gain matrix would say the exact opposite pairing of controlled and manipulated variables.

The interaction between outlet and inlet temperature controllers for kilns that manipulate firing rate and exit gas flow can be reduced by the addition of an oxygen controller. A feedforward of firing rate to the outlet temperature controller as a decoupler can reduce the remaining interaction.

A more complex example showing the importance of first addressing gas and liquid inventory control is distillation column control. When both the top and bottom composition need to be tightly controlled (two point composition control) full interaction occurs between the temperature loops on the top and bottom of the column. Here temperature is an inferential measurement of composition usually in terms of the impurities (e.g., high boiling point key component on the top and low boiling point key component on the bottom of the column).

First, column pressure and levels must be controlled. Tight pressure control is needed so that boiling points of components throughout the column change with temperature, not pressure. The pressure controller should have a relatively large gain and reset time to rapidly manipulate the split ranged between nitrogen and vent valve. If the gain is decreased and the reset time is not correspondingly increased, slow rolling oscillations will result from integral action dominating proportional action in this integrating process as detailed in Chapter 4 on the effect of process dynamics. The slow rolling oscillations can cause the rest of the loops to perpetually oscillate.

Since reflux flow back to the column is often greater than distillate flow from the column, overhead receiver level PID often manipulates reflux flow. This leaves distillate flow from the receiver to be manipulated for temperature control. Because the column is not affected by a change in distillate flow until the reflux changes, tight level control is essential. Fortunately, tight level control is possible and desirable for the reasons such as inherent compensation of overhead vapor or reflux temperature changes. As with gas pressure control, the level controller should have more proportional than integral action for the integrating process. For the bottom of the column, the manipulation of bottoms flow is preferred for sump level control. However, if the bottoms flow is too small, then sump level PID manipulates steam flow to the reboiler. Here, tight sump level control is needed because a change in bottoms flow does not affect the column until the steam flow changes. Unfortunately, a slow secondary lag from heat transfer and an inverse response from bubble formation and collapse, prevents tight sump level control. A feedforward signal of bottoms flow, that is in reality a half decoupler, provides the immediate effect needed in terms of a change in steam flow.

The tray for top and bottom temperature control is selected that provides the largest change in temperature for a change in the manipulated flow in both directions (increase and decrease). By choosing the largest change in both directions, the composition at other trays in the column is more tightly regulated. The manipulated flow paired with the temperature is the one with the largest effect on the temperature. Normally, the top temperature is most affected by reflux flow (directly or indirectly from receiver level control) rather than steam flow so the top temperature is often not paired with steam flow.

After proper pairing, there is still an interaction with the two point composition control. The addition of feedforward of column feed to both loops can provide much of the decoupling needed. Typically the top temperature trims the reflux to feed or distillate to feed ratio and the bottom temperature trims the steam to feed or bottoms to feed ratio.

Periodic disturbances should be eliminated where possible by better process and automation system design and better tuning. These oscillations can trigger interactions causing a confusing situation where oscillations at different frequencies are spreading throughout the process.

Recycle streams and heat integration create more opportunities for interactions where disturbances come back with an integrating response from accumulation and at worse a runaway response (snowballing effect) from positive feedback. A good example of this problem is continuous reaction and recovery system. A fixed setpoint flow controller based on production rate needs to be set somewhere in the recycle stream path from a reactor through a recovery system and back to the reactor in terms of recovered reactant.

Separation of dynamics by tuning, making fast loops faster, can stop full interaction. In particular, the fast loops can correct quickly for a slow disturbance from the slow loop. A feedforward of the fast loop to the slow loop eliminates the rest of the interaction. The remaining concern is setpoint changes to the slow loop that result in large abrupt changes in the slow loop output that upset the fast loop.

10.1.2 OVERVIEW

The measurement with the greatest threshold sensitivity for the PV with the largest effect on quality should be selected as the controlled variable. The most notable example is the selection of the distillation column tray that shows the largest change in temperature for changes in both directions of the manipulated flow or flow ratio.

The manipulated variable with the largest effect should be paired with the associated controlled variable. A classic example is where the composition PID should manipulate the small pure component stream flow for a blend composition setpoint of less than 50 percent. A decoupling feedforward signal may be needed to prevent a PID output from saturating.

Given adequate effect on the controlled variable, the manipulated variable that is least affected by disturbances should be paired with the most important controlled variable. A common example is a pressure loop and flow loop in series with a large and small valve in the same pipeline. The flow loop should manipulate the small valve because the small valve has the larger pressure drop and hence is less affected by pressure disturbances.

Gas pressure and level loops should be paired with the manipulated flows that will keep the gas and liquid inventory within the desired range. Tight pressure control is almost always desired. Tight level control may be needed to enforce material balances and residence time requirements. Loose level control to use up available volume may serve to slow down as much as possible the manipulated flow as a disturbance to downstream operations. For a distillation column receiver, tight level control is needed when reflux is manipulated and loose level control is recommended when distillate flow is manipulated.

Fast loops should be tuned faster to reduce the effect of interaction with slow loops. The feedforward of fast loop PVs to slow loops can eliminate the remaining interaction. The most common example is the use of feed flow feedforward to preemptively change the manipulated flow to maintain a flow ratio. Dynamic compensation is applied so the change in the manipulated flow arrives at the same point at the same time in the process as the change in feed flow. The manipulated flow is corrected by composition, pH, and temperature loops. The same rules applying for setting the feedforward gain and the lead-lag apply for decoupling.

If interaction is still problematic after improvements in control strategy in terms of pairing and mitigating recycle effects, in process and system design and tuning to eliminate oscillations, by the use feedforward control for decoupling and handling production rate changes, and making fast loops faster, the final step is detuning to reduce the remaining propensity for oscillations.

Relative gain analysis is a powerful technique for accessing the type and degree of interaction. The relative gain for a given loop is the *open loop* steady state gain for the other loops *open* divided by the *open loop* steady state gain for the other loops *closed*. Note that the given loop is open for the tests with the other loops open and closed. A loop is considered open if the mode is manual or remote output or output tracking is enabled or the PID output is at an output limit. For integrating processes, the PV is translated to a rate of change so the concept of steady state can be used. The relative gain is dimensionless and thus does not depend upon the engineering units of the loops involved. The relative gain does not change when a flow ratio rather than a flow is manipulated. Operating point nonlinearities will affect the analysis but not changes in time constants and dead times. Ideally, the pairing of loops should have a relative gain close to one or slightly higher. Negative relative gains are disastrous.

Full interaction creates additional hidden feedback loops. The type oscillation observed is a clue to whether the loop has positive or negative feedback and to what degree.

If the subject loop's oscillation is severe showing signs of instability, there is a strong positive feedback loop created from other closed loops. The subject controller action can be reversed and tuned to stop the oscillations but this is not safe because if the offending loop becomes open (could be just temporary residence at output limit), the action sign will be wrong again. Process or system redesign is needed to eliminate this extreme type of interaction. Most severe interaction problems are caused by poor process or system design. This scenario corresponds to a negative relative gain.

If the subject loop's oscillation period and amplitude are larger but with a normal rate of return back to setpoint, there is a parallel negative feedback path created by the other closed loops. The open loop gain and dead time in the subject loop has increased, requiring that the PID gain be decreased and the reset time be increased. This scenario corresponds to a relative gain between 0 and 1.

If the subject loop's oscillation period does not appreciably change but the return back to set point is protracted, there is a slight positive feedback path created by the other closed loops. For setpoint changes, the subject loop will peak below setpoint and very slowly approach setpoint. For both disturbances and setpoint changes, the approach to setpoint is so slow, there appears to be an offset and no integral action. For loops with similar dynamics, the subject loop PID gain should be decreased. This scenario corresponds to a relative gain greater than one.

Furnaces with multiple burners for zone temperature control are a medium scale interaction problem with a medium relative gain matrix (e.g., 5 × 5). Sheets with multiple actuators for cross directional thickness control create an interaction problem grand in scale with a huge relative gain matrix (e.g., 100 × 100). Some multi-variable sheet thickness control systems seek to use a model of the deflection of the die lip when an actuator is manipulated. This is more easily said than done because an exact match is needed to effectively decouple the actuator actions from each other. The implementation requires a lot of trial and error in model adjustment.

10.1.3 RECOMMENDATIONS

1. Pick the measurement with the greatest threshold sensitivity for the PV that shows the greatest sensitivity to both increases and decreases in the manipulated flow or flow ratio.
2. Pair manipulated variables with controlled variables that have the greatest effect on the controlled variable and are least affected by disturbances.

3. Improve process and system design to eliminate oscillations.
4. Tune fast loops faster and slow loops slower to reduce interaction between the loops by providing a greater separation of dynamics. If the loops have similar dynamics, tune the less important loops much slower to provide the separation.
5. Add flow feedforward control to decouple and facilitate production rate changes.
6. Do a relative gain analysis to access the type and degree of remaining interaction.
7. Look at oscillation pattern to access the type and degree of remaining interaction.
8. Detune loops whose oscillation period and amplitude is larger due to interaction.
9. If decoupling involves complex dynamics, consider Model Predictive Control (MPC).
10. If decoupling is more than flow feedforward, consider MPC.

10.2 PAIRING

The relative gain array (RGA) developed by Ed Bristol and extensively applied by Greg Shinskey has been an extremely effective tool for selecting the proper pairing of controlled and manipulated variables to minimize the consequences of interaction. Most of the material presented is a result of an understanding gained from the fourth edition of *Process Control Systems* (Shinskey 1996).

The relative gain given by Equation 10.1 is the ratio of steady state process gains for the pairing of process variable 1 (ΔPV_1) and manipulated variable 1 (ΔMV_1). The numerator is the steady state process gain for other loops open $\left[\Delta PV_1 / \Delta MV_1\right]_o$ and the denominator is the steady state process gain for the other loops closed $\left[\Delta PV_1 / \Delta MV_1\right]_c$ where the subscript "o" denotes the other loops are open (e.g., in manual) and "c" denotes the other loops are closed (e.g., in automatic). Note that the effect of valve gains and measurement gains do not need to be included since these gain factors would appear in both the numerator and denominator and cancel out. Also, the change in the manipulated variable (e.g., change in manipulated flow), is used rather than the change in controller output to eliminate the nonlinearity of the control valve installed flow characteristic that would change the relative gain for different setpoints or loads. This focus enables process gains to be evaluated based on first principle equations for the process as seen for the static mixer example.

$$\lambda_{11} = \frac{\left[\Delta PV_1 / \Delta MV_1\right]_O}{\left[\Delta PV_1 / \Delta MV_1\right]_C} \tag{10.1}$$

Whether the manipulated variable is a flow or a ratio of flows does not change the relative gain. However, the use of a flow ratio makes more sense for composition, pH, and temperature control because these PVs are all functions of flow ratio and the process gain is the slope on the plot of the PV versus the ratio. For pH the process gain for small changes in flow is the slope of the titration curve that is the plot of pH versus ratio of reagent to feed flow. For large changes in flow, the process gain is the slope of the line connecting the old and new points on the titration curve as explained in Chapter 9. Consequently, tests done to identify the process gain with the other loops open and closed should employ the same manipulated variable step direction and step size and start out at the same operating point. Also, tests should be done when there are no load changes and no other disturbances.

Engineering units and plant size does not affect the results. Thus the results can be used for other plants and parallel trains that have the same type of equipment as long as the composition, pH, and temperature setpoints are the same.

10.2.1 RELATIVE GAIN ARRAY

The RGA is typically shown to have the controlled variables assigned as rows and the manipulated variables assigned as columns. The sum of relative gains in each column must equal one and the sum in each row must equal one. Consequently, for the 2 × 2 RGA (Λ) shown in Equation 10.2, only one relative gain must be found. For example if the first relative gain (λ_{11}) for the pairing of process variable 1 (PV_1) and manipulated variable 1 (MV_1) is found we can very simply compute the other relative gains. The relative gain for the complementary pairing of process variable 2 (PV_2) and manipulated variable 2 (MV_2) is equal to this first relative gain ($\lambda_{22} = \lambda_{11}$). The other possible pairings are equal to one minus the first relative gain ($\lambda_{21} = 1 - \lambda_{11}$ and $\lambda_{12} = 1 - \lambda_{11}$).

$$\Lambda = \begin{array}{c} \\ \Delta PV_1 \\ \Delta PV_2 \end{array} \begin{array}{c} \begin{array}{cc} \Delta MV_1 & \Delta MV_2 \end{array} \\ \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \end{array} \tag{10.2}$$

Most PID control problems can be simplified to the 2 × 2 RGA shown by realizing that inventory control loops such as gas pressure and level, must be closed before the RGA is computed and are outside of the RGA analysis, reducing the problem to two key PVs. Since the RGA matrix must be square, the study comes down to selecting two manipulated variables that give the best relative gains in a pairing of variables for PID control. For example, consider the classic case of two point composition control for distillation columns. The RGA analysis comes down to selecting what manipulated variables should be paired with the top and bottom temperatures that are inferential measurements of composition. Different combinations of manipulated variables are inserted into the matrix such as distillate flow, reflux flow, steam flow, and bottoms flow. Often flow feedforward is used so that effectively the PID is manipulating a ratio of flows. The most common ratio is with respect to feed flow by virtue of a feed flow feedforward signal. For startup and batch operations, ratios such as reflux to steam may prove useful since there is no steady state or feed. Often columns are started up on ratio control. The temperature control loops are not put in automatic until the column reaches operating conditions. For batch operations, the differentiation of the temperature creates a steady state PV. To provide a better signal to noise ratio, the same technique used to predict future values described in Chapter 1 and 15 that employs a dead time block to create old values is used to compute the slope of the temperature profile with respect to time. The slope can achieve a steady state (constant slope) and be the PV for the PID and the RGA.

10.2.2 DISTILLATION COLUMN EXAMPLE

For distillation columns, the best trays for top and bottom composition control are found from process simulation and field tests. The best tray is the one that shows the largest and most symmetrical temperature change for a change in the possible manipulated variables.

The column sump and distillate receiver level controllers must be able to control the inventories in these volumes. Most often the sump level is controlled by manipulating bottoms flow and receiver level is controlled by manipulating the reflux flow. If the bottoms flow is much less than the boil-up, then steam may need to be manipulated for sump level control despite the introduction of an inverse response. If the reflux flow is much less than the distillates flow, then receiver level may need to manipulate distillate flow despite the loss of inherent compensation of internal reflux for changes in ambient temperature. The point is that some process understanding can simplify the matrix in this case to be an evaluation of the pairing of top and bottom temperature with distillate or reflux to feed ratio and a bottoms or boil-up to feed ratio. Process simulation and field test results may also show that for the best pairing only one temperature really needs to be controlled. In fact many two point composition control problems simplify to a single point control system after the best tray and pair is chosen for the dominant temperature loop and improvements such as better temperature and flow measurements and feedforward control are implemented. Frequently the two point composition control problem is the result of deficiencies in the implementation of the control strategies, measurements, and PID controllers. Often ad-hoc control system fixes don't address the inherent interaction problem and just add complexity making the situation more confusing.

10.2.3 STATIC MIXER EXAMPLE

Consider the quality and production control of a static mixer shown in Figure 10.1a with feed stream 1 with pure component 1 and feed stream 2 with pure component 2. The quality control is accomplished by a concentration control loop of component 1 in the static mixer outlet. Production rate control is done by a total flow control loop. Each stream has a measurement of the manipulated flow so that the nonlinearity of the control valve does not affect the RGA. A cascade control loop where the output of the composition controller AC-1 is the setpoint for a secondary flow loop FC-1 on stream 1 is used to isolate the valve nonlinearity and

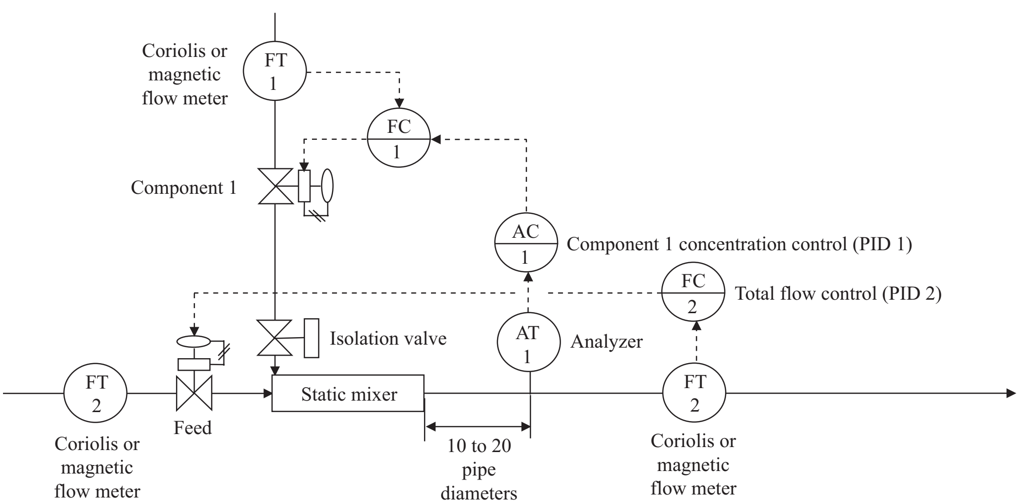


Figure 10.1a. Best pairing of static mixer loops to achieve quality control by concentration control and production rate by total flow control.

pressure disturbances from AC-1. The total flow is controlled by FC-2 manipulating a valve in stream 2.

In this 2 × 2 RGA we just need to find one relative gain. The relative gain for the control of total flow by the manipulation of stream flow 2 can be easily computed from a simple mass balance. All flows are mass flows and the concentration is a mass fraction. The pairing for finding this relative gain is opposite of the best pairing shown in Figure 10.1a.

As shown in Equations 10.3a and 10.3b, the total flow (F) is the sum of the two stream flows and the mass fraction concentration of component 1 (X_1) at the outlet of the mixer is simply the mass flow of stream 1 (F_1) as a fraction of the total mass flow. Equation 10.3b can be solved for total flow to give Equation 10.3c. We can use these equations to find the relative gain for the control of total flow by the manipulation of stream 1 flow. The numerator is the total flow control with the other loop open, which means the concentration control loop is in manual and stream 2 flow is held constant. The numerator is the partial derivative of Equation 10.3a (change in total flow with respect to stream 1 flow). The result is simply 1 as found by Equation 10.3d. The denominator is total flow control with the other loop closed, which means the concentration control loop is in auto and concentration of component 1 is held constant. The denominator is the partial derivative of Equation 10.3c (change in total flow with respect to stream 1 flow with the composition constant). The result is simply the inverse of the component 1 concentration as found by Equation 10.3e. The relative gain as shown in Equation 10.3f is the component 1 concentration (X_1); Equation 10.3d divided by Equation 10.3e.

$$F = F_1 + F_2 \tag{10.3a}$$

$$X_1 = \frac{F_1}{F} \tag{10.3b}$$

$$F = \frac{F_1}{X_1} \tag{10.3c}$$

The process gain with the concentration control loop open is the partial derivative of Equation 10.3a with respect to stream 1 flow with stream 2 flow constant:

$$\left[{}^{\Delta F}\!/\!{}_{\Delta F_1} \right]_O = 1 \tag{10.3d}$$

The process gain with the concentration control loop closed is the partial derivative of Equation 10.3c with respect to stream 1 flow with static mixer concentration of component 1 constant:

$$\left[{}^{\Delta F}\!/\!{}_{\Delta F_1} \right]_C = \frac{1}{X_1} \tag{10.3e}$$

The relative gain for the paring of total flow as the controlled variable and stream flow 1 as the manipulated flow is Equation 10.3d divided by Equation 10.3e:

$$\lambda_{21} = \frac{\left[{}^{\Delta F}\!/\!{}_{\Delta F_1} \right]_O}{\left[{}^{\Delta F}\!/\!{}_{\Delta F_1} \right]_C} = X_1 \tag{10.3f}$$

Since the rows and columns must add up to 1, we can solve for the other relative gains giving us Equation 10.3g for the RGA. Note that the relative gains are equal for the two choices for pairing of controlled variables and manipulated variables. If the mass fraction setpoint is less than 0.5, the best pairing is the control of static mixer concentration by the manipulation of stream 1 flow which is correspondingly smaller than stream 2 flow. This solution is intuitive in that the larger stream is manipulated to control the total flow. The relative gains for this pairing are both greater than 0.5 but less than 1.0, which will be seen to facilitate decoupling.

$$\Lambda = \begin{array}{c} \\ \Delta X_1 \\ \Delta F \end{array} \begin{array}{c} \begin{array}{cc} \Delta F_1 & \Delta F_2 \end{array} \\ \begin{bmatrix} 1 - X_1 & X_1 \\ X_1 & 1 - X_1 \end{bmatrix} \end{array} \tag{10.3g}$$

Nomenclature for 2 × 2 RGA:

ΔMV_1 = change in manipulated variable 1 (e.u.)
ΔMV_2 = change in manipulated variable 2 (e.u.)
ΔPV_1 = change in process (controlled) variable 1 (e.u.)
ΔPV_2 = change in process (controlled) variable 2 (e.u.)
F = total flow at outlet of static mixer (e.u.)
ΔF = change in total flow at outlet of static mixer (e.u.)
F_1 = manipulated stream 1 flow (e.u.)
F_2 = manipulated stream 2 flow (e.u.)
ΔF_1 = change in stream 1 flow (e.u.)
ΔF_2 = change in stream 2 flow (e.u.)
ΔX_1 = change in component 1 concentration at outlet of static mixer (mass fraction)
X_1 = component 1 concentration at outlet of static mixer (mass fraction)
λ_{11} = relative gain for the pairing of PV_1 and MV_1 (dimensionless)
λ_{12} = relative gain for the pairing of PV_1 and MV_2 (dimensionless)
λ_{21} = relative gain for the pairing of PV_2 and MV_1 (dimensionless)
λ_{22} = relative gain for the pairing of PV_2 and MV_2 (dimensionless)
Λ = relative gain array (dimensionless)

10.2.4 HIDDEN CONTROL LOOPS

For a 2 × 2 RGA, there is one hidden loop that has a path through both of the interaction process gains and the PID controllers. For the static mixer example, consider a change in concentration controller output. The change in stream 1 flow to control concentration passes through a process gain of one and causes an equal change in total flow. The total flow PID then responds and decreases the stream 2 flow, which passes through a concentration process gain that is the inverse of the total flow to cause a concentration change. The concentration PID then responds and changes the stream 1 flow completing the hidden loop. For a 3 × 3 RGA, there are three hidden control loops.

Hidden control loops are obviously confusing since the effect is not recognized and there is no PID whose options and tuning can be used to control what is going on in the loop. The hidden loop can cause a closed loop positive or negative feedback response depending upon the value of the relative gain. As emphasized in Chapter 1, positive feedback causes instability.

Processes with a positive feedback open loop response (e.g., runaway response or open loop unstable response) rely upon the PID controller to provide enough negative feedback action so that the closed loop response has negative feedback. To better understand the creation and consequences of hidden loops and to provide better guidance as to pairing and tuning without excessive complexity, the effect of different relative gain ranges is considered for a 2 × 2 RGA.

10.2.5 RELATIVE GAINS LESS THAN ZERO

If the relative gain is negative, the numerator and denominator in the relative gain have opposite signs. The hidden loop in a 2 × 2 RGA has a dominant positive feedback closed loop response. Instability can be prevented by reversing the control action of one of the visible loops PID but the response is poor due to all the dynamics being in series in the hidden loop. This reversal of sign is only valid if the other loops are in automatic mode and functioning properly. If one of the visible loops PID is put in manual or hits an output limit, the other visible loop PID will have a positive feedback closed loop response. *For these and other reasons loops should not be closed if the relative gain is negative*.

10.2.6 RELATIVE GAINS FROM ZERO TO ONE

The ideal relative gain is one because this means the closing of the other loops has no effect on the control loop. However a relative gain of one can easily increase due to unknowns and non-linearities to become much greater than one which could lead to a hyper sensitivity to decoupling errors noted in Section 10.2.8. The good news is that a relative gain of about one means decoupling is not needed. The conclusion is that decoupling should not be used on a loop with a relative gain close to or exceeding one.

If the process gain when the other loops are in manual is zero, the numerator is zero and the relative gain is zero. Consequently the subject loop will not be able to control unless the other loops are in automatic. When this situation occurs because the subject loop depends upon an inventory control loop being closed (e.g., level control in automatic), there is no general concern because inventory loops must be in automatic. For other situations, the loop can become dysfunctional when other loops are put in manual or hit an output limit. Also the loop is perilously close to having a negative relative gain and a disaster in terms of positive feedback and reversal of the sign of process action. Thus, unless the zero relative gain is solely due to an inventory loop, the pairing is a bad choice.

The subject loop will be stable but the response will be slower for a relative gain greater than zero but less than one. The ultimate gain will be larger due to interaction necessitating a decrease in the maximum PID gain. If the subject loop is much slower than the other loops, then the ultimate period and the corresponding minimum reset time and maximum rate time stay the same and just the maximum allowable PID gain must be reduced. If the interacting loops have similar dynamics the ultimate period is larger. The minimum allowable reset time and maximum allowable rate time should be increased. Equations 10.4a through 10.4d in Section 10.5 estimate these changes in tuning settings for linear interacting loops. For loops that are not aggressively tuned, further detuning PID could be unnecessary and even detrimental. Excessive detuning of fast loops can make interaction worse by reducing the separation of the closed loop response dynamics.

10.2.7 RELATIVE GAINS GREATER THAN ONE

The ultimate gain is less for a loop with a relative gain greater than one that is much slower than the other loops. While theoretically, the PID tuning could be made more aggressive as per Equation 10.4e in Section 10.5, a higher PID gain is not practical. If the other loops are put in manual or hit an output limit, the more aggressive tuning is no longer valid. For interacting loops with similar dynamics, the PID gain should be decreased as per Equation 10.4f in Section 10.5.

For load disturbances, interacting loops with relative gains greater than one have a good load response by sharing the burden. The response is non oscillatory if interacting loops with similar dynamics have their gain reduced.

The setpoint response is problematic for interacting loops with relative gains greater than one. For the loop with a setpoint change, the PV falters in its approach to the new setpoint and takes a long time to reach the new setpoint. The other loops that did not have their setpoint change falter in the recovery and take a long time to return to setpoint. Decoupling is not a good option due to hypersensitivity to decoupling errors. If a relative gain greater than 1 must be used, the best solution to improve the setpoint response is to preposition PID outputs to their final resting value (FRV) and hold the PID output at the FRV until the future value of the PV is predicted to reach setpoint. For near-integrating loops, the full throttle response described in Chapters 12 and 15 can be used where the PID output is held at an output limit and positioned to the FRV for one dead time when the future value is predicted to reach setpoint. This action is extremely disruptive to the other loops, so these other loops must be held at their respective FRV until the near-integrating loop is released for feedback control.

10.2.8 MODEL PREDICTIVE CONTROL

The difficulty of the RGA analysis and solution escalates as the size of the matrix increases. For a 3×3 matrix, four relative gains must be identified. If ratios as well as flow differences are candidates, there are 72 viable combinations that can be configured in six possible ways giving 452 different pairings. Fortunately, the interactions in most liquid continuous and batch unit operations, such as columns, crystallizers, neutralizers, and stirred reactors can be reduced to a 2×2 RGA by simulation and plant test results.

For matrices that cannot be reduced to 2×2, MPC is most likely the better solution. This is generally the case for gas reactors and furnaces particularly if there are multiple burners. MPC also provides sophisticated optimization based on changing market conditions and feed stocks. For large volume processes, a fraction of a percent improvement in process capacity or efficiency translates to huge financial gains. These conditions help explain the success of MPC in oil and gas and the front end of hydrocarbon processes. However, the use of manipulated variables that are ratios of flows and the implementation of flow ratio control is not part of the typical implementation shown in the literature. There may be a need for special formulations of the matrix and different models for different production rates to handle the nonlinearity.

10.3 DECOUPLING

The addition of one feedforward signal to one of the interacting loops in a 2×2 RGA is sufficient to break the hidden loop. This half decoupler has the same setup and dynamic compensation

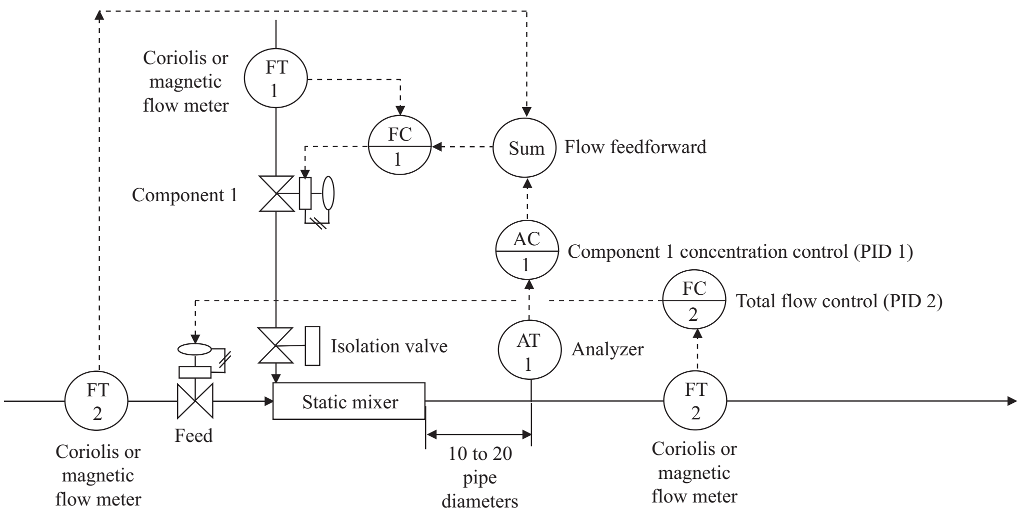


Figure 10.1b. Best pairing of static mixer loops with half decoupler to achieve quality control by concentration control and production rate by total flow control.

requirements of a feedforward signal. The feedforward correction needs to arrive at the same point and at the same time in the process as the effect of the interaction disturbance but with an opposite sign to cancel out the interaction. Figure 10.1b shows that for the static mixer, the half-decoupler is simply a flow feedforward where flow stream 2 measurement is multiplied by a ratio to become the setpoint for the stream 1 flow controller. The ratio is trimmed by the addition of a correction in terms of the concentration controller output. A bias of −50 percent is used so that the PID can make a full positive and negative correction. With this bias at 50 percent PID output, there is no correction. The actual ratio should be displayed along with access to set the desired ratio.

This feedforward strategy is commonly used not realizing the objective is decoupling besides preemptive coordination to maintain flow ratios. This strategy is useful for nearly all composition, pH, and temperature loops since the operating point is on a process curve plotted versus a ratio of the manipulated flow to the feed flow. The process can go on flow ratio control during startup or when the analyzer is out of service. This is particularly important for distillation columns and complex analyzers. If the flow measurement does not have enough rangeability, there should be a bumpless transfer to an inferential flow measurement computed from valve position and an estimated installed flow characteristic. Wherever possible, magnetic flow meters and Coriolis meters should be used because of their extraordinary rangeability and excellent signal to noise ratio.

Figure 10.2 provides a block diagram for the general depiction of a 2 × 2 RGA with a half decoupler. The biases in the interaction paths are used to make the cross term zero after a multiplication of the output by the relative gain for the normal output. This zeroing of the cross term enabled the manipulated variable to have a complete range of control (a PV response of 0 to 100 percent to the manipulated variable change of 0 to 100 percent).

This diagram would be representative of a two point distillation column control system. Loop 1 has a slow response which would be typical for a top column temperature controller (PID 1) doing material balance control by manipulating the distillate flow. Loop 2 has a fast

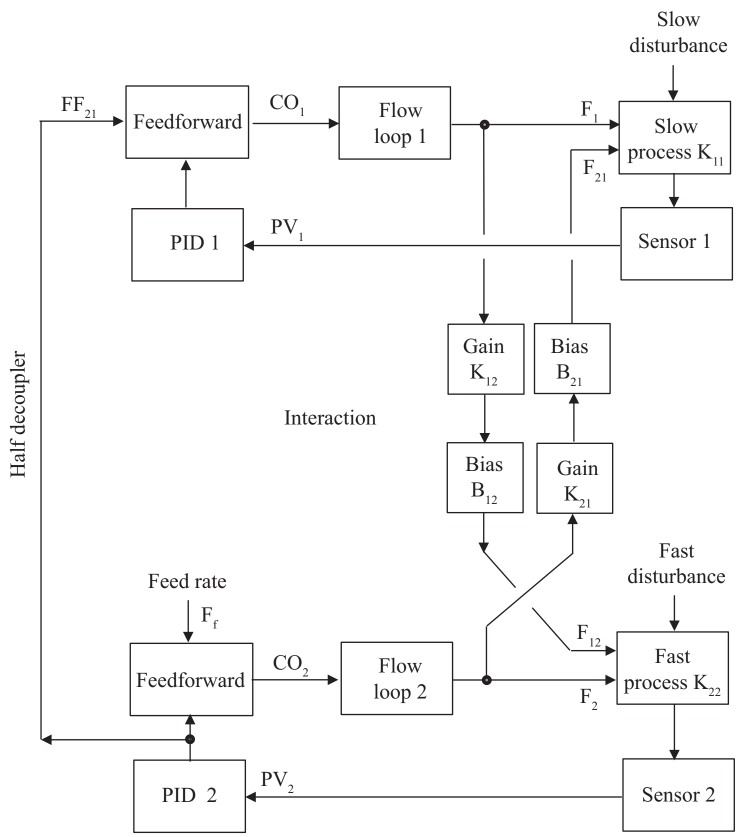


Figure 10.2. General block diagram of two interacting loops with a half decoupler that is the fast loop PID 2 output added as feedforward to the slow loop PID 1 output.

response which would be typical for bottom temperature controller (PID 2) doing separation (boil-up) control by manipulating steam flow. The general strategy is to send a decoupling signal from the fast loop to the slow loop. The fast loop can correct for disturbances and interactions much faster than the slow loop. The slow loop benefits the most from a decoupling signal.

The decoupling signal is the PID 2 output in engineering units added as a feedforward to PID 1 output in engineering units. Since these PID outputs are the setpoint of secondary flow loops (e.g., distillate and steam flow), what we have is flow ratio control with a correction by the PID 1 output. As mentioned before, the use of controller outputs that go directly to a control valve as decoupling signals is undesirable because of the severe decoupling error introduced by the valve's installed flow characteristic.

A feedforward of column feed rate as an indicator of production rate is added to the bottom temperature PID 2 output to provide steam to feed flow ratio control. The feedforward should not be added to the PID 1 output because a feed flow change will result in a steam flow change that results in distillate flow change through the decoupler. The feed flow feedforward already exists in PID 1 courtesy of the decoupling signal.

If the feedforward gain and dynamic compensation is accurate the half decoupler signal breaks the hidden loop in Figure 10.2 that passes through the interaction and process dynamics and PID blocks. A change in PID 2 output in engineering units should propagate as a change in feedforward gain that is equal in magnitude but opposite sign as to the change that passed through the interaction dynamics and affected the process. The arrival of the correction by the feedforward should be at the same point and same time in the process as the interaction. If the flow control loops are tuned for the same closed loop response and both the feedforward correction flow (FF_{21}) and interaction flow (F_{21}) are process inputs, dynamic compensation is typically not needed and the feedforward gain is simply the interaction gain. If the process actions (e.g., direct or reverse are the same for the correction and interaction flows, the feedforward gain is the negative of the interaction gain ($K_{ff} = -K_{21}$).

For the two point column composition example being discussed, the correction from the decoupler as a change in distillate flow will affect the control tray after the interaction effect from the change in steam flow. A lead-lag with a lead time equal to the lag associated with the liquid material balance response and a lag equal to the separation (boil-up) response can be applied to the dynamic compensation to help with timing.

If the relative gain is greater than one, the decoupling becomes very sensitive to feedforward errors. For a relative gain of four, the loop is destabilized for decoupling error of just 15 percent. The sensitivity increases with relative gain. Decoupling should only be applied for relative gains between zero and two.

10.4 DIRECTIONAL MOVE SUPPRESSION

The simple addition of a slow set point rate limit on the manipulated variable (e.g., secondary flow loop setpoint) in the direction where correction can be slow can reduce interaction from an optimizing loop. External reset feedback must provide knowledge of the action of production or efficiency variable (e.g., feed flow or chiller temperature) to the optimizing PID. This strategy is effectively used by VPCs with external reset feedback to provide a gradual optimization via a slow setpoint rate limit for an increase in feed rate or chiller outlet temperature that minimizes interaction with the composition and temperature loops. The gradual optimization reduces the interaction between control loops for utilities, production rate, and quality.

10.5 TUNING

Equations 10.4a through 10.4d developed by Shinskey show how much the PID tuning settings must be changed in accordance with the slower dynamics introduced by interaction. The decrease in PID gain and increase in reset time is intended to give about the same robustness (e.g., same gain and phase margin) as seen with the other loops open. If the PID has more than enough robustness, this detuning may not necessary.

For a loop with a relative gain between 0 and 1 that is much slower than the other loops:

$$K_c' = \lambda * K_c \tag{10.4a}$$

For interacting loops with a relative gain between zero and one and similar dynamics:

$$K_c' = (0.22 + 0.78 * \lambda) * K_c \tag{10.4b}$$

$$T_i' = \frac{T_i}{(0.22 + 0.78 * \lambda)} \tag{10.4c}$$

$$T'_d = \frac{T_d}{(0.22 + 0.78 * \lambda)} \tag{10.4d}$$

For a loop with a relative gain greater than one that is much slower than the other loops if the other loops are in automatic and not at output limits:

$$K'_c = \frac{K_c}{\lambda} \tag{10.4e}$$

For interacting loops with a relative gain greater than one and similar dynamics:

$$K'_c = \lambda * \left(1 - \sqrt{1 - 1/\lambda}\right) * K_c \tag{10.4f}$$

Inventory control loops should be in automatic and well-tuned before tuning the interacting loops. The interacting loops with the relative gains closest to one should be tuned first to reduce the number of iterations of tuning. As the relative gain gets further away from one, loops are more affected by the tuning of other loops causing more iteration. Loops with a relative gain close to one may only need to be tuned once.

For loops with similar relative gains, the fastest loop should be tuned first for a fast response. Separation of dynamics can reduce the oscillations. The rule to minimize interaction is similar to the rule for cascade control where the lower (secondary) loop is tuned first to have a closed loop response that is five times faster than the upper (primary) loop. If the fast loop cannot be tuned faster, the slow loop may need to be tuned slower to provide the separation in dynamics.

Nomenclature for retuning interacting loops:

K_c = maximum PID gain possible with other loops in manual (dimensionless)
K'_c = retuned PID gain due to interaction (dimensionless)
T_i = minimum PID integral (reset) time possible with other loops in manual (sec)
T'_i = retuned PID integral (reset) time due to interaction (sec)
T_d = maximum PID derivative (rate) time possible with other loops in manual (sec)
T'_d = retuned PID derivative (rate) time due to interaction (sec)
λ = relative gain for subject loop (dimensionless)

10.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and a step change in load (ΔF_L) or setpoint (ΔSP) to loop 1. Table 10.1 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The tests are for two loops (2 × 2 RGA) that are either half coupled or full coupled. Load disturbances and setpoint changes are made to loop 1. PID 1 has a process response that is five times slower (Figures 10.3a, b, c, d and 10.4a, b, c, d) or two times slower (Figures 10.5a, b, c, d

Table 10.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	Effect
10.3a, b, c, d	Moderate self-reg.	1 dimensionless	10 2	20 4	$\Delta F_L = 10\%$	Coupling and Lambda
10.4a, b, c, d	Moderate self-reg.	1 dimensionless	10 2	20 4	$\Delta SP = 10\%$	Coupling and Lambda
10.5a, b, c, d	Moderate self-reg.	1 dimensionless	10 5	20 10	$\Delta F_L = 10\%$	Coupling and Lambda
10.6a, b, c, d	Moderate self-reg.	1 dimensionless	10 5	20 10	$\Delta SP = 10\%$	Coupling and Lambda

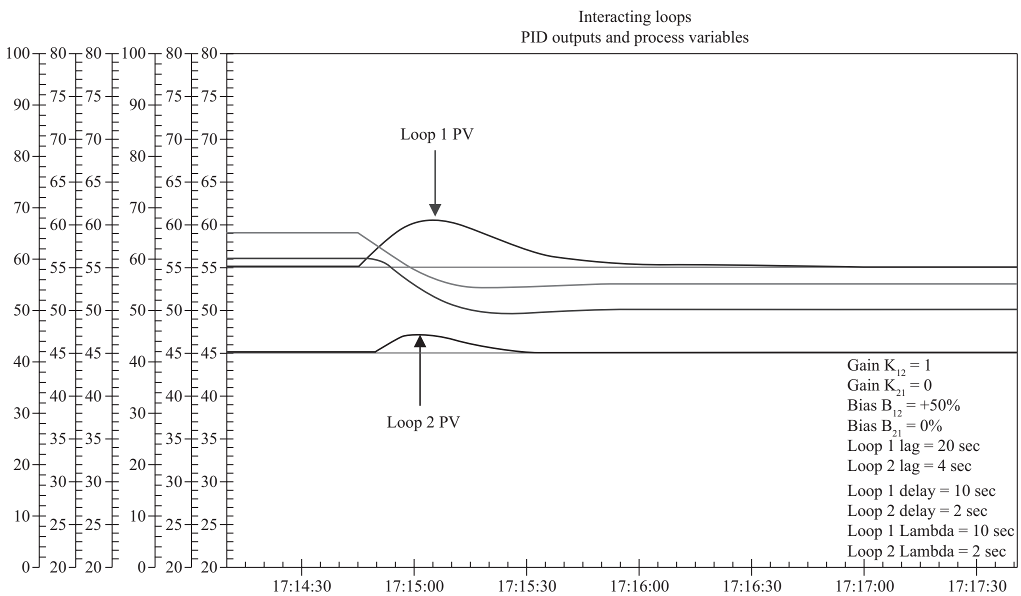


Figure 10.3a. *Load* response in *half* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 2 *seconds*.

and 10.6a, b, c, d) than the process response for PID 2 to provide some initial separation of dynamics before tuning. The half coupled case is the result of setting the interaction gain (K_{21}) of the cross term from the fast to the slow loop equal to zero. This would correspond to a prefect half decoupler where the fast loop PID 2 output (manipulated flow) is a feedforward to the slow loop PID 1 output (e.g., flow feedforward) as discussed in the distillation column and static mixer examples.

The load and setpoint response is evaluated for both half and full coupled cases. For the half coupled cases, aggressive tuning (PID 1 and PID 2 Lambda = 1 dead time) is used for both

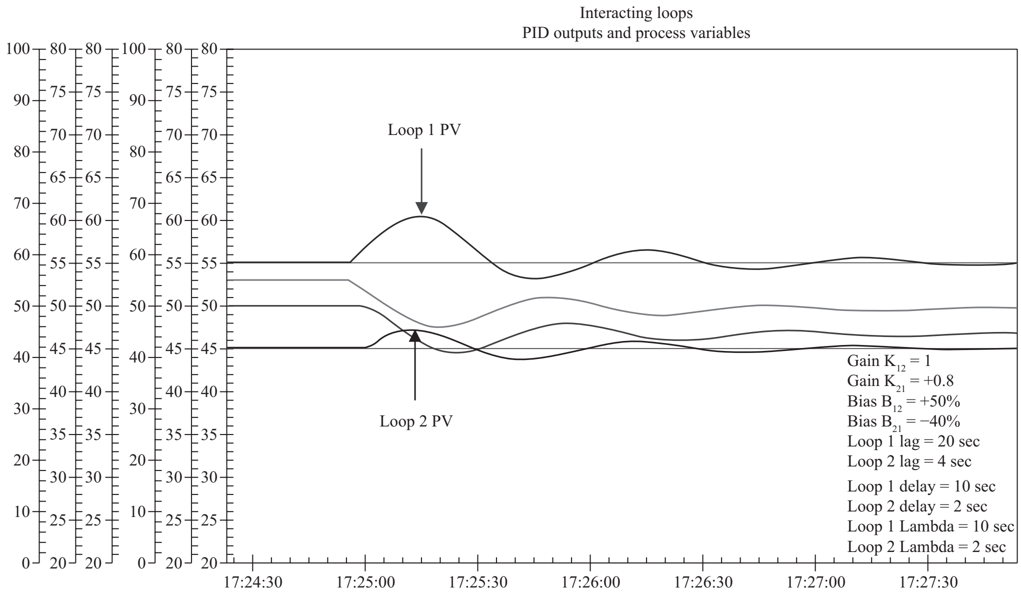


Figure 10.3b. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 2 *seconds*.

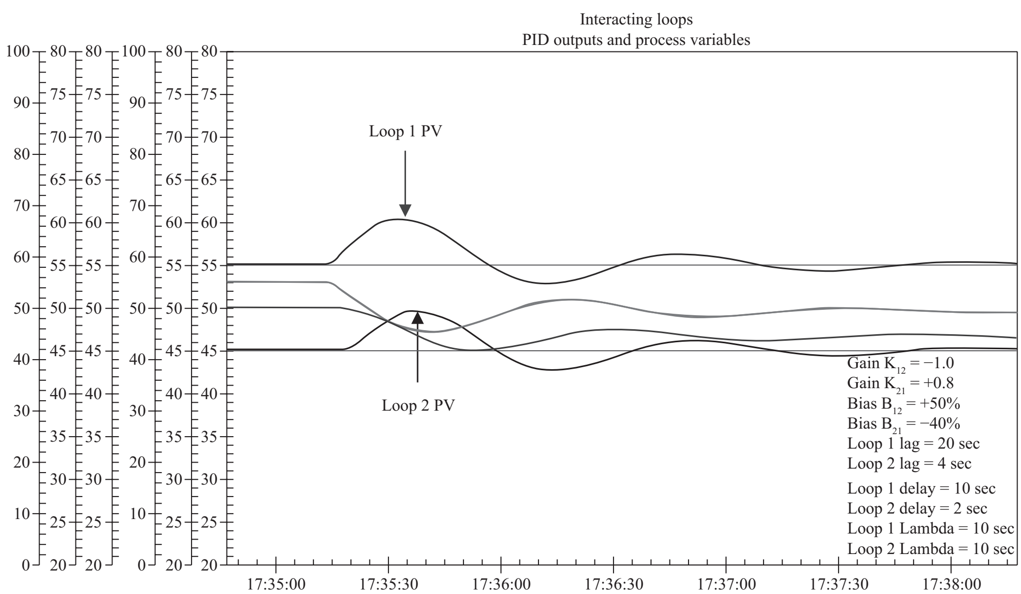


Figure 10.3c. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 10 *seconds*.

PID loops. For the full coupled cases, the effect of slower tuning is individually tested by making the Lambda of the fast loop five times slower (PID 2 Lambda = 5 dead times) or the Lambda of the slow loop three times slower (PID 1 Lambda = 3 dead times). Moderate self-regulating processes are used so that the final response is seen quickly after the dead time and Lambda is the closed loop time constant of the setpoint response.

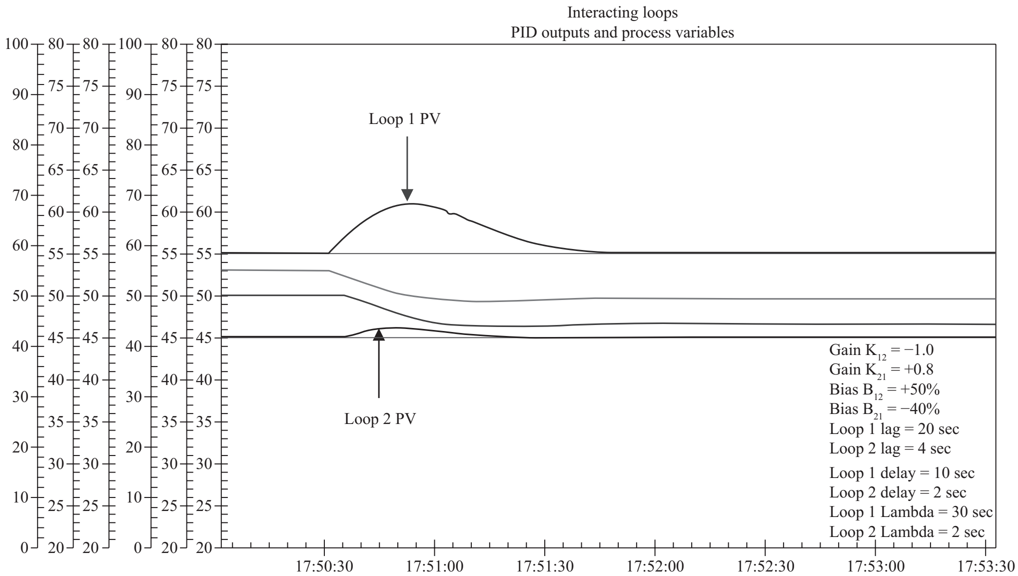


Figure 10.3d. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 30 *seconds* and loop 2 Lambda 2 *seconds*.

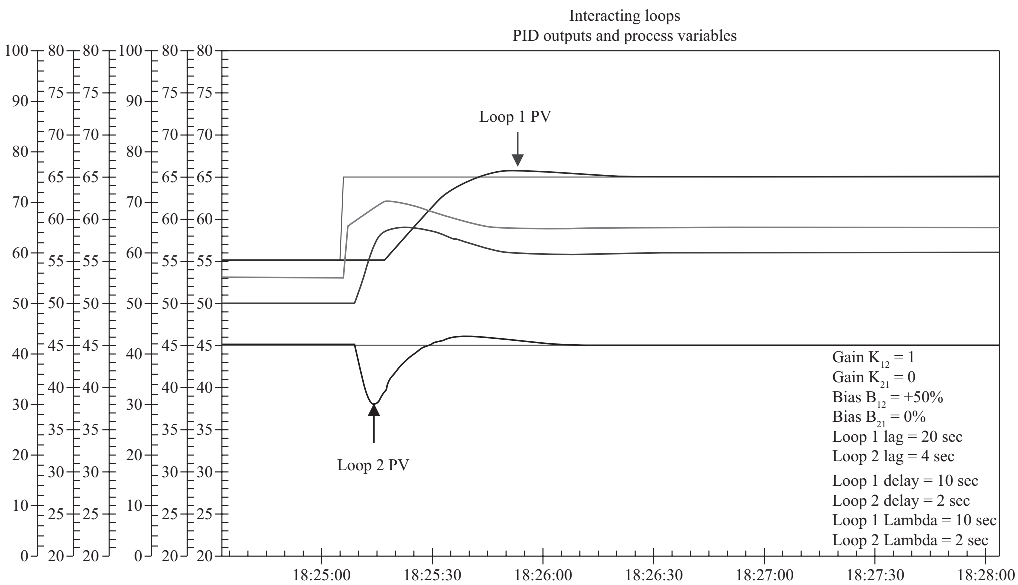


Figure 10.4a. *Setpoint* response in *half* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 2 *seconds*.

Figures 10.3a, 10.4a, 10.5a, and 10.6a show tests for a half coupled system where the fast loop 2 process input saw 100 percent of the change in the slow loop 1 output (K_{21} = 1.0) but there was no effect of the fast loop 2 on the slow loop 1 (K_{21} = 0). This would be about the same result if a signal of loop 2 output change was used as a feedforward to loop 1 perfectly compensating for the effect of loop 2 on loop 1. Note that loop 1 does well but loop 2 takes a hit

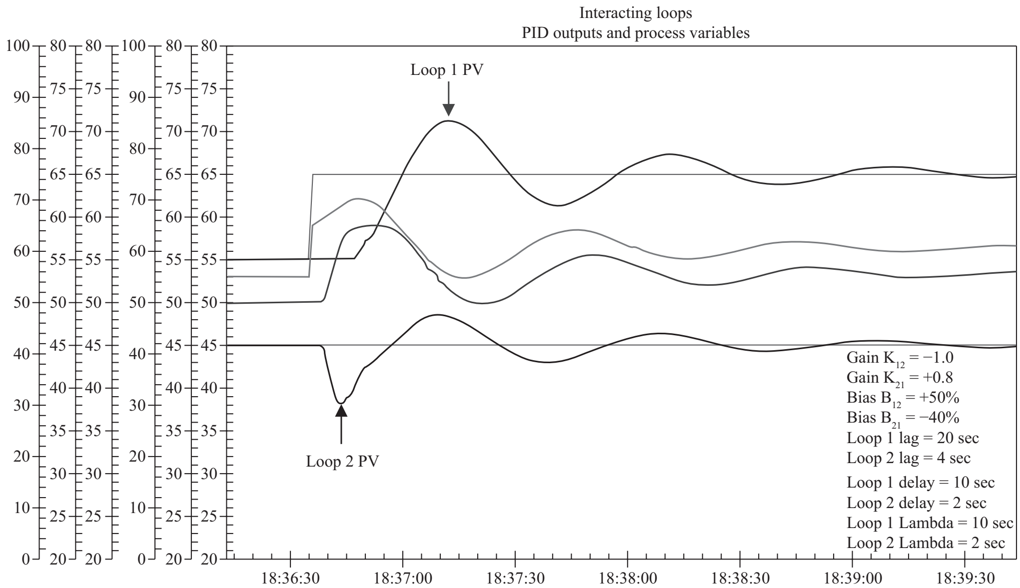


Figure 10.4b. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5 × loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 2 *seconds*.

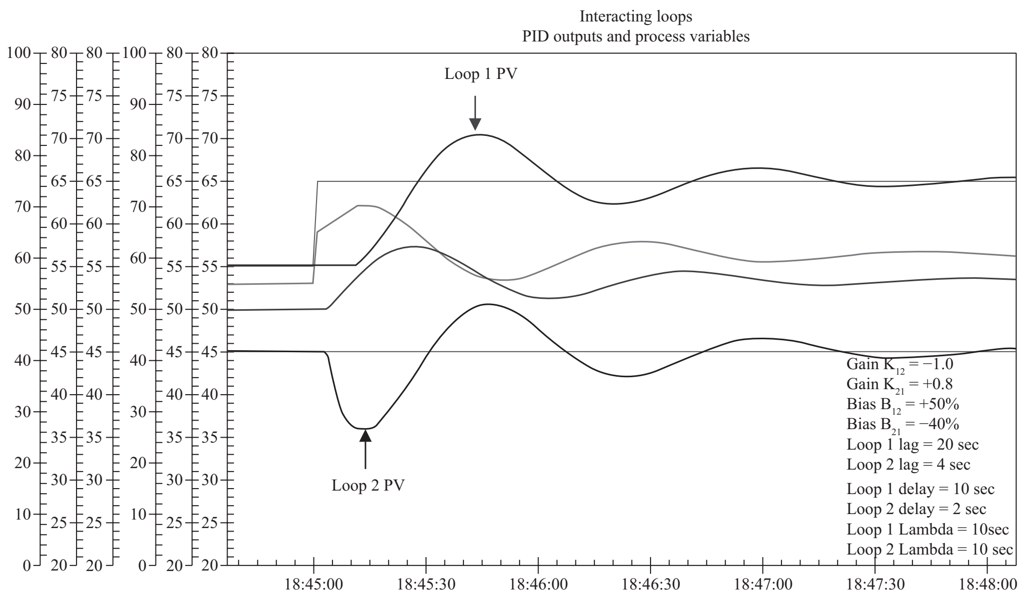


Figure 10.4c. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5 × loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 10 *seconds*.

for a load or setpoint change. Loop 2 uses a Lambda equal to the dead time (minimum Lambda) so the effect is minimized.

The Figures 10.3b, c, d, 10.4b, c, d, 10.5b, c, d, and 10.6b, c, d tests were for a full coupled system where the slow loop process input now saw 80 percent of the change in the fast loop output and as before the slow fast loop process input saw 100 percent of the change in the slow

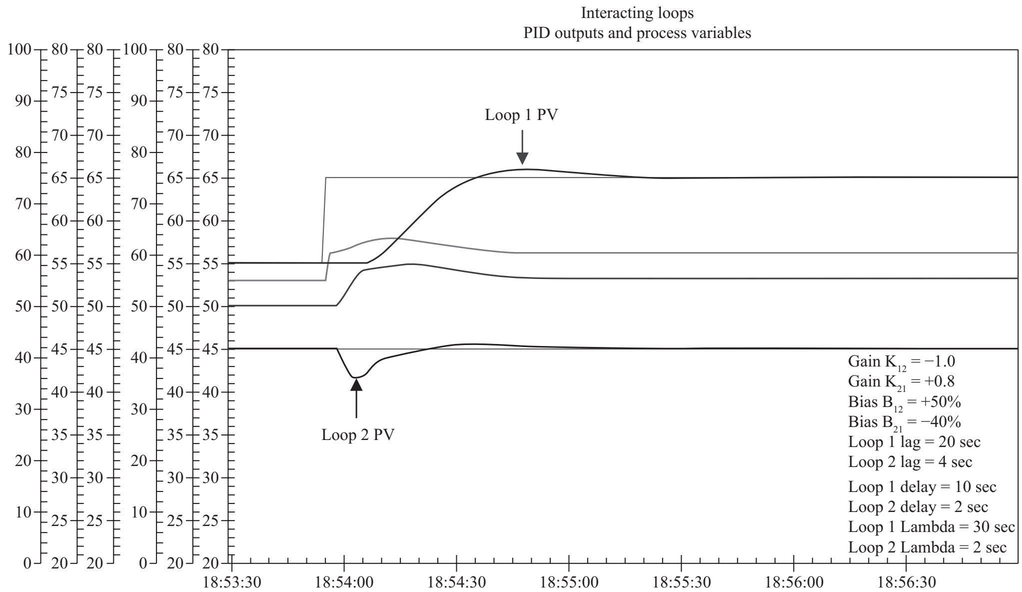


Figure 10.4d. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 5× loop 2 with loop 1 Lambda 30 *seconds* and loop 2 Lambda 2 *seconds*.

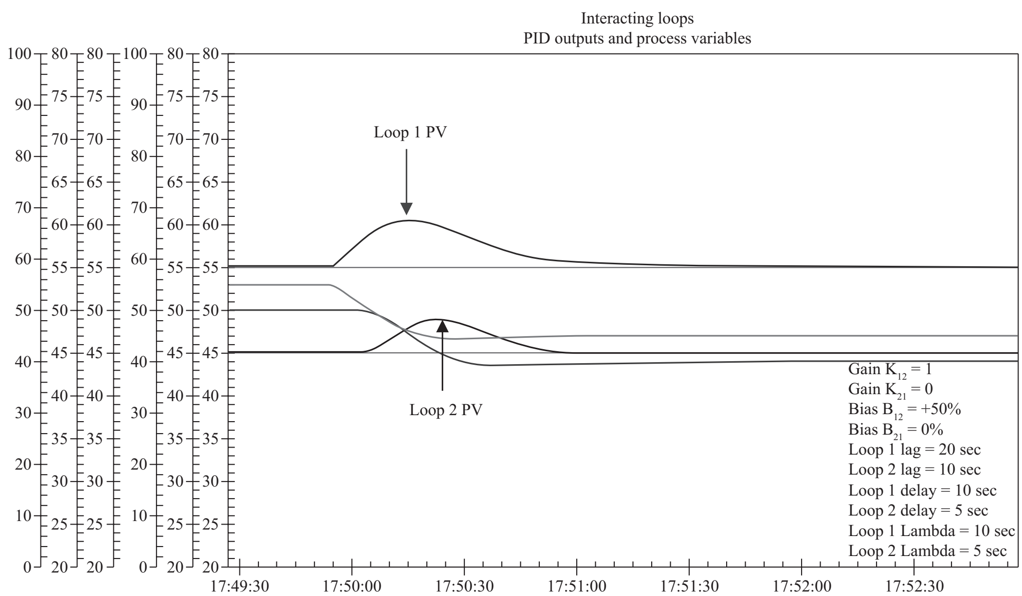


Figure 10.5a. *Load* response in *half* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 5 *seconds*.

loop output. The oscillations seen in Figures 10.3b, 10.4b, 10.5b, and 10.6b for full coupled interactions are quite slow and severe in both loops for the aggressive tuning settings used in the half decoupled case.

Increasing the Lambda from one dead time to five dead times in PID 2 made the responses slightly worse for the slow loop and more noticeably worse for the fast loop in Figures 10.3c

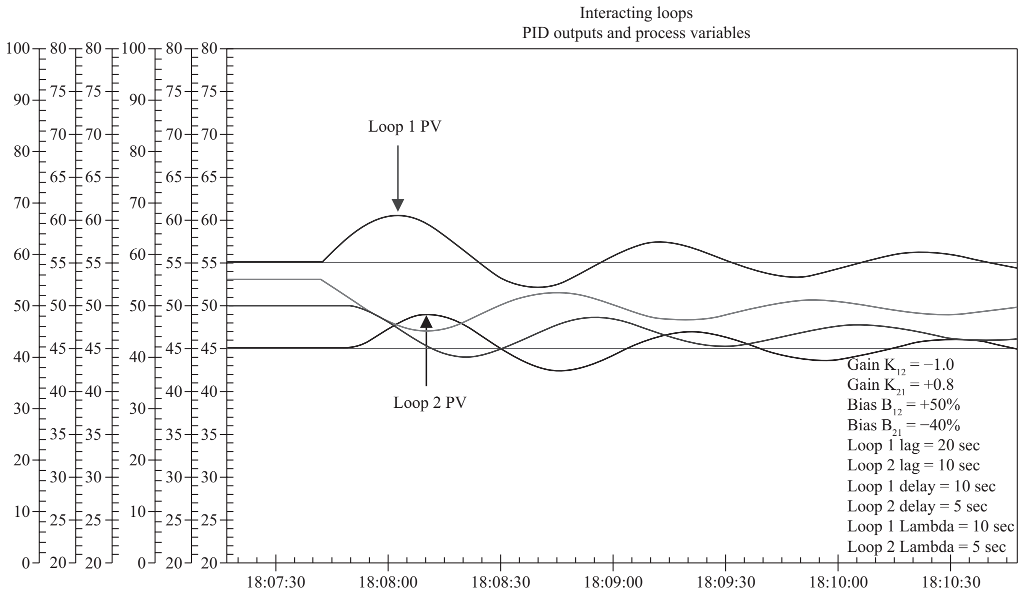


Figure 10.5b. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 5 *seconds*.

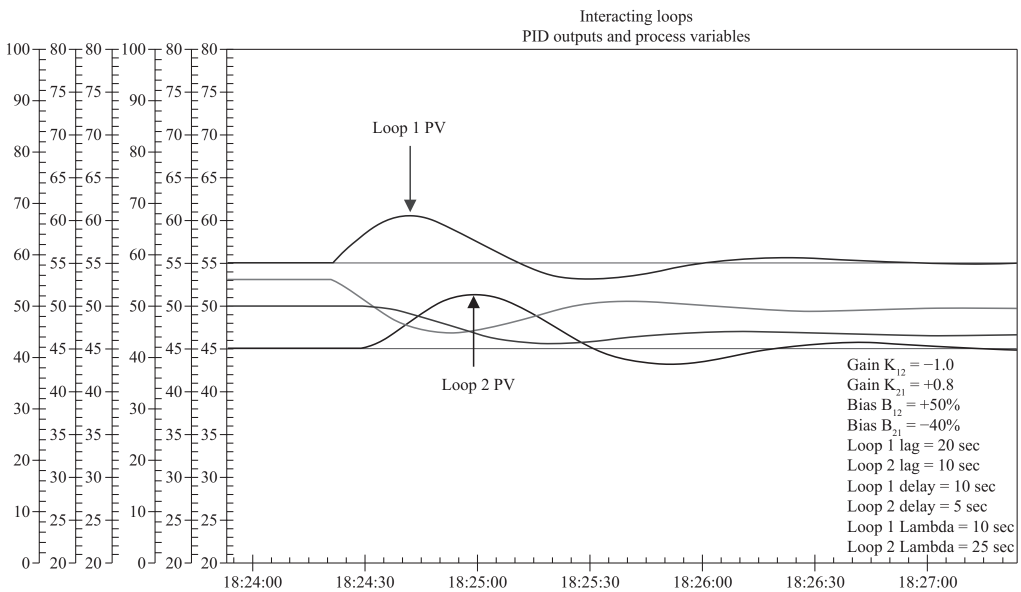


Figure 10.5c. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda *10 seconds* and loop 2 Lambda *25 seconds*.

and 10.4c where the loop 2 process was 5× faster than the loop 1 process. By detuning PID 2 we had made the closed loop response of fast loop close to the speed of response of the slow loop making the interaction more oscillatory.

The opposite was true for Figures 10.5c and 10.6c, where the loop 2 process was just 2 times faster than the loop 1 process. The detuning of the PID 2 made this loop response slower

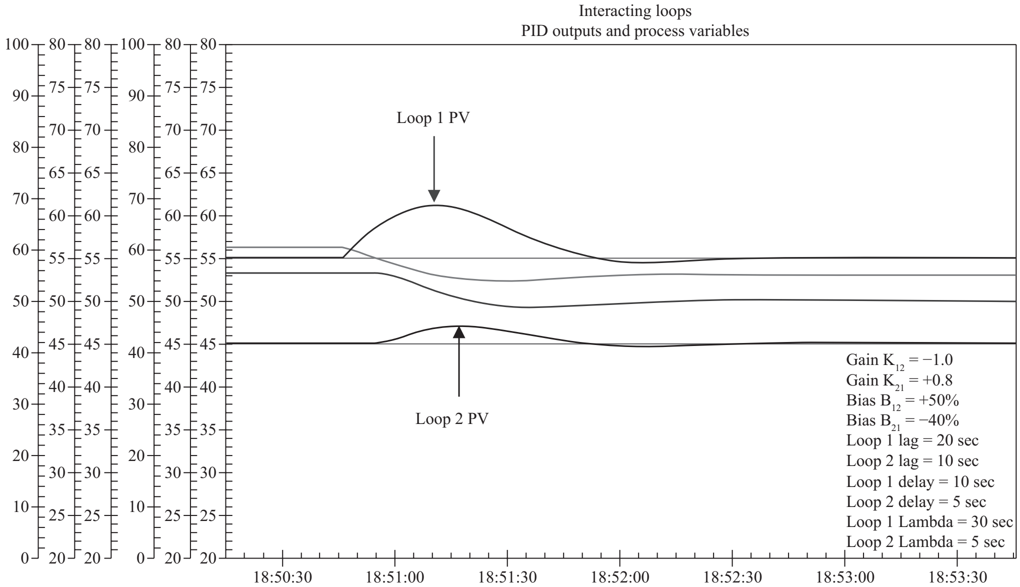


Figure 10.5d. *Load* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 30 *seconds* and loop 2 Lambda 5 *seconds*.

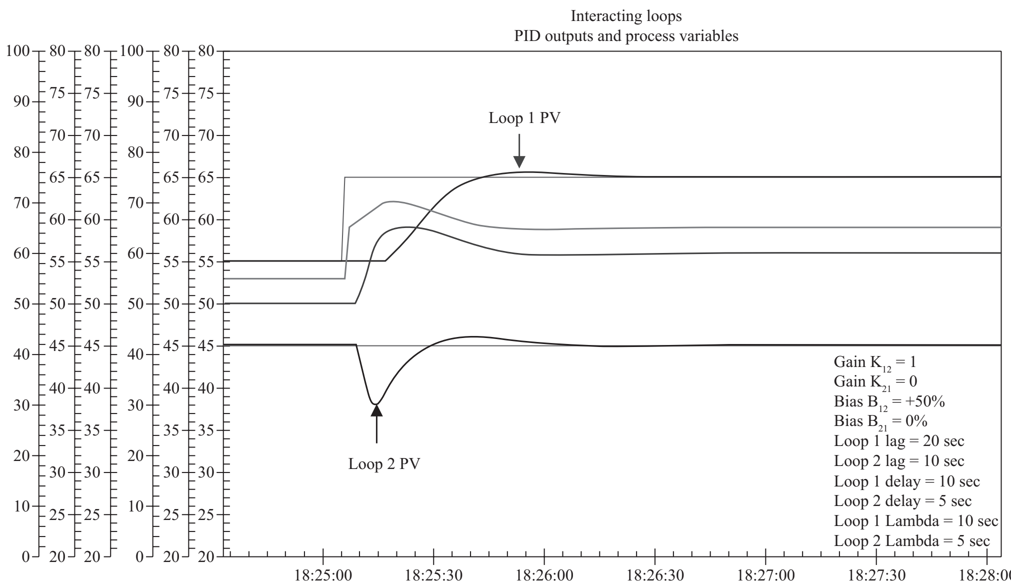


Figure 10.6a. *Setpoint* response in *half* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 5 *seconds*.

than the PID 1 loop response providing some separation of dynamics making the response less oscillatory than if both loops were tuned aggressively for a loop 2 process that was just slightly faster than the loop 1 process. The interaction is greater with a much larger disruption to the fast loop than the half decoupled case or slow loop detuned case, but the oscillations decay quickly. The oscillations could have been eliminated by further detuning but the PID 2 integrated error and time to return setpoint would be larger.

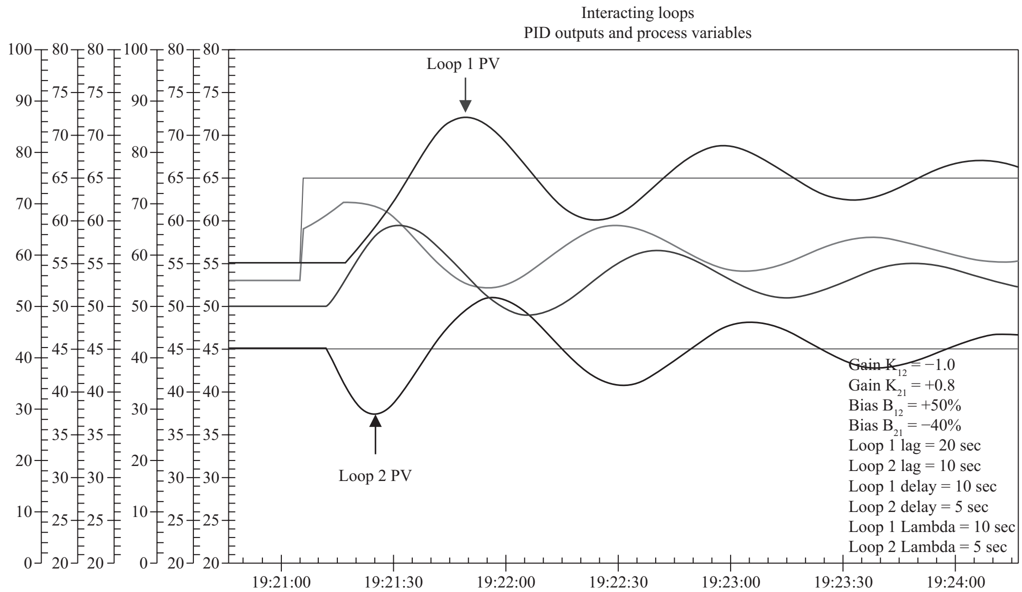


Figure 10.6b. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 5 *seconds*.

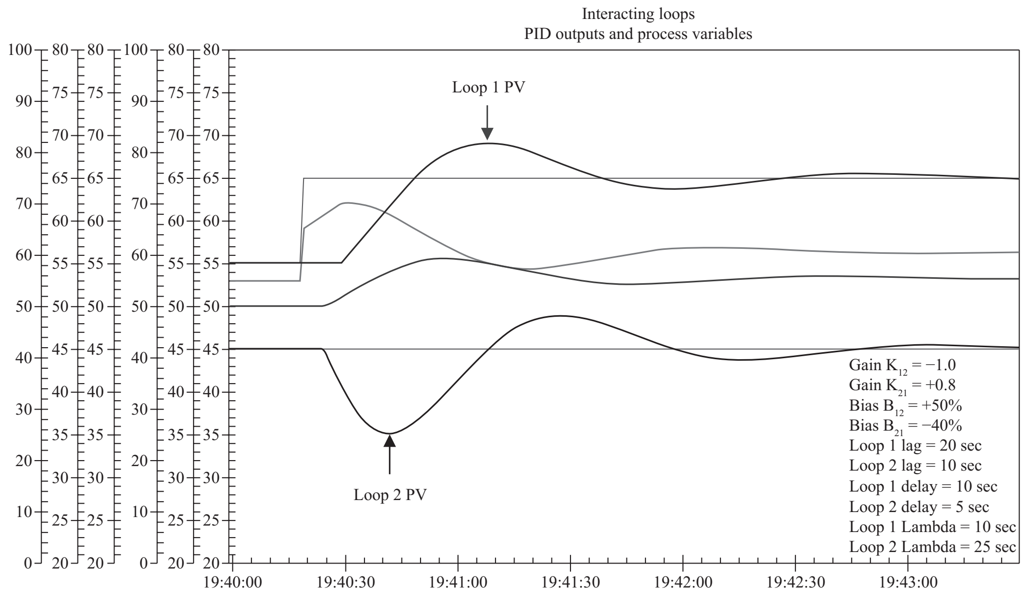


Figure 10.6c. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2 × loop 2 with loop 1 Lambda 10 *seconds* and loop 2 Lambda 25 *seconds*.

When the slow loop Lambda is increased from one dead time to three dead times in Figures. 10.3d, 10.4d, 10.5d, and 10.6d, the resulting responses are almost as good as a perfect feedforward for half decoupling. For detuning rather decoupling, the peak and integrated error is slightly larger for the slow loop but are less for the fast loop. Detuning is a simple solution compared to the identification of decoupling dynamics. The performance shown for the half decoupled test case is difficult to achieve in practice because this case requires perfect feedforward gain and lead-lag settings.

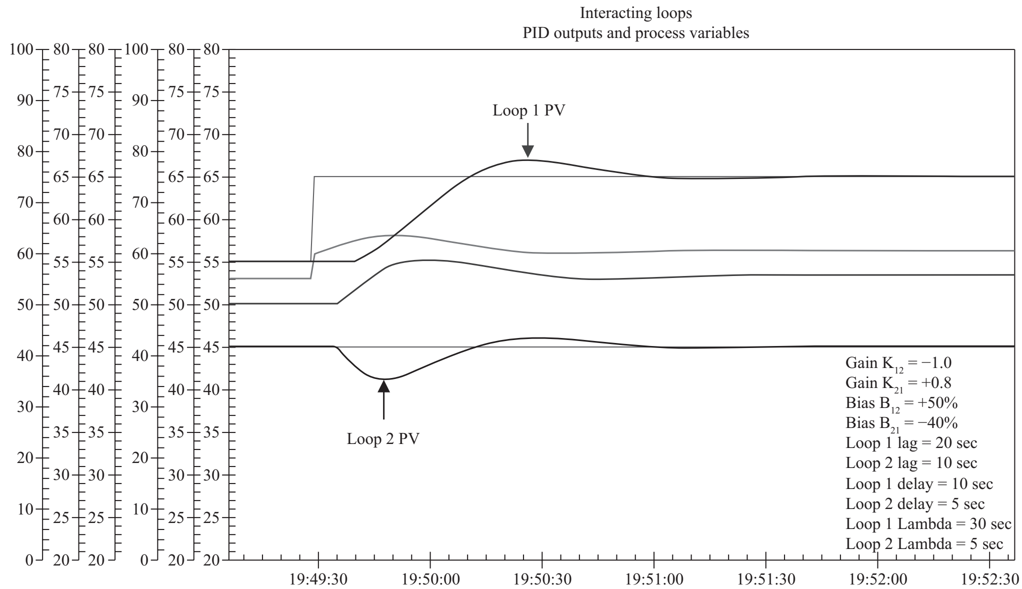


Figure 10.6d. *Setpoint* response in *full* coupled moderate self-regulating processes loop 1 *lag and delay* 2× loop 2 with loop 1 Lambda 30 *seconds* and loop 2 Lambda 5 *seconds*.

KEY POINTS

1. The interactions in PID applications on well-mixed liquid volumes can in most cases be analyzed by different cases of a 2 × 2 RGA.
2. Inventory control loops (e.g., gas pressure and liquid level) should be closed (PIDs in automatic) for any interaction analysis.
3. Tight liquid and gas pressure control is critical for preventing the propagation of disturbances from one user to another user of a utility or raw material.
4. Tight level control for column distillate receiver by manipulation of reflux control is essential for a manipulated change in distillate flow by a temperature controller to translate into a change in reflux flow that affects the column. Tight level control by manipulation of reflux flow also provides internal reflux control where a change in internal traffic temperature causes a change in overhead vapor flow which through the level loop results in a compensating change in reflux flow.
5. A relative gain slightly less than one is ideal. In general, the pairing that provides gains close to but not exceeding one is best.
6. A pairing with a negative relative gain is not a safe solution due to positive feedback and resulting instability.
7. A loop with a relative gain greater than one is not a good candidate for decoupling because of the hyper sensitivity to decoupling errors.
8. A loop with a relative gain greater than one will exhibit a slow setpoint response. The approach to the new setpoint falters and is protracted. The prepositioning and holding of PID outputs till a future PV is projected to reach setpoint is useful for achieving a good setpoint response with minimal disruption.

9. Tuning a fast loop faster or slow loop slower to provide a separation of dynamics that is greater than 3:1 can be extremely effective. The first choice is to tune a fast loop faster but nonlinearities and unknowns may preclude this option. The next choice of tuning the slow loop slower generally adds robustness.
10. Decoupling is often simply flow feedforward because the most common manipulated variable that ends up as a process input is flow.

CHAPTER 11

Cascade Control

11.1 INTRODUCTION

Nearly every control loop has cascade control by virtue of a valve positioner. Often there is a triple cascade as a result of the addition of a flow control loop with a remote setpoint from a composition, level, pH, or temperature controller. Cascade control is beneficial in most applications but there can be implementation problems that lead to oscillations. In most cases, there are fixes to prevent an oscillatory response but there are a few instances where a cascade loop is worse than the original single loop. This chapter will discuss how to get the most out of cascade control, benefits, and watch-outs with a special focus on how to recognize when oscillations could reduce process performance. The solutions are often quick fixes but there situations where a level of cascade needs to be abandoned.

11.1.1 PERSPECTIVE

In cascade control, the output of an upper loop proportional-integral-derivative (PID) is the setpoint to a lower loop PID. While the discussion often centers on a single cascade, there can be multiple cascades. If you consider that a valve positioner is really a digital valve controller (DVC) and in this DVC there may be cascade control where the output of a position controller is the setpoint for a relay travel or actuator pressure controller, you could have a triple cascade with just a flow controller. If you consider a reactor temperature controller whose output is the setpoint of a jacket temperature controller whose output is the setpoint of a flow controller whose output is the setpoint of a DVC, you have a quintuple cascade. Since today's control valves have positioners, every loop has a cascade control system and most have multiple cascades.

Here we will use the terms *upper* and *lower* to denote the relative position in the cascade control system. Upper loops are typically associated with quality control (e.g., composition, pH, and temperature control) and inventory control (e.g., gas pressure and level). Given that cascade loops are much more common than realized, we need to be aware of when and how cascade can help and what are the potential problems. The upper loop is often termed the primary or master loop and the lower loop termed the secondary or slave loop. While the primary loop may enclose the primary process time constant and the secondary loop may enclose the secondary process time constant, this is not necessarily the case. To avoid confusion as in the location or subscripting of process time constants and to facilitate multiple levels of cascade, we will refer to the primary loop as the upper loop and the secondary loop as the lower loop.

For cascade control to be effective, a lower loop must be faster than the loop immediately above it that is the source of its setpoint. There is an unofficial cascade rule that a lower loop must be five times faster than an upper loop. While stated in terms of a single cascade, the rule is applicable as you progress to higher loops. We will focus here on a single cascade with an upper and a lower loop.

By faster, we mean the closed loop response of the lower loop must be smaller than the upper loop. This can be achieved by detuning the upper loop but the best overall results are obtained by making the lower loop faster. Since the closed loop time constant and arrest time and the ultimate period are all a function of the total loop dead time, a lower loop dead time that is less than the upper loop dead time enables the best results if the tuning takes advantage of the opportunity offered. The ultimate limits to the peak and integrated errors decrease as the ratio of the lower to upper loop dead time decreases.

If the cascade rule is violated, peak error and integrated error suffers but oscillations can be prevented by tuning the lower loop faster and the upper loop slower. Here the closed loop time constant or arrest time of the lower and upper loops are made smaller and larger, respectively. Also, the external reset feedback of the lower loop process variable (PV) to the upper loop PID can be used to prevent the burst of oscillations when an upper PID tries to change the setpoint of a lower PID, faster than the lower loop can respond.

Lower loops can make the upper loop process gain more linear and ultimate period smaller, compensate for lower disturbances before they appreciably affect the upper process, deal with valve and variable frequency drive (VFD) deadband, and resolution limits on a faster more direct basis, and offer flow feedforward (e.g., flow ratio control) when the lower loop is flow. If the lower loop does not give these advantages and is not as fast as needed, the cascade loop might best be abandoned in favor of a single loop with the rate time set equal to the lower loop process time constant.

In the days of analog electronic controllers and pneumatic positioners, the cascade rule was violated by putting a positioner on a fast loop (e.g., liquid pressure and flow). Volume boosters were recommended based on Nyquist plot studies. There were several factors not considered in this theoretical analysis that are not valid today. The net result is that positioners are recommended on all control valves.

Cascade control has the interesting property of converting a detrimental dynamic term in a single loop to a beneficial term in a cascade loop. If you consider the original single loop, a secondary time constant can seriously degrade performance either taken as an increase in dead time in self-regulating processes or even worse as a time constant in integrating and runaway processes. If a lower loop is created enclosing the secondary time constant, this time constant is effectively removed from the upper loop and becomes the largest time constant in the lower loop. A larger secondary time constant increases the maximum allowable PID gain in both loops by increasing the time constant to dead time ratio in both loops. The upper loop can become faster (ultimate period smaller) and better able to deal with upper loop disturbances. The lower loop can become more aggressive (ultimate gain smaller) and better able to deal with lower loop disturbances.

The greatest improvement is seen for disturbances that are inputs to the lower process. Potentially the lower loop can correct for them before these disturbances affect the upper loop. Lower flow loops can correct for pressure disturbances before they affect the upper composition, level, pH, or temperature loop. Lower jacket temperature loops can correct for changes in coolant temperature before they affect a reactor temperature loop.

As the process time constant in the upper loop increases and the lower loop dead time decreases, the lower loop period decreases and the filtering action of short term transients by the upper process time constant increases. In fact, lower loops tuned for an oscillatory response may provide the best disturbance rejection for upper loop process time constants that are more than 10 times the lower loop dead time. An interesting unexpected example is the tuning of a DVC in a temperature loop to be oscillatory by the addition of integral action in the DVC. The amplitude of the fast oscillations (e.g., two second period) in the DVC, filtered by large time constant (e.g., 60 minutes) in the temperature loop are less than the resolution limit of the sensor. These oscillations are preferable to an offset in the DVC that will show up in the temperature. However such tuning will wear out the valve. The better solution is a valve with a much smaller stick-slip (resolution and threshold sensitivity limit) and hence smaller offset.

The upper loop has a smaller ultimate period than the original single loop. The lower loop can also isolate nonlinearities from the upper loop. Flow loops isolate the nonlinearity of the installed characteristic. Jacket and coil temperature loops isolate the process nonlinearity where the process gain and dead time increase as the jacket flow decreases as well as valve nonlinearities and discontinuities at the split range point from the transition from heating to cooling and vice versa. These loops can also enforce temperature limits to prevent hot or cold spots on heat transfer surfaces, particularly important for crystallizers and bioreactors.

If the cascade rule is obeyed, the upper PID gain and rate time can be increased and the reset time decreased. Consequently, the cascade loop can better handle disturbances that originate in the upper loop than in the original single loop.

We can make the lower loop faster by making the measurement, controller, and final control element (e.g., valves and VFDs) faster. For measurements this corresponds to decreasing the delay and lag of the sensor (e.g., thermowell or electrode) and of the transmitter (e.g., damping and wireless default update rate). For controllers, this means making the PID execution rate and signal filter time faster. For final control elements this translates to decreasing the deadband and resolution and threshold sensitivity limits. It also means keeping speed control in the VFD rather than moving it to the Distributed Control Systems (DCS) so the controller execution is faster. For valves, we need to decrease the pre-stroke dead time and stroking time. For VFDs, we need to increase the allowable rate of change of speed.

The upper closed loop time constant or arrest time (Lambda) should be about five times greater than the lower loop Lambda. If this cannot be done by making the lower loop faster, it can be done by tuning the upper loop Lambda to be larger.

The addition of a lower flow loop enables flow feedforward. A watch-out is whether the flow measurement has enough rangeability. Magnetic flow meters have good turndown (e.g., 50 to 1) and Coriolis meters have great turndown (e.g., 200 to 1). Differential head meters may get too noisy and vortex meters signals may drop out at low flows in which case logic needs to be added to the cascade loop to switch to a flow measurement computed from valve position or VFD speed.

VFDs with negligible deadband and setpoint rate limits may be necessary for polymer pressure control to prevent a serious violation of the cascade rule. Polymer pressure control often needs to be tight to maintain polymer properties. Polymer pressure process dynamics are faster than valve dynamics.

Process flow diagrams, process and analysis, production metrics, data analytics, and simulations (e.g., steady state and dynamic) all depend upon getting the flows right. Lower flow loops enable greater process knowledge and process control improvement.

11.1.2 OVERVIEW

Cascade control is most effective when the lower loop is five times or more faster than the upper loop (cascade rule). If the lower loop is too slow, the upper PID must be prevented from changing the lower PID setpoint faster than the lower loop can respond to prevent a burst of oscillations for large and fast disturbances or setpoint changes.

Cascade control can greatly reduce the effect of disturbances entering the lower loop. If the cascade rule is not violated, disturbances are also reduced that enter the upper loop. The ultimate period of the cascade loop is less than the ultimate period of a single loop with the same dynamics.

Lower loops can also isolate nonlinearities from the upper loop and better enforce limits on lower PVs (e.g., coolant temperature). Lower flow loops enable flow feedforward and process simulation, metrics, and analysis.

11.1.3 RECOMMENDATIONS

1. Use valve positioners (DVCs) on all control valves.
2. If the valve stroking time (time for 100 percent stroke) is significantly greater than the reset time of the PID manipulating the valve, add volume booster(s) on valve positioner output(s). Open the booster bypass enough when stroke testing the valve to prevent high frequency cycling of valve position (e.g., 1 cps).
3. Tune valve positioner for fast response avoiding the use of integral action.
4. Use lower flow loops wherever possible to compensate for nonlinear installed flow characteristics and to provide the measurements needed for mass, mole and energy balances, and cost analysis.
5. Use jacket and coil temperature control loops for bioreactor, crystallizer, and chemical reactors. If boiler feedwater is used for high temperature highly exothermic reactors, use a pressure control loop on coil outlet.
6. Make the measurement and execution of the lower loop as fast as possible. Ensure the sensor type, condition, and installation minimizes the sensor delay and lag. Minimize transmitter damping and signal filters. Use a PID execution rate that is at least five times faster in the lower loop than in the upper loop.
7. Tune the lowest loop first, then the next lowest loop, and ending up tuning the upper most loop last. For example, tune the valve positioners first, then the flow loop, then the jacket temperature loop, and finally the reactor temperature loop.
8. Tune lower loops to be as fast as possible emphasizing more proportional action than integral action. Use a PID structure that has proportional action on error.
9. Do not use setpoint filters on lower loops. Use Lambda tuning for coordination of flow loops (e.g., maintaining stoichiometry for inline blending and reactions).

10. If the lower loop cannot be made five times faster than the upper loop, tune the upper loop slower by making the upper loop Lambda five times larger than the lower loop Lambda. Consider abandoning cascade control using either the lower or upper loop PV for single loop control depending upon whether the disturbance size and speed is more problematic in the lower or upper process. The temperature control of some bioreactors is best done by simple jacket temperature control because the jacket volume is comparable to the process volume and process disturbances from cell growth are incredibly small and slow.
11. Use external reset feedback of lower loop PV to upper loop PID so the upper loop PID output does not try to change faster than the lower loop can respond.
12. If flow measurement rangeability is insufficient, there may need to be a switch to direct throttling of the control valve, a common practice in boiler drum level control at low steam rates and start up. A better solution is a computed flow from the installed valve characteristic and a bumpless transition to an inferential flow measurement maintaining cascade control.

11.2 CONFIGURATION AND TUNING

The implementation and commissioning of a cascade control system can make or break the application. Here we have a list of things to make sure this is addressed in the configuration and tuning of the PIDs used in cascade control.

1. Decide whether cascade control is the right solution. In general, do not use cascade control where the lower loop cannot be made fast enough and slowing down the upper loop is detrimental. For example, cascade control of liquid or polymer pressure control to flow control must be avoided. Similarly cascade control of gas pressures in inches of water (e.g., furnace pressure control) to flow control can be disastrous. The output of these pressure controllers should go directly to a variable speed drive or a valve or damper with volume boosters.
2. Set the upper PID output scale to match the lower PID setpoint scale. If the PID output is in engineering units, make sure the same engineering units are used for the upper PID output and lower PID setpoint. Note that PID output scale could only be in percent (e.g., 0 to 100 percent) in most analog controllers and 1980s vintage DCS. In migration to a modern day DCS, it is critical to make sure the upper PID output scale and limits are converted to the engineering units of the lower PID. While the PID algorithm internally uses signals in percent of scale, the PID output besides the PID input and setpoint are configured and displayed in engineering units as noted at the bottom of Figure 11.1, the block diagram for a cascade loop.
3. Set output limits and anti-reset windup (ARW) limits of the upper loop PID to match the setpoint limits of the lower loop PID. If the upper loop PID output scale is in percent, convert the lower PID setpoint limits to percent of scale for the upper PID.
4. Use a PID structure of proportional-integral (PI) on error and D on PV for all lower loops.
5. Setup external reset feedback (e.g., dynamic reset limiting) so that the upper loop output cannot change faster than the lower loop can respond. For valve positioners, this may

not be possible or necessary for small sliding stem throttling valves because the stroke is precise and the slewing rate is so fast. For large valves or valves with backlash or stick-slip, add a fast readback of actual valve position as the external reset feedback (e.g., BKCAL_IN). For lower process loops, use the measurement of the PV as the external reset feedback. For VFDs with tachometer feedback, use the speed as the external reset feedback. Make sure the external reset feedback is properly conveyed back through configuration blocks between the lower PID output and upper PID or analog output blocks for valves and variable speed drives (e.g., blocks for signal characterization, signal selection, or for split ranged operation).

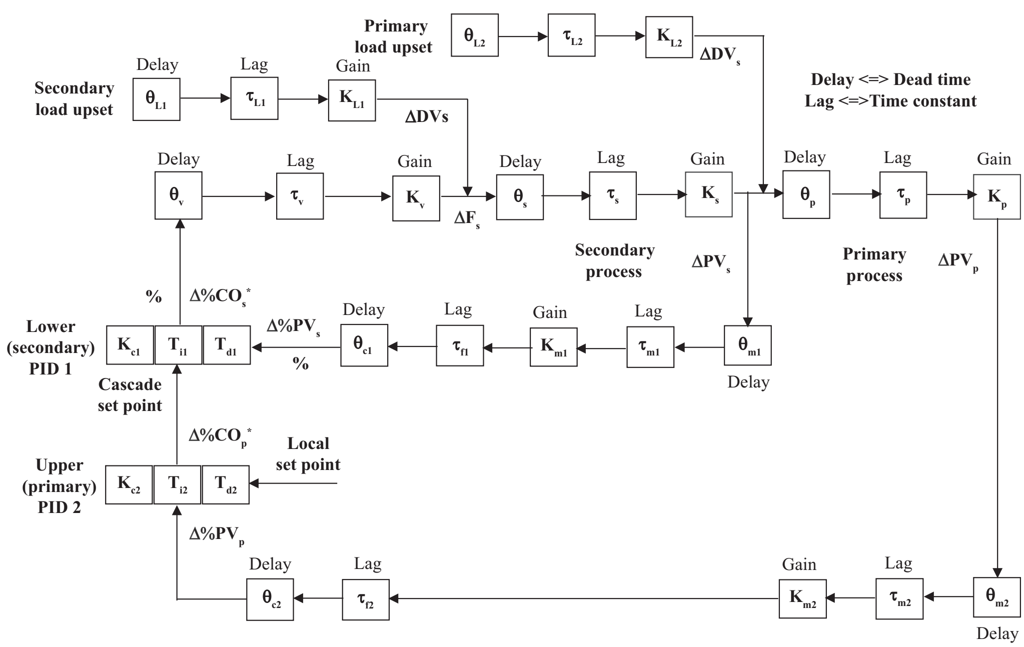
6. For large valves and dampers (e.g., 12 inch or larger) being manipulated by a fast process loop (e.g., flow or pressure loop), add volume booster(s) to the valve positioner output(s) and adjust the booster bypass to eliminate any high frequency oscillations. If there is no integral bypass valve in the booster, one must be added.
7. For VFDs with speed control, keep the speed control in the field and fast. Taking the speed control into the DCS makes this lower loop too slow especially for flow and pressure control.
8. Tune the valve positioner (e.g., DVC) or speed controller first to have a fast response. Do not use integral action unless there is a special reason (e.g., valve is manipulated by a temperature loop and the resulting limit cycle from stiction after filtering by a large process volume reduces the temperature offset).
9. Set the lower loop execution, filter time, and transmitter damping to be as fast as possible. Tune the lower process PID to be as fast as possible while dealing with nonlinearities and meeting process objectives. Realize that if the lower PID is not at least five times faster than the upper PID, the upper PID whose response is most important must be detuned. Remember that the slope of the installed flow characteristic and hence the local process gain changes by more than a factor of 5:1 for most control valves. Use adaptive tuning or signal characterization or both, if the flow control loop response needs to be improved.
10. For ratioed flows, each flow PID should have the same closed loop time constant (Lambda) with no overshoot or oscillations. If the flows are not coordinated, unbalances can occur that are disruptive to blending operations and can accumulate in back mixed volumes throwing off the stoichiometry reducing yield and product quality. Consequently, each flow controller is tuned to have the same fastest possible Lambda. If the flow loops are ratioed to a setpoint rather than a flow measurement and the Lambdas are identical or a setpoint filter is used, the flow loops should respond in unison to a change in production rate. The PID should be configured so that the feedforward is active with the upper PID in manual so that the lower PID can be operated with manually corrected ratios for startup and analyzer failure.
11. Use proportional action to provide a more immediate response and to get through valve deadband. Some applications benefit from an extremely aggressive lower loop response since oscillations in the lower loop are filtered and not seen in the upper loop due to a large primary time constant (e.g., large liquid process volume). Highly exothermic polymer and specialty chemical reactors have used oscillating jacket temperature loops to achieve tighter reactor temperature control.
12. With the lower loops in cascade mode, tune the upper PID. If the lower loop cannot be made faster than five times the upper loop, use a Lambda for the upper PID that is at least five times larger than the lower PID Lambda. Test the upper PID for the ability to

handle load disturbances and setpoint changes in both directions of about the same size expected. To simulate a load change, put the upper PID in manual just long enough to make a step change in the upper PID output.

11.3 PROCESS CONTROL BENEFITS

The process control benefits of cascade loops are numerous. Here is a list of the ones that come to mind.

1. Lower loop isolates nonlinearities (e.g., valve and process) and stream and utility disturbances (e.g., pressure and temperature) from the upper loop.
2. As seen in Figure 11.1, the lower loop encloses secondary time constant that converts the secondary time constant that would have been a detrimental term in a single loop to being beneficial term as the largest time constant in a lower loop.
3. The cascade upper loop ultimate period is smaller than the original single loop enabling a faster upper loop and better rejection of disturbances in the upper loop.
4. The peak error in the upper loop for a lower loop disturbance can be reduced to be as small as 12 percent for self-regulating, 2 percent for integrating, and 1 percent for runaway of the peak error for a single loop as seen in Figures 11.2a, b, c for a lower to upper dead time ratio of 0.6. For lower dead time ratios the improvement would be even more impressive. The best reduction in peak error for a given dead time ratio is achieved for time constant ratio approaching one where the secondary time constant was as large as



*While the PID algorithm internally uses signals in % of scale, the PID output besides the PID input and setpoint of most modern DCS are configured and displayed in engineering units.

Figure 11.1. Types and locations of dynamics in cascade control loop.

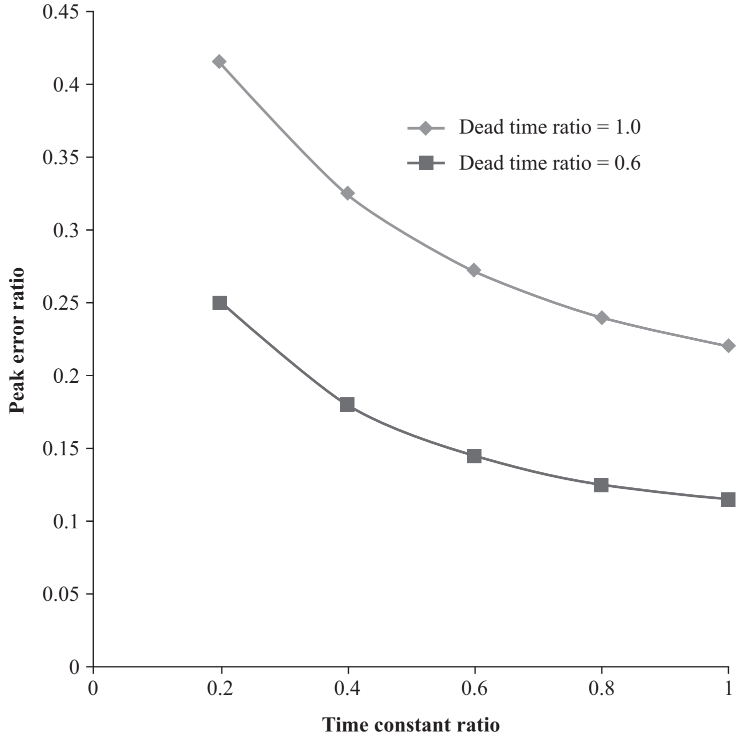


Figure 11.2a. Ratio of cascade to single loop peak errors for *self-regulating* upper loop process.

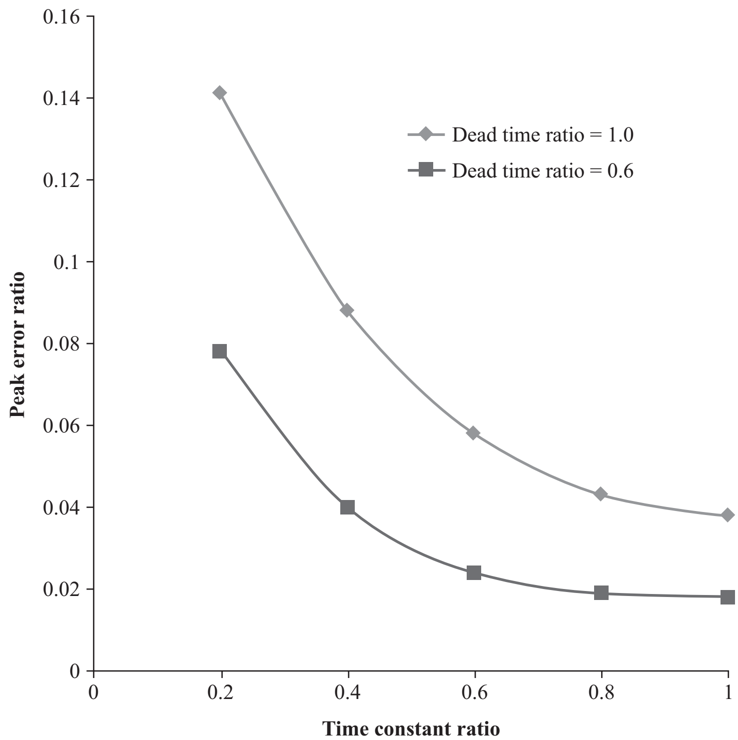


Figure 11.2b. Ratio of cascade to single loop peak errors for *integrating* upper loop process.

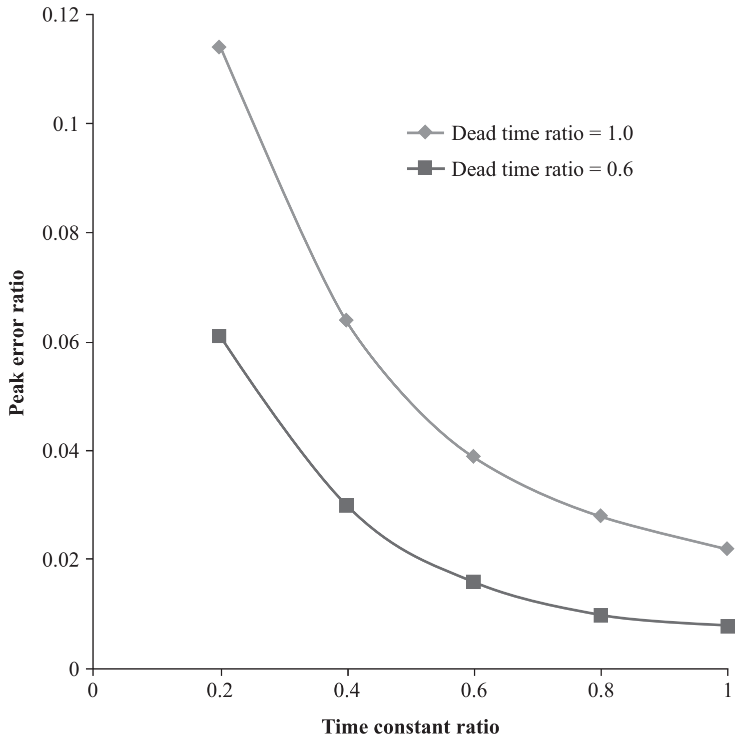


Figure 11.2c. Ratio of cascade to single loop peak errors for *runaway* upper loop process.

the primary time constant. This latter relationship is often not recognized because it is counter intuitive and contradicts the cascade rule. The user must realize the cascade rule pertains to the closed loop response and not the open loop response. Since the ultimate period and Lambda is a factor of the total loop dead time, the loop dead time sets the limits on the closed loop response.

5. Better regulation of process stoichiometry leading to better composition control by the use of lower flow loops and coordinated flow ratio control.
6. More accurate feedforward control to preemptively correct for feed and utility disturbances (e.g., flow and temperature changes).
7. For startup, the cascade control system can be operated with just the lower loop in service (e.g., flow ratio control) until operating conditions are reached (e.g., distillation columns). The feedforward should be configured to be active with the upper loop in manual, which means the lower loop stays in cascade mode.
8. For an upper loop measurement (e.g., analyzer) failure, the cascade control system can be operated with just the lower loop in service (e.g., flow ratio control) until the measurement is fixed.

11.4 PROCESS KNOWLEDGE BENEFITS

The process knowledge benefits of cascade control are not discussed much in the literature but can be just as important and more extensive. The improvement in the recognition and identification of relationships and the creation of models can translate to more intelligent setpoints and operator and process engineers understanding of confusing situations. An increase in process knowledge can be far reaching.

Since nearly all manipulated process inputs are flows, the addition of flow measurements for lower flow control loops offers many advantages. A control loop transfers variability in the PV to the manipulated variable (e.g., flow). If the upper loop (e.g., level, composition, pH, or temperature) is tightly controlled, nearly all of the variability caused by changes in the process is seen in the manipulated flow rather than the PV. The PV in these loops stay right at setpoint. The process knowledge is in the size and pattern of changes in the manipulated flows.

All process simulations need to be compared to the plant and corrected. Process simulations have a difficult time getting the pressure drops and hence the installed flow characteristics right because of the incredible amount of detail on the geometry and characteristics of piping and valve systems needed besides the changes in interior surfaces (e.g., roughness and coatings). The control valve positions in a simulation (e.g., virtual plant) will not match up with the actual plant. The only way to improve simulations that are both doing a good job of control at setpoint is to match up the flows. The addition of flow measurements and flow control in the actual plant enables more accurate process simulations and hence process analysis and improvements.

Virtual plants that are high fidelity of the process including the automation system running real time have been adapted online by matching manipulated flows in the virtual plant to the actual plant. In each case model predictive control (MPC) was setup where the setpoint (target) was an actual plant flow and the controlled variable was the corresponding virtual plant flow. The MPC manipulated variable was a virtual plant process model parameter. The MPC models were identified offline by perturbing the virtual plant parameter. For the identification and adaptation, the upper and lower PID in the cascade control system that end up manipulating the flows must be in service providing closed loop control. In one application the MPC adapted tray efficiencies and pressure drops to match up the virtual and actual plant reflux and steam flows manipulated by the distillation column control system. In another application, the MPC adapted a waste stream acid concentration to match up the virtual and actual plant base reagent flow manipulated by the pH control system.

Nearly all online metrics as to process efficiencies and capacity depend upon flow measurements. For energy release or consumption metrics, temperature measurements are also useful. For example, a jacket inlet temperature subtracted from the jacket outlet temperature and multiplied by the jacket coolant flow can provide an online measurement of heat release and hence conversion for an exothermic reactor. The inlet jacket temperature is passed through a dead time block to mimic the transportation delay in the jacket for synchronization of the inlet and outlet jacket temperatures.

Here is a summary of the benefits of flow measurements beyond cascade control.

1. More linear, accurate, and representative inputs for data analytics and neural networks particularly when loops are tightly controlled. Better analysis of whether correlations represent causes and effects or coincident occurrences.
2. More accurate process simulations for better process analysis and improvements.
3. Greater recognition of the source and path of disturbances and abnormal operation to reduce the consequences and frequency of disturbances and failures.

4. More accurate online process metrics for process efficiency and capacity.
5. Adaptation of parameters in a virtual plant synchronized with an actual plant.
6. More effective interaction analysis by the use of manipulated flows instead of PID outputs in the computation of relative gains for a Relative Gain Array (RGA) (Chapter 10).

11.5 WATCH-OUTS

Most of the problems with cascade control stem from poor PID tuning, measurement problems, configuration mistakes involving the propagation of the external reset signal and the setting of setpoint and output limits, signal filters, and transmitter damping.

1. The setpoint limits and output limits do not match the range of operation needed.
2. The external reset signal on the output of a downstream block (BKCAL_OUT) is not connected to the corresponding input (BKCAL_IN) of the upstream block.
3. A big actuator, large transmitter damping setting, slow wireless update rate, large signal filter, or slow PID execution time cause lower loop to be too slow.
4. Lower measurement has insufficient rangeability losing signal to noise ratio (e.g., orifice flowmeters) or dropping out (e.g., vortex flowmeters).
5. The lower PID Lambda is not five times faster than the upper PID Lambda.
6. Upper measurement (e.g., analyzer) has insufficient reliability to be used for closed loop control.

11.6 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and a step change in lower PID load (ΔF_L) or upper PID setpoint (ΔSP). Table 11.1 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The tests in Figures 11.3a, b, c, d were for a pH to reagent flow cascade control system with wireless transmitters on a static mixer. The wireless update rates were set so that upper PID loop for pH control was approximately five times slower than the lower PID loop for flow control. The default update rates add a dead time to the loop that is half of the update rate. The resulting wireless dead time is much larger than the process dead time. A traditional PID must be detuned to prevent instability. The enhanced PID described in Appendix E enabled tuning that was even more aggressive than the tuning if a wired transmitter was used. The enhanced PID gain was set equal to the inverse of the open loop gain. Each test compares the performance of an enhanced PID to a traditional PID by trending the wireless transmitter output (wireless PV) and the sensor PV input (sensor PV) for each PID. The sensor PV shows the actual PV without any delay or lag. The wireless PV includes the effect of the wireless default update rate.

Table 11.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	Effect
11.3a, c	Moderate self-reg.	1 dimensionless	6 0.5	6 1	ΔSP = 20%	Wireless control
11.3b, d	Moderate self-reg.	1 dimensionless	6 0.5	6 1	ΔF_L = 20%	Wireless control
11.4a, b, c, d, e, f	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	Lower PID Lambda
11.5a, b, c, d, e, f	Moderate self-reg.	1 dimensionless	10 1	40 4	ΔSP = 10%	External reset on
11.6a, b, c, d	Moderate self-reg.	1 dimensionless	10 8	40 32	ΔSP = 10%	Upper PID Lambda
11.7a, b, c, d	Moderate self-reg.	1 dimensionless	10 8	40 32	ΔSP = 10%	External reset on

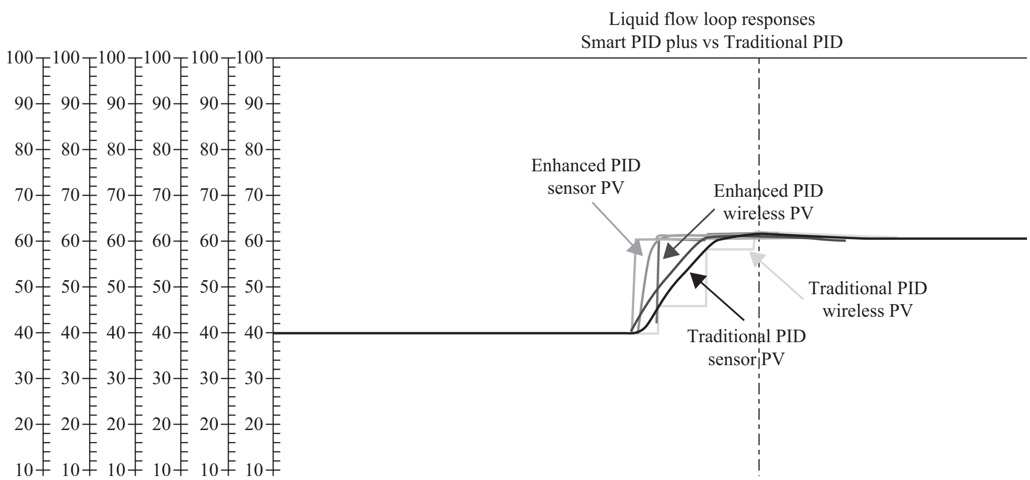


Figure 11.3a. Lower reagent flow loop setpoint response of enhanced PID and traditional PID for wireless default update rate = 16 *seconds*.

There is considerable confusion as to the effect of wireless update rate due to the test setup. Tests that have the disturbance arrive just before the wireless update will not show the effect. Since there is no control over the timing of a disturbance relative to the update in an actual plant, the disturbance can be considered on the average to occur half way through the update time interval. Setpoint response tests will not show the effect of the wireless update delay in the initial response since the PID sees the change in setpoint immediately. The effect of wireless update delay on subsequent corrections depends upon the timing of the setpoint change and update similar to what was mentioned in terms of disturbance timing. Finally, the last source

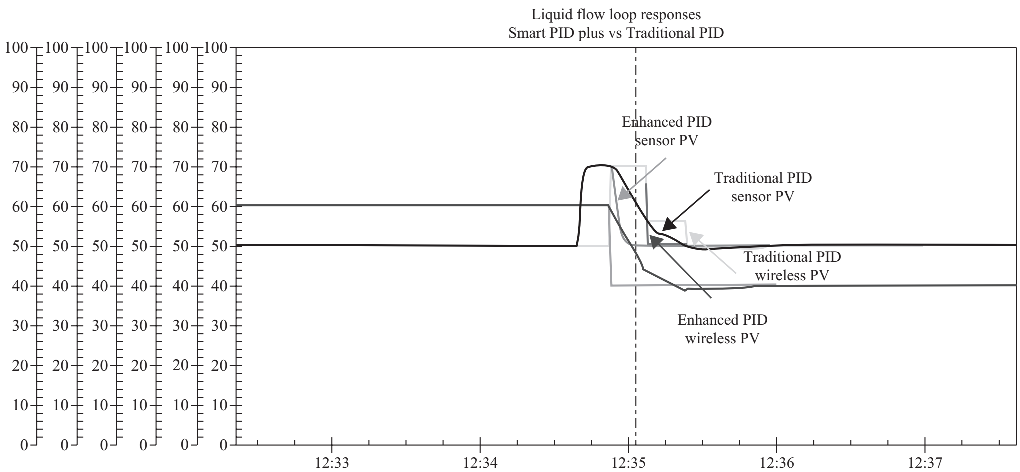


Figure 11.3b. Lower reagent flow loop load response of enhanced PID and traditional PID for wireless default update rate = 16 *seconds*.

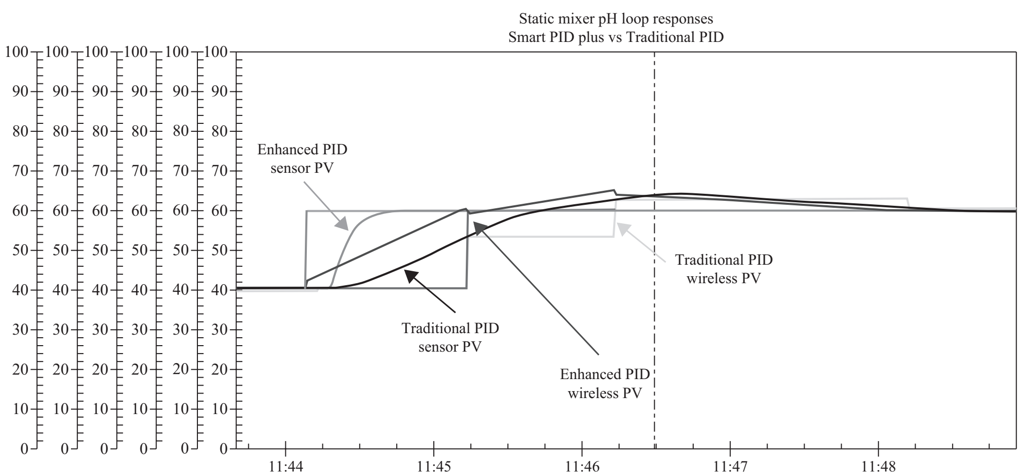


Figure 11.3c. Upper pH loop setpoint response of enhanced PID and traditional PID for wireless default update rate = 60 *seconds*.

of confusion is that the detrimental effect of a wireless update rate upon the integrated error depends upon the PID reset time as detailed by the equations in Chapter 6 on the effect of measurement dynamics. The effect of wireless update rate is negligible if the update rate is small compared to the reset time.

Figures 11.3a and 11.3c show the setpoint response for the lower flow and upper pH loop. The enhanced PID is able to make a single correction in its output that exactly matches the change needed to put the actual PV at the new setpoint after the process reaction delay and lag. Note that the wireless delay does not come into play if the correction is perfect. If

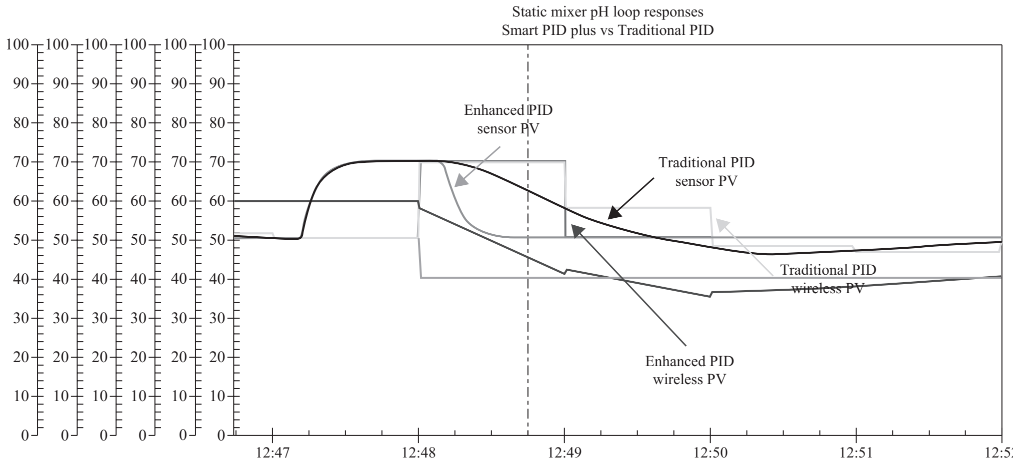


Figure 11.3d. Upper pH loop load response of enhanced PID and traditional PID for wireless default update rate = 60 *seconds*.

the PV did not reach setpoint or went past the sent point as a result of the single correction, the response to subsequent corrections would be delayed by the wireless delay. The dead time from the wireless update rate is seen in the trend of the wireless PV but not in the sensor PV. Unfortunately, the control room does not see the immediate response and thus operations may not fully appreciate the improvement in the setpoint response from the use of the enhanced PID.

Figures 11.3b and 11.3d show the load response for the lower flow and upper pH loop. The enhanced PID is able to make a single correction in its output that exactly matches the correction needed to return the actual PV to the existing setpoint after the wireless delay in recognizing the upset and the process reaction delay and lag. Here the wireless delay does affect the peak error and integrated error.

Figures 11.4a, b, c, d, e, f show for a *fast* lower loop the effect of increasing the lower PID Lambda (closed loop time constant) on the upper PID setpoint response with the external reset *off*. The rise time is quite fast but there is some overshoot in the upper loop. This could have been eliminated by the use of a setpoint lead-lag or 2 degrees of freedom structure as seen in the test results for Chapter 2. The response becomes noticeably oscillatory when the lower PID Lambda is increased to be one-fifth the upper PID Lambda.

Figures 11.5a, b, c, d, e, f show for a *fast* lower loop the effect of increasing the lower PID Lambda (closed loop time constant) on the upper PID setpoint response with the external reset *on*. The external reset feedback of the lower PID PV prevents the upper PID from trying to change the lower PID setpoint faster than the lower loop can respond. The response never becomes oscillatory. The effect of an increase in the lower loop PID is an increase in the upper PID rise time (time to reach setpoint). The increase is gradual at first but accelerates as the lower PID Lambda approaches the upper PID Lambda. For a lower PID Lambda that is four fifths of the upper PID Lambda, the rise time is double the rise time for a lower PID Lambda that is one twentieth of the upper PID Lambda.

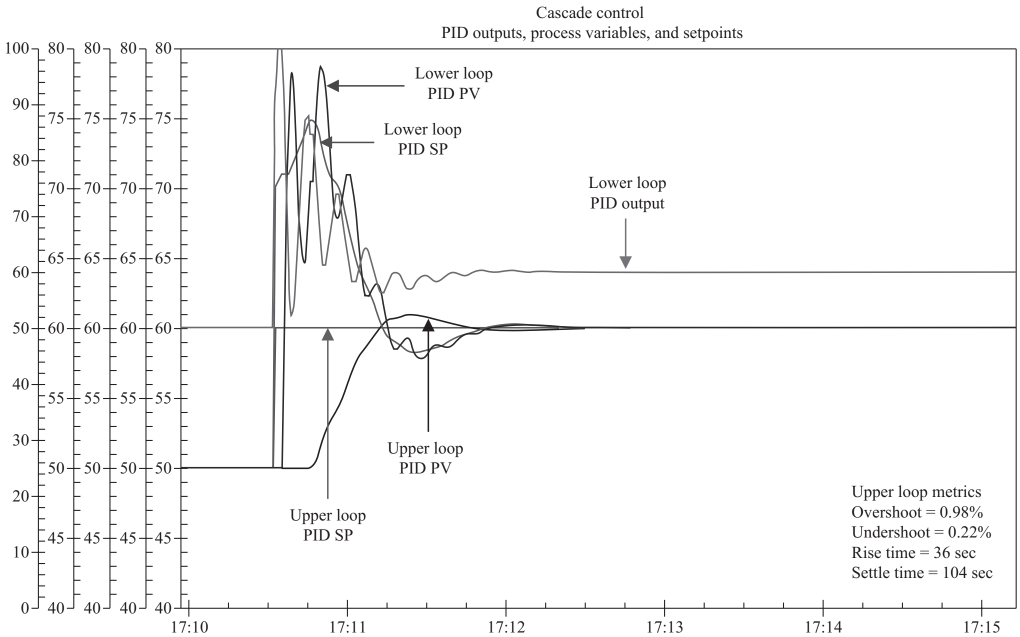


Figure 11.4a. Cascade control with fast lower loop (lower Lambda = 0.02 × upper time constant and external reset *off*).

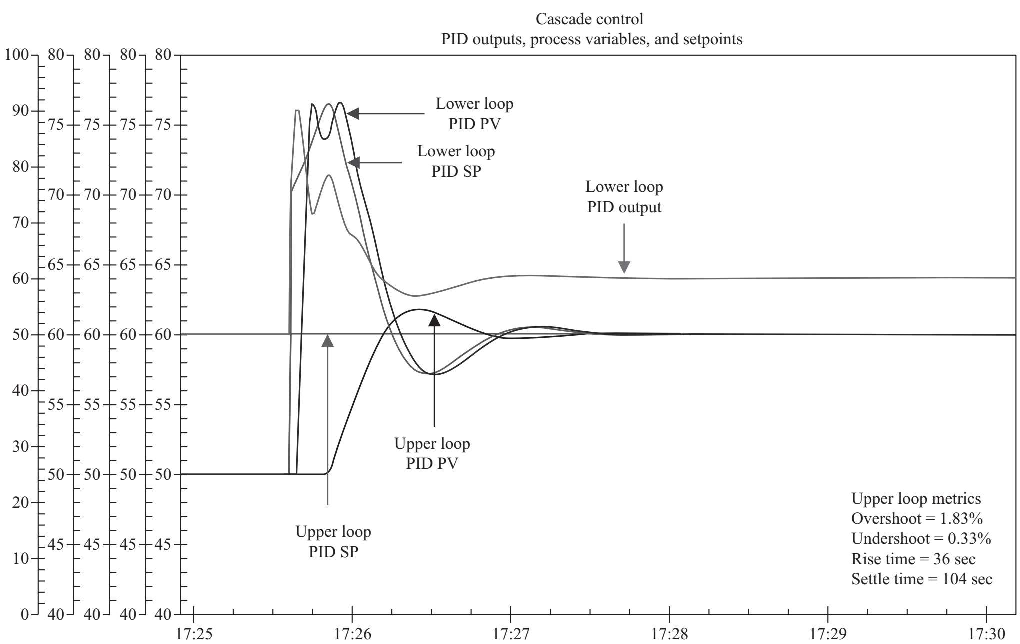


Figure 11.4b. Cascade control with fast lower loop (lower Lambda = 0.05 × upper time constant and external reset *off*).

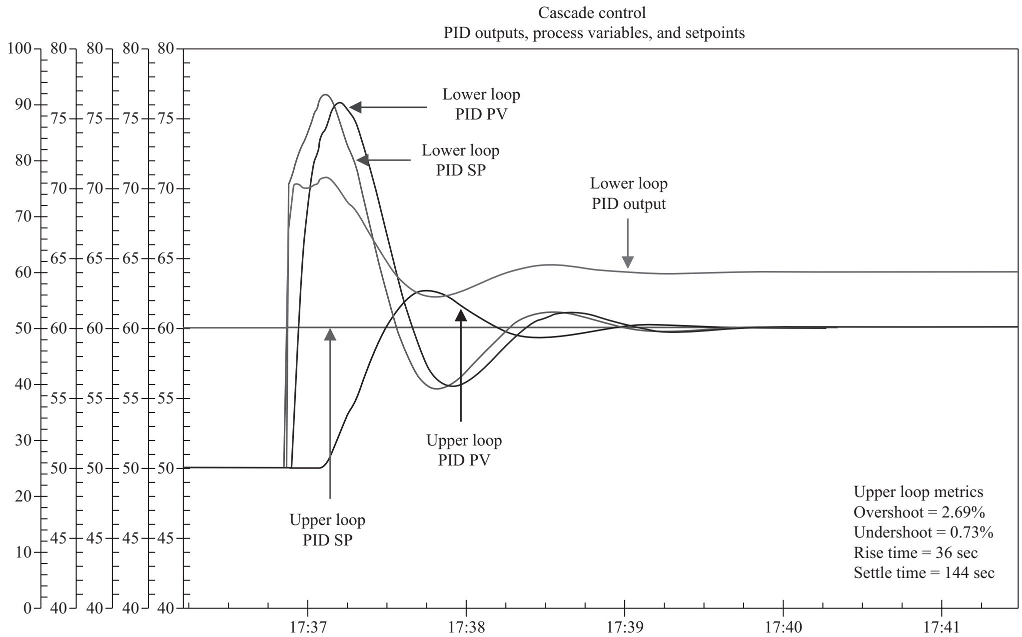


Figure 11.4c. Cascade control with fast lower loop (lower Lambda = 0.10 × upper time constant and external reset *off*).

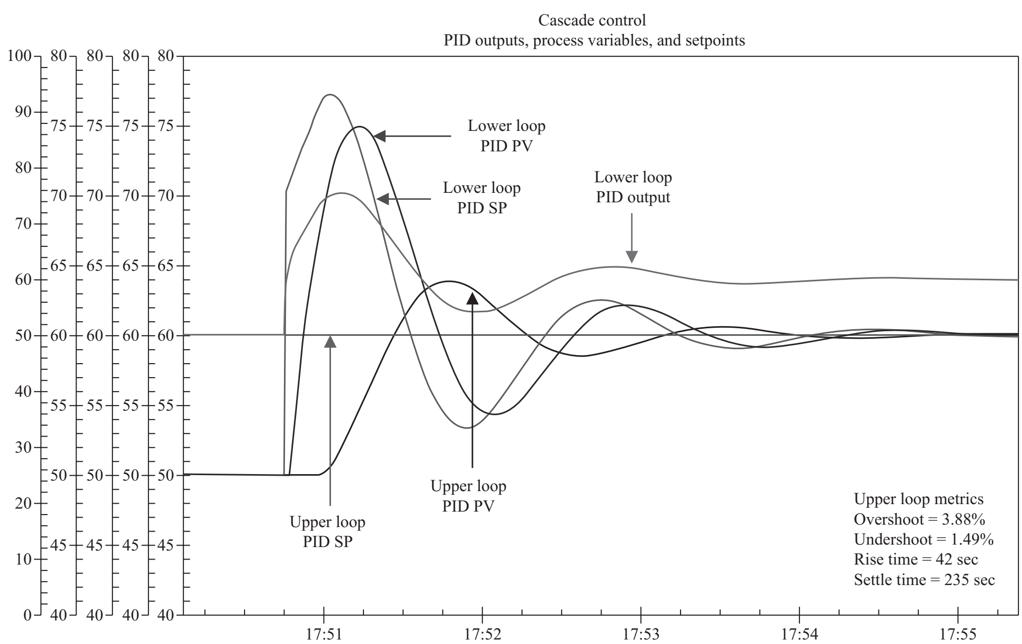


Figure 11.4d. Cascade control with fast lower loop (lower Lambda = 0.20 × upper time constant and external reset *off*).

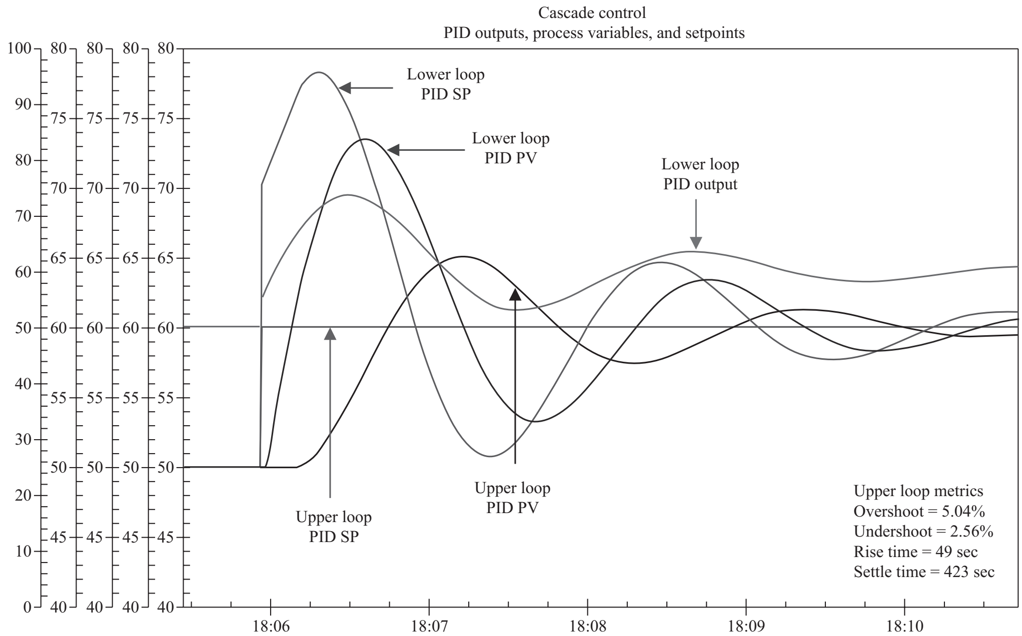


Figure 11.4e. Cascade control with fast lower loop (lower Lambda = 0.40 × upper time constant and external reset *off*).

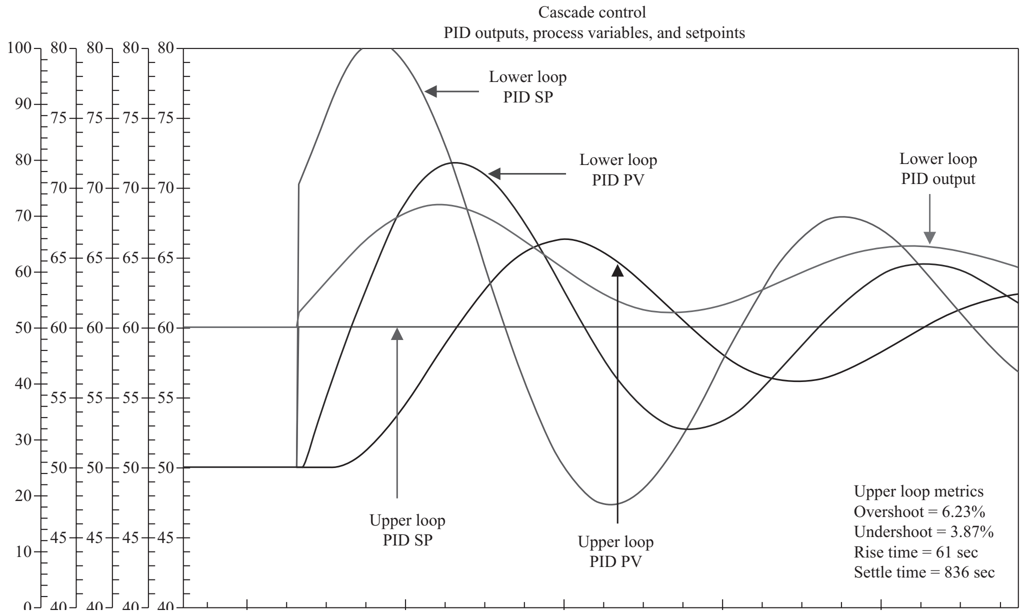


Figure 11.4f. Cascade control with fast lower loop (lower Lambda = 0.80 × upper time constant and external reset *off*).

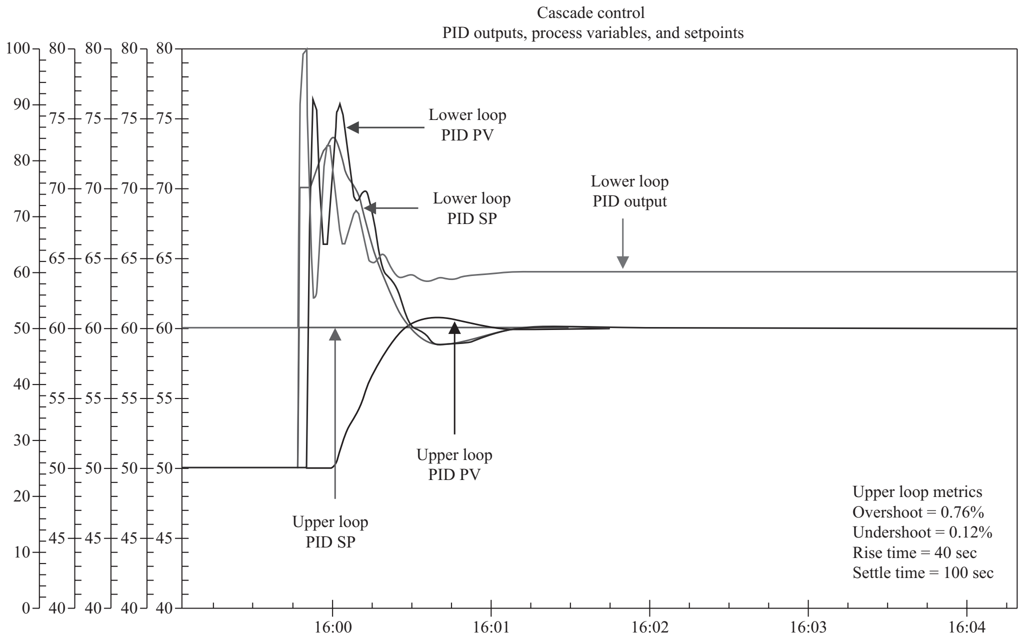


Figure 11.5a. Cascade control with fast lower loop (lower Lambda = 0.02 × upper time constant and external reset *on*).

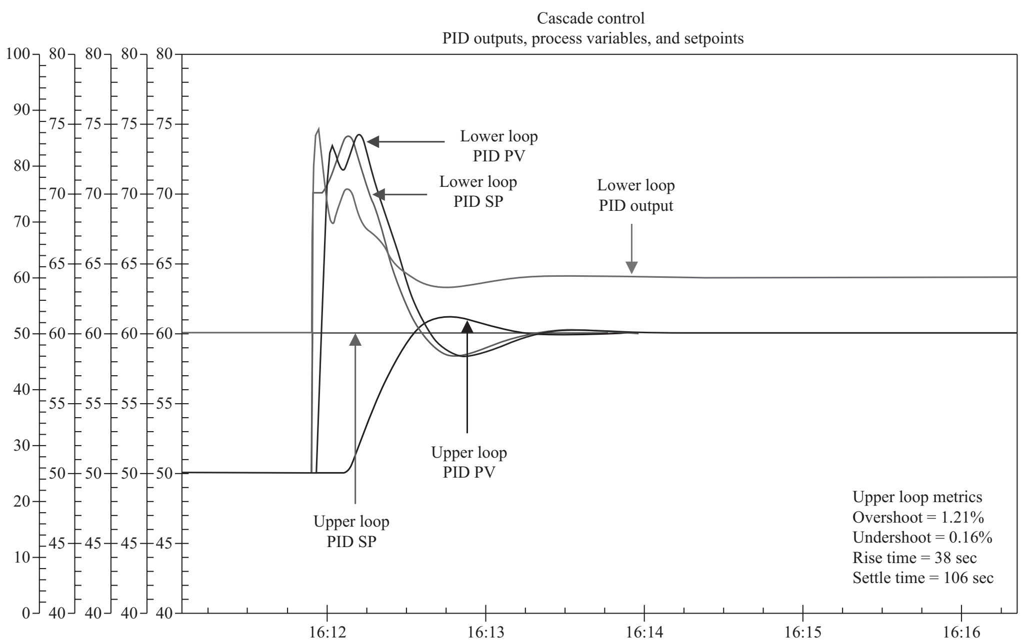


Figure 11.5b. Cascade control with fast lower loop (lower Lambda = 0.05 × upper time constant and external reset *on*).

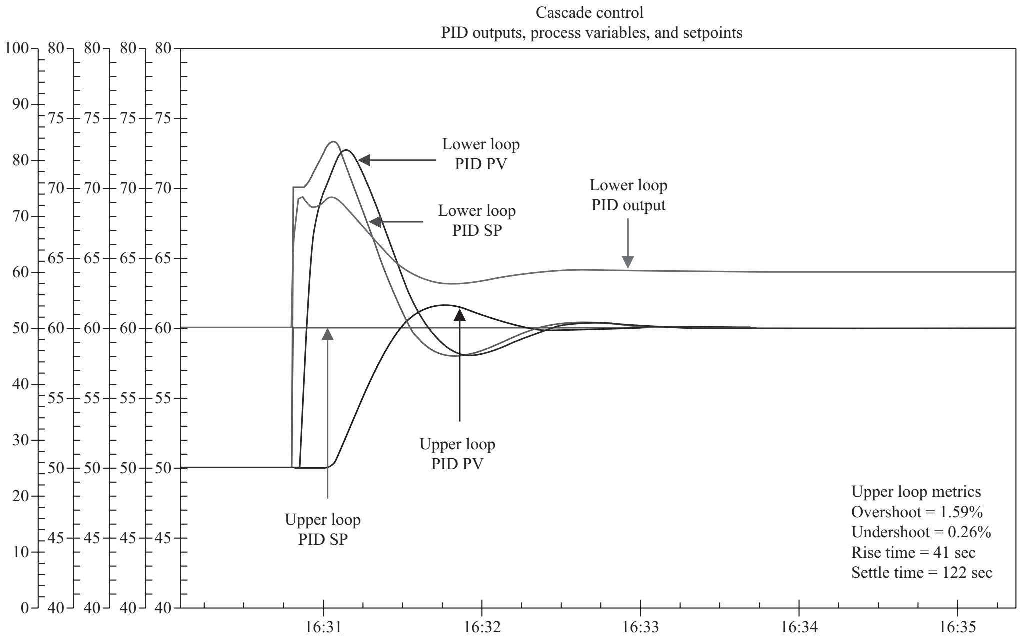


Figure 11.5c. Cascade control with fast lower loop (lower Lambda = 0.10 × upper time constant and external reset *on*).

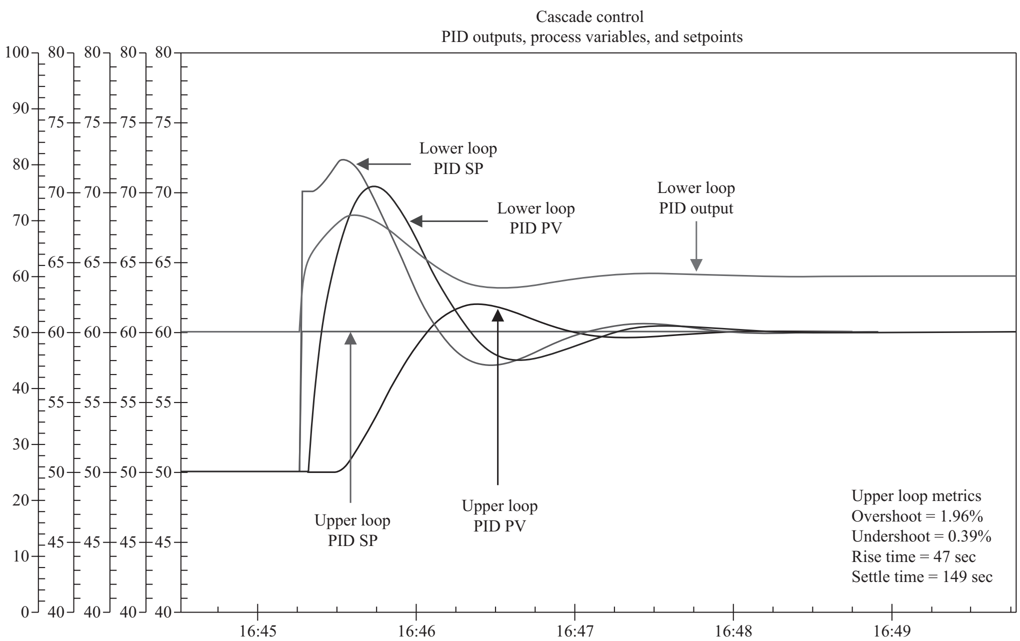


Figure 11.5d. Cascade control with fast lower loop (lower Lambda = 0.20 × upper time constant and external reset *on*).

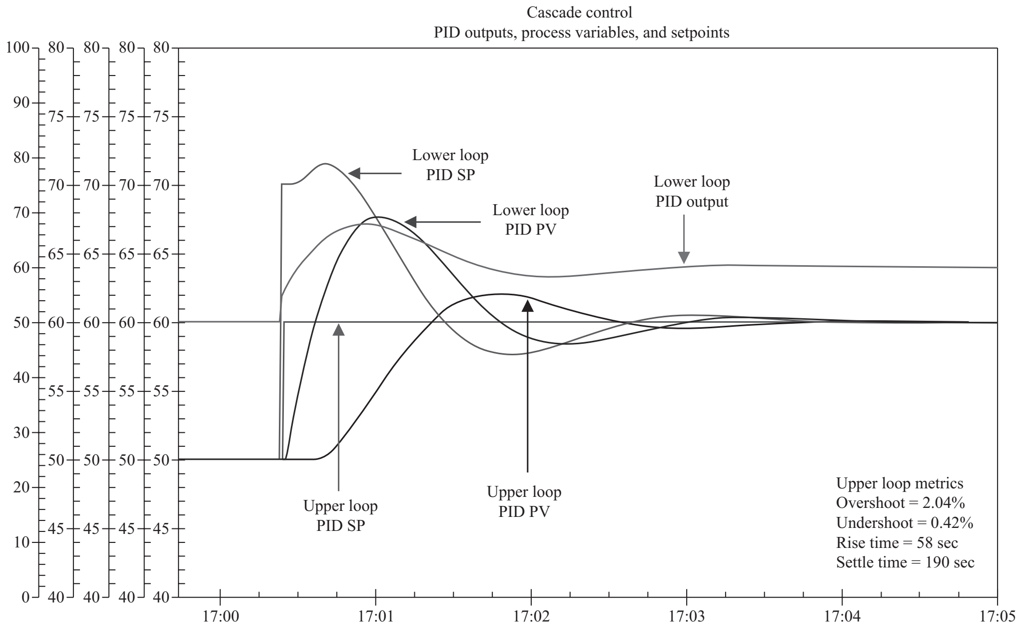


Figure 11.5e. Cascade control with fast lower loop (lower Lambda = 0.40 × upper time constant and external reset *on*).

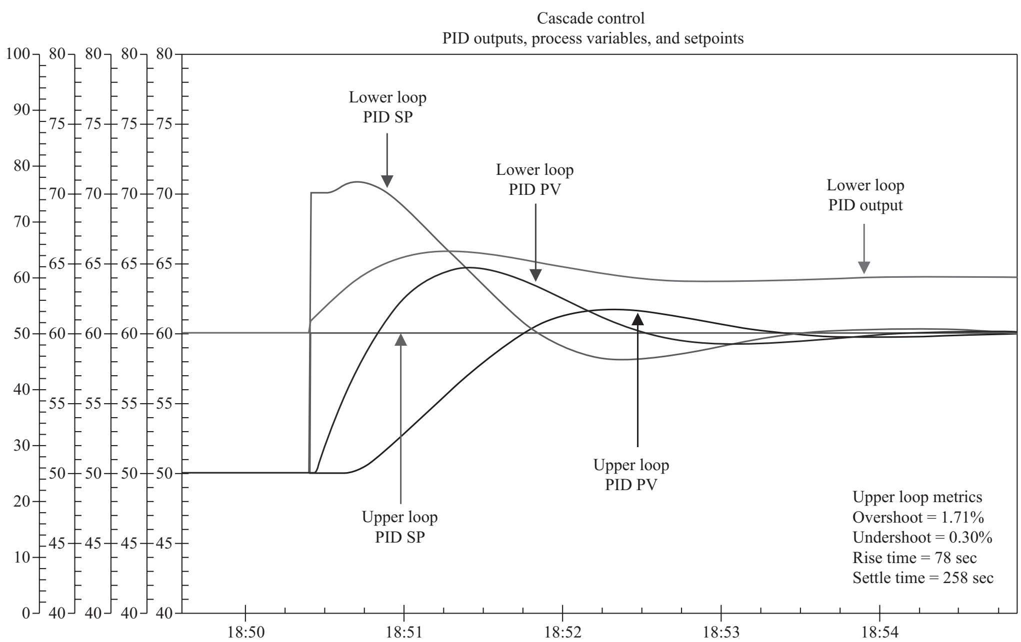


Figure 11.5f. Cascade control with fast lower loop (lower Lambda = 0.80 × upper time constant and external reset *on*).

Figures 11.6a, b, c, d show for a slow lower loop the effect of decreasing the upper PID Lambda (closed loop time constant) on the upper PID setpoint response with the external reset *off*. The rise time is incredibly slow but there is no overshoot in the upper loop for the first test where the upper PID Lambda is large enough to give a lower PID Lambda that is one tenth of the upper PID Lambda. The succeeding tests show the effect of decreasing the upper Lambda so that the lower PID Lambda is one fifth, two fifths, and four fifths the upper PID. Overshoot starts for a lower PID Lambda that is one fifth the upper PID Lambda but the rise time has decreased by over 60 percent compared to the first test. Oscillations start and get severe for a lower PID Lambda that is two fifths and four fifths of the upper PID Lambda, respectively. The results show that excessively detuning the upper PID is detrimental and the optimum ratio of lower to upper PID Lambda is one fifth. The tests do not show the adverse effect of upper PID detuning on the ability to reject disturbances that originate in the upper loop that adds weight to the guidance of avoiding as much as possible the detuning of upper PID. An increase in the upper PID lambda decreases the upper PID gain.

Figures 11.7a, b, c, d show for a slow lower loop the effect of decreasing the upper PID Lambda (closed loop time constant) on the upper PID setpoint response with the external reset *on*. The external reset feedback of lower PID PV prevents the start of a noticeably oscillatory response until the ratio of lower to upper PID Lambda increases to four fifths. For a ratio of two fifths, the setpoint response is comparable to that for a ratio of one fifth with external reset off. Also, the external reset feedback prevents severe oscillations even if the upper PID is only detuned to a ratio of four fifths, which is the last case. Thus, the use of external reset feedback enables the detuning of the upper Lambda to be considerably moderated.

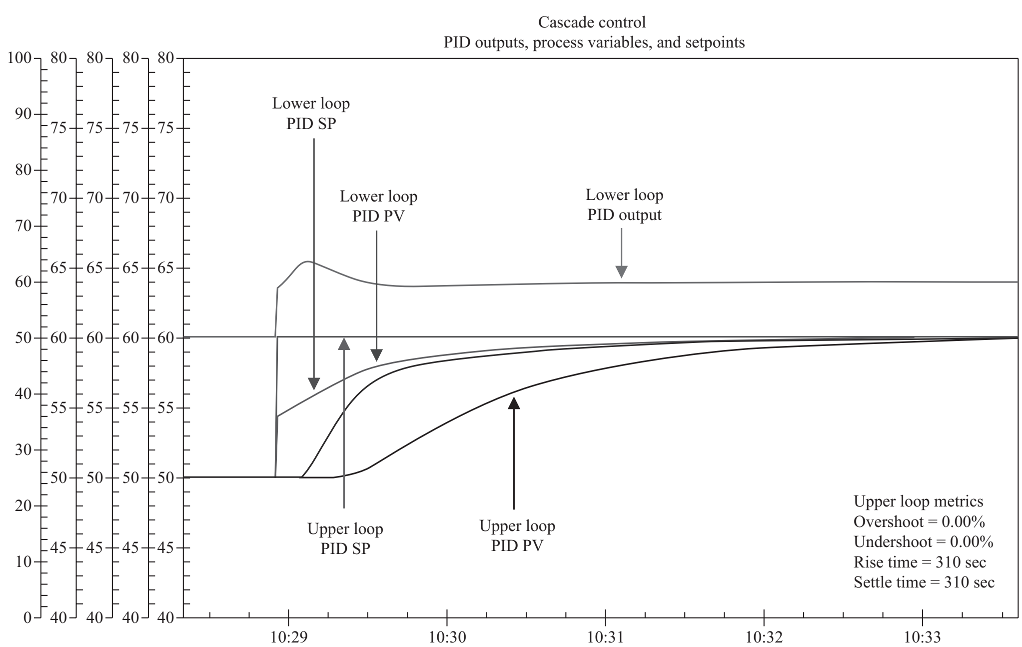


Figure 11.6a. Cascade control with slow lower loop (lower Lambda = 0.10 × upper Lambda and external reset *off*).

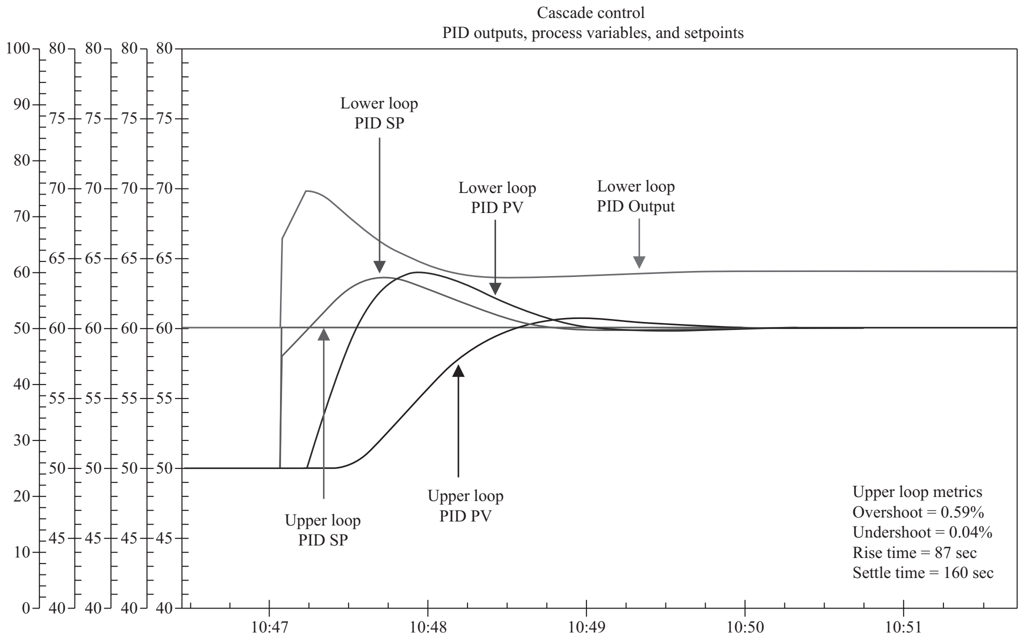


Figure 11.6b. Cascade control with slow lower loop (lower Lambda = 0.20 × upper Lambda and external reset *off*).

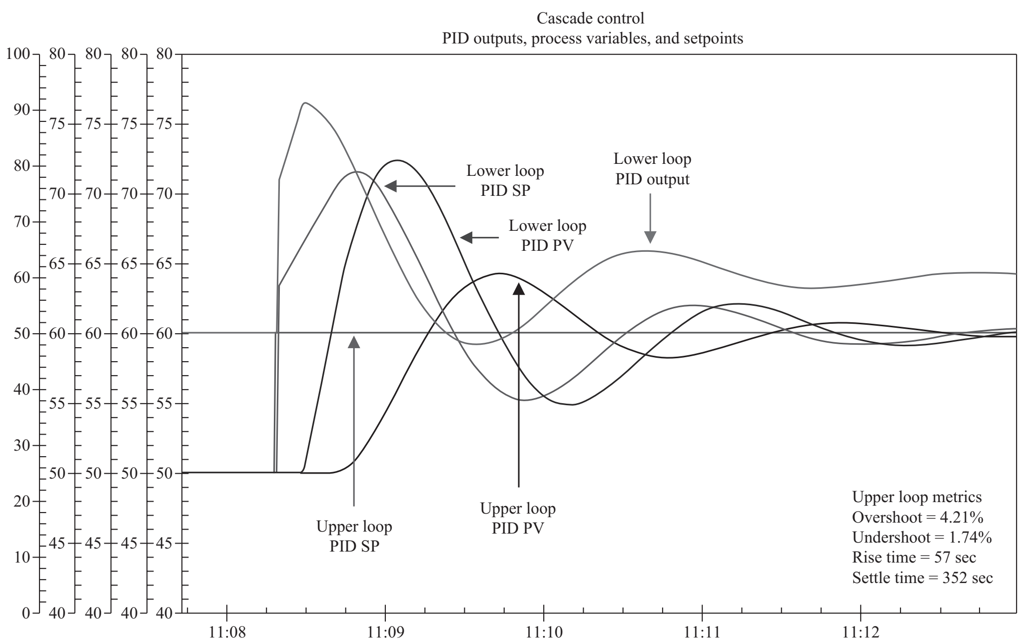


Figure 11.6c. Cascade control with slow lower loop (lower Lambda = 0.40 × upper Lambda and external reset *off*).

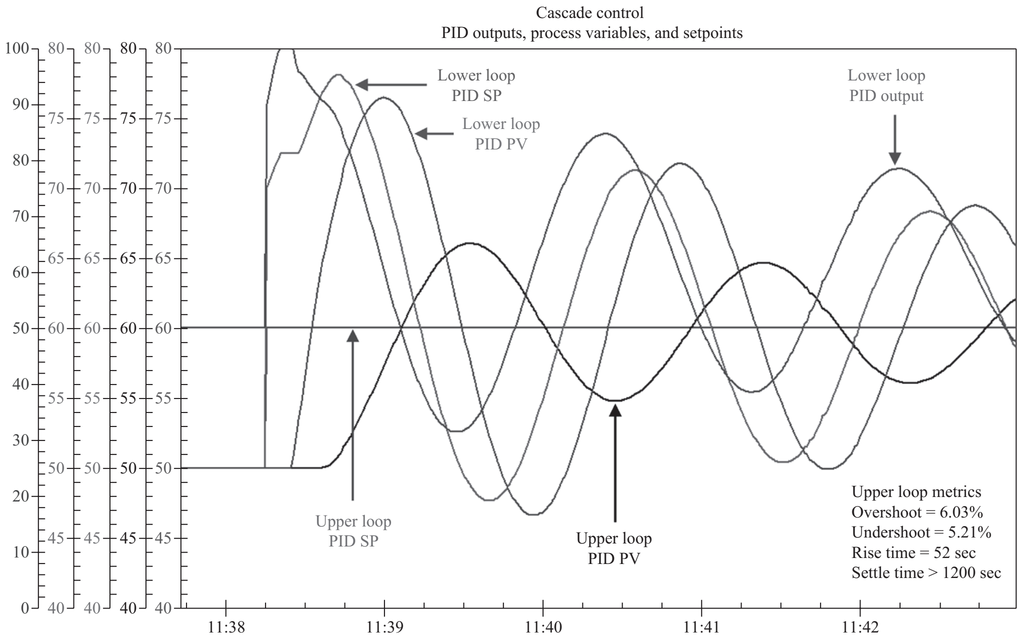


Figure 11.6d. Cascade control with slow lower loop (lower Lambda = 0.80 × upper Lambda and external reset *off*).

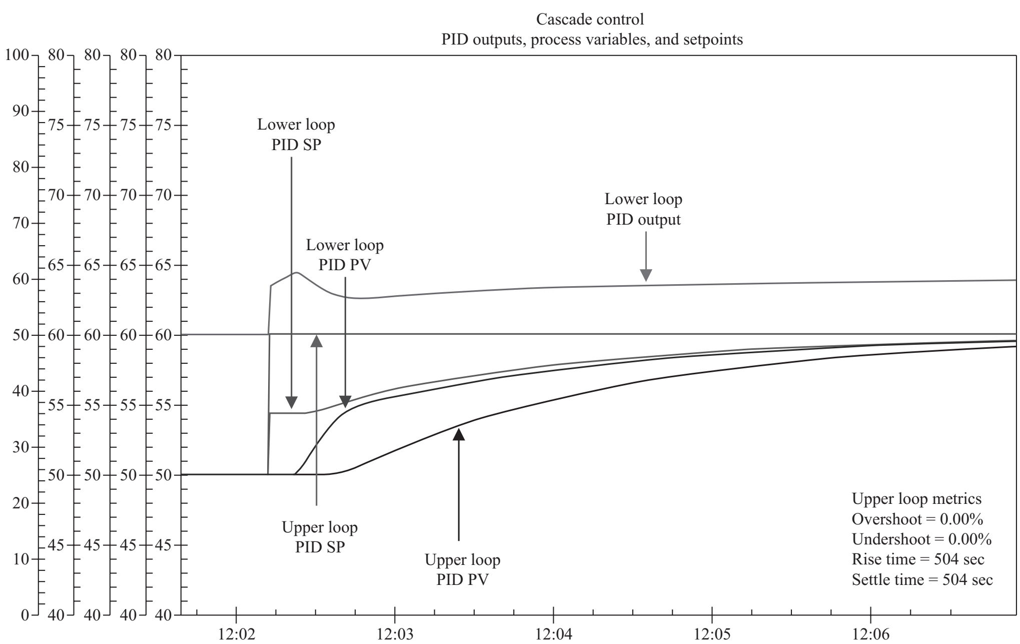


Figure 11.7a. Cascade control with slow lower loop (lower Lambda = 0.10 × upper Lambda and external reset *on*).

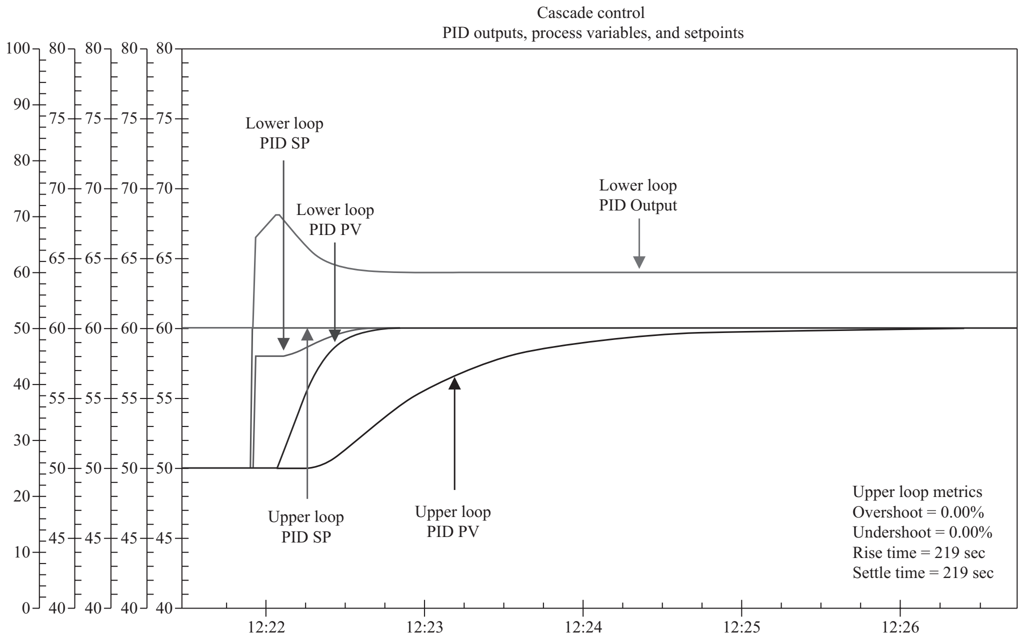


Figure 11.7b. Cascade control with slow lower loop (lower Lambda = 0.20 × upper Lambda and external reset *on*).

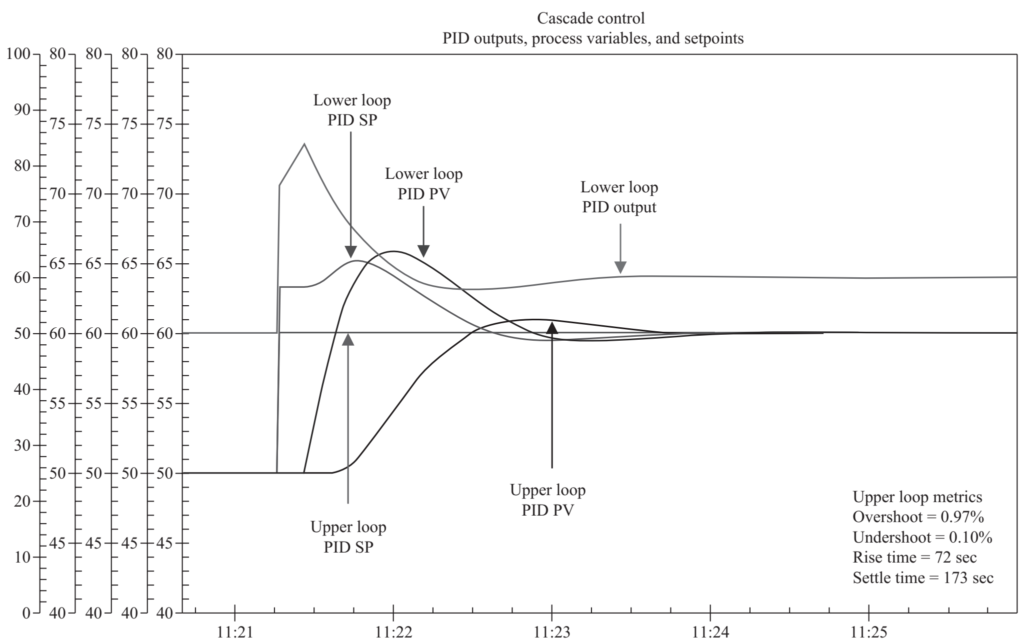


Figure 11.7c. Cascade control with slow lower loop (lower Lambda = 0.40 × upper Lambda and external reset *on*).

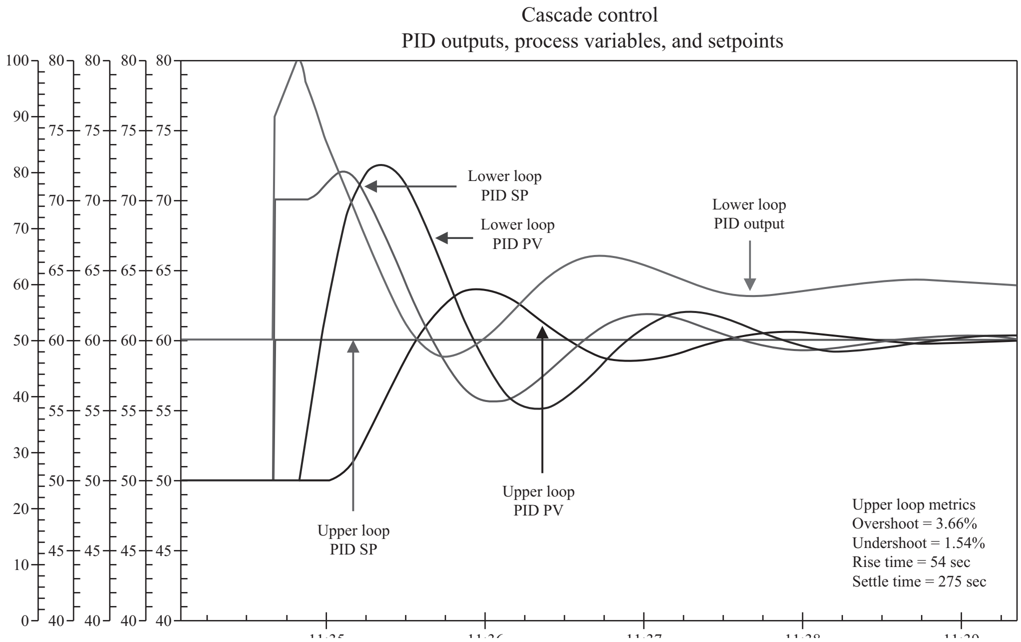


Figure 11.7d. Cascade control with slow lower loop (lower Lambda = 0.80 × upper Lambda and external reset *on*).

KEY POINTS

1. A fast automation system (PI on error structure and a fast PID execution, final control element, and measurement) is particularly important for the lower loop.
2. The cascade rule of the lower loop being five times faster than the upper loop pertains to the closed loop response and not the open loop response. The greatest improvement in going from single loop control to cascade control occurs when the lower loop process time constant (e.g., secondary time constant) approaches the upper loop process time constant (e.g., primary time constant).
3. If an upper PID must be detuned, the objective is a ratio of lower to upper PID constant that is two fifths and one fifth if the PID external reset is off and on, respectively.
4. The use of external reset feedback offers considerable advantage in performance and forgiveness for tuning that violates the cascade rule.
5. Cascade control offers considerable process control and process analysis benefits.
6. The principle problems with cascade control originate from a lower loop measurement with poor rangeability or signal to noise ratio (e.g., orifice meter without sufficient straight pipe upstream), an upper loop measurement with poor reliability (e.g., at-line analyzer with sample system problems), incorrect configuration of limits and scales, absence or incorrect configuration of external reset feedback, and a PID that is not tuned properly (e.g., slow lower PID tuning or fast upper PID tuning).

CHAPTER 12

ADVANCED REGULATORY CONTROL

12.1 INTRODUCTION

The power of the proportional-integral-derivative (PID) largely remains underutilized. Most of the options and parameters other than scale ranges and tuning settings remain at their defaults. Here we look how to tap into the incredible capability of the PID to minimize batch cycle time and startup time and to maximize production capacity, flexibility, and efficiency while reducing disruption to utility, raw material, and reagent systems. Also detailed is a simple method to compensate for dead time and to eliminate limit cycles.

12.1.1 PERSPECTIVE

Advanced regulatory control seeks to incorporate knowledge of process dynamics, disturbances, constraints, and objectives to increase process efficiency and capacity. The PID power and flexibility enables an incredible spectrum of creative opportunities to achieve these goals.

The most common advanced control technique is feedforward control that is designed to preemptively correct for a disturbance. In industrial applications, feedforward control is far from perfect so feedback control is essential. Most of the problems with feedforward control stem from incorrect and missing dynamic compensation or poor measurement and valve resolution or rangeability. Separate open loop tests should be conducted where the disturbance variable and manipulated variables are stepped. If a time constant plus dead time model of both responses is adequate, the dynamic compensation simplifies to a feedforward dead time and lead-lag block.

The feedforward dead time is the disturbance variable dead time minus the manipulated variable dead time. If this value is negative, dynamic compensation of dead time is not used and the feedforward gain needs to be reduced because the feedforward correction will be late and some feedback action will already have occurred.

The feedforward lag time is set equal to the disturbance variable time constant. The lead time is then set equal to the manipulated variable time constant. The ratio of manipulated variable to disturbance variable time constant is the lead-lag factor. If this factor is greater than one, there is a kick in the feedforward to compensate for the excessive slowness in the manipulated variable. If open loop tests cannot be done or higher order dynamics (e.g., secondary lag) alter the initial response or an integrating or runaway response occur, then an addition of lead or lag time to make the feedforward response faster or slower may be the simplest solution. Test results will show how the closed loop response to a disturbance can be used to decide what term to adjust and in what direction.

If the feedforward arrives too soon, an inverse response is created that is particularly disruptive. In this case, the feedforward can do more harm than good. Since timing cannot be perfect, a feedforward signal that is a little late is better than one that is early especially considering PID controllers are typically tuned with a gain that is far below the maximum. Thus, the better solution is to make sure the feedforward signal can handle major disturbances with the largest dead time and slowest time constant by being liberal with the use of dead time and lag time in the manipulated variable path. This is not to say there is no problem with a feedforward dead time or lag that is excessive. If the feedforward correction arrives late by more than half the ultimate period, the feedforward correction can get in phase with the feedback correction, creating undershoot. If the feedforward correction is more than twice the ultimate period, a second upset in the opposite direction of the original upset is created. The amplitude of undershoot or a secondary upset can approach the amplitude of the original disturbance depending upon the feedforward gain. Knowledge of how time constants and dead times change with production rate should be used to schedule the settings in the feedforward dynamic compensation to avoid the feedforward signal arriving early or too late.

Setpoint feedforward does not require dynamic compensation. Setpoint feedforward only offers a significant advantage if the step change in PID output from the PID modes for a setpoint change is less than 50 percent of the final change needed as a step change in PID output. This scenario can be caused by a structure of proportional on process variable (PV) (I on error and PD on PV), the product of the beta setpoint weight factor and PID gain less than 0.5 for a two degrees of freedom (2DOF) structure, or the product of the setpoint lead to lag ratio and PID gain less than 0.5 for a structure of proportional on error (e.g., PI on error and D on PV). These values are offered as a guideline only. Specific application conditions need to be considered.

The other major considerations to be addressed are disturbance measurement noise, error, and rangeability. Noise can be amplified by a feedforward lead time greater than the lag time. If the rangeability of a flow measurement is insufficient, a flow calculated from valve position should be used for low flows. A flow setpoint instead of a flow measurement for the flow feedforward will eliminate noise and provide coordination of flow changes for a change in production rate.

Intelligent output action via output tracking and the remote output mode enable the PID output to be scheduled to deal with abnormal situations. This function is called an open loop backup because the feedback control is temporarily suspended to guarantee a fast correction. Control is returned to the PID feedback control when a protective state is assured. An open loop back up has been extensively used to prevent surge and environmental violations.

The PID output can be set at a limit to achieve the fastest possible setpoint response without upsetting other loops by setpoint rate limiting of the manipulated flow. The PID output is then set to a final resting value (captured from last batch or startup) and control is returned to the PID when a PV value one dead time or more into the future is projected to reach the setpoint. This full throttle response can reduce batch cycle and startup times.

Intelligent integral action can stop limit cycles (equal amplitude oscillations) from deadband and resolution or threshold sensitivity limits, and stop variable amplitude oscillations from wireless devices, at-line and off-line analyzers, slow secondary loops, or slow final control elements, and enable directional move suppression for optimization. Move suppression is set with analog output (AO) block or lower loop setpoint up and down rate limits to provide a gradual and less disruptive approach to an optimum and a fast getaway for abnormal conditions. Move suppression is also useful to prevent unnecessary crossing of the split range point that is the location of severe process and valve discontinuities in the transition from one manipulated variable to another. The elimination of unnecessary crossings not only reduces process variability but also improves process efficiency from eliminating cross neutralization of reagents or repetitive heating and cooling.

The principal methods of optimization in advance regulatory control involve the use of valve position control (VPC) and override control. Often these work together to providing a sequential optimization addressing the most pressing constraint.

In VPC, process control loop valve(s) are pushed to a maximum or minimum throttle position to reduce energy use by lowering a compressor discharge pressure or variable pump speed, reduce raw material costs by reducing expensive reagent flow, or maximize production rate by increasing feed flow. The main problems with VPC is the slow response from integral only control and the lack of tuning rules to prevent interaction and the response to limit cycles in valve position. The limit cycle period is large and only apparent when there are no disturbances and a long trend chart time frame is used. Intelligent integral action, directional move suppression, and feedforward control can make VPC more effective and easier to implement.

In override control, the lowest or highest output of VPC representing valve constraints or multiple PID controllers representing PV constraints and targets, such as high level or high pressure or a desired flow, is selected either as a setpoint for a PID feed or to directly manipulate a final control element (e.g., control valve or variable speed drive [VSD]). The principle difficulty with override control is the proper setup of external reset feedback to prevent an unselected controller from winding up or walking off, tuning to deal with constraint dynamics, and operator understanding of the order of constraints in the sequential optimization.

VPC is also used to simultaneously stroke a small control valve in parallel with a large control valve. The small valve provides better resolution and threshold sensitivity and the large valve provides capacity. The VPC problem of being too slow is addressed by the use of a compound feedforward signal. The first part of the feedforward signal is typically a production flow feedforward. The second part is an innovative scheduling of a feedforward to make a fast correction for the approach of the small valve to an output limit. The implementation uses an "If... Then" statement in the expression of a Calculation (CALC) block. The VPC feedforward is incremented similar to an open loop backup for environmental protection except the VPC stays in Auto. If the small valve position (VPC PV) is predicted one dead time into the future to be at an output limit, the VPC feedforward is incremented or decremented in the direction to bring the small valve off the limit. This feedforward action stops immediately when the future small valve position is no longer predicted to be at a limit.

Model predictive control (MPC) of continuous processes offers an inherent dynamic compensation of feedforward signals and constraints, a simultaneous optimization of constraints, incorporation of economics (e.g., feed and energy costs), and a potential analysis of the contribution of each constraint to the MPC output. Advanced regulatory control is used for batch operations and as a low cost and quick solution for simple optimization problems since only a configuration change is needed.

12.1.2 OVERVIEW

Feedforward control can be used to improve regulatory control and optimization by preemptive correction or coordination of process control and VPC. The feedforward signal gain and dynamic compensation should be designed to provide a correction in the process at the same point and same time as the disturbance. A dead time block and lead-lag block is generally all that is needed for this dynamic compensation of feedforward signals. Given an error in timing is inevitable, preferably the feedforward signal arrives late rather than early in order to prevent inverse response.

The most common feedforward signal is flow because of production rate changes and because nearly all fast and large disturbances originate as flow rather than composition or temperature disturbances. To attain the process flows on the process flow diagram (PFD), the manipulated flow is ratioed to a wild or leader flow. In most cases, the feedforward is the primary feed flow. A ratio station is used to provide the operator with the ability to set the desired ratio and see the actual ratio. The process PID provides a feedback bias correction to the manipulated flow. The actual ratio displayed is the corrected manipulated flow divided by the feedforward flow.

Intelligent integral action is the key to preventing oscillations. The suspension of integral action when final control elements are not moving (e.g., deadband and resolution limits) and measurements are not updating (e.g., communication failure, wireless default update rate, and analyzer cycle time), will stop oscillations and reset windup. The slowing down of integral action when secondary loops or final control elements are slow to respond will prevent the burst of oscillations from large and fast disturbances and setpoint changes.

A signal selection block chooses the lowest or highest of outputs from override PID controllers to honor constraints. Override process controllers prevent violation of PV constraints. Override VPCs prevent violation of control valve position constraints by keeping the control valve position at the most desirable throttle range. Output limits serve as a backup. Intelligent integral action and feedforward action are keys to a fast smooth override response for process protection and optimization.

Dead time compensation can be simply implemented by the addition of a dead time block in the external reset feedback signal (PID + θ_o). This PID achieves about the same performance as a Smith Predictor without having to set a predictor model gain and time constant. The dead time setting must always be within +5 and −20 percent of the actual dead time so that the PID reset time can be significantly decreased toward the limit of the dead time error or module execution time to see an improvement in control. If the reset time is not decreased, the performance is worse than if there was no dead time compensation. Overestimates of the dead time will lead to fast oscillations whereas low estimates just cause a loss of performance advantage. This is the opposite effect for the tuning of the PID where an overestimate of dead time causes sluggish response and an underestimate mostly causes an oscillatory response. Contrary to popular opinion the improvement by dead time compensation is greatest for lag dominant rather than dead time dominant loops. Ironically, lag dominant loops (low dead time to time constant ratio) can have a high controller gain and have less of a need for dead time compensation.

The enhanced PID developed for wireless that suspends integral action when there are no measurement updates can simplify the tuning for at-line and off-line analyzers. If the cycle time is much larger than 63 percent process response time, the reset time can be set as small as the process dead time rather than the open loop dead time and the PID gain can be set as large as the inverse of the open loop steady state gain for self-regulating processes. The tuning is not as simple for integrating processes and the improvement is not as dramatic but the enhanced PID adds a significant degree of robustness. The advantages and rules also pertain to large wireless update times on fast processes. While composition control is more robust and setpoint response much faster by the enhanced PID, the peak error and integrated error for unmeasured load disturbances is increased due to the increase in total loop dead time by amount that is 1½ times the analyzer cycle time and half of the wireless update time.

12.1.3 RECOMMENDATIONS

1. Use feedforward control to give preemptive action for large and fast disturbances.
2. Use ratio and bias and gain stations to provide flow ratio control with feedback bias correction and the actual flow ratio viewable and desired flow ratio accessible by operations. Enable startup on ratio control without feedback correction until the process (e.g., distillation column) reaches operating condition.
3. Provide dynamic compensation of feedforward signals by dead time and lead-lag blocks to give the right feedforward timing for disturbances.
4. If open loop tests cannot be done or the process response is complex (e.g., high order or recycle effects), integrating, or runaway, the dead time and lead-lag times can be adjusted based on oscillation patterns observed.
5. Ensure the feedforward correction does not arrive too soon causing an inverse response that can result in the feedforward doing more harm than good.
6. If the feedback correction will unavoidably arrive late, decrease the feedforward gain to prevent a secondary oscillation and undershoot that result from feedback control having already done some of the correction.
7. Use intelligent output action via output tracking or remote output to schedule output changes for an open loop backup to protect equipment or full throttle setpoint response (bang-bang control) to minimize batch cycle times.
8. Use a rate limit on valve, VSD, or flow loop setpoints (move suppression) with external reset feedback to the primary PID (e.g., composition, pH, or temperature) manipulating these setpoints to prevent disruption to utility, raw material, and reagent systems from high PID gain or rate time settings.
9. Use intelligent integral action via integral deadband, external reset feedback, and/or an enhanced PID developed for wireless to prevent oscillations from split range discontinuities, deadband, resolution or threshold sensitivity limits, coated electrodes, slow secondary loops or final control elements, communication failures, and large wireless update times, and analyzer cycle times.
10. If the wireless update time or analyzer cycle time is much greater than the 63 percent process response time, use an enhanced PID to suspend integral action between measurement updates and increase the PID gain toward a maximum gain that is the inverse of the open loop steady state gain for self-regulating processes.
11. Use innovative feedforward control to help a VPC deal with large and fast disturbances with parallel large and small valves to achieve best resolution and threshold sensitivity and rangeability.
12. Use override control to honor PV and valve position constraints.

13. Use up and down setpoint rate limits with intelligent integral action to provide a directional move suppression to enable a slow approach to an optimum or a split range point and a fast getaway for abnormal operation and to prevent safety instrumentation system (SIS) activation.
14. If the dead time can always be accurately set within +5 and −20 percent of actual dead time, add a dead time block in the external reset feedback signal path and decrease the reset time by a factor of four or more to improve performance.
15. Use setpoint feedforward to reduce the time to reach setpoint (rise time) when the initial step change in PID output due to PID tuning and structure for a setpoint change is less than 50 percent of the final PID output change needed.
16. Move up to MPC when the feedforward and feedback dynamics are complex and multivariable, the optimization involves more than one objective, or simultaneous constraint control is needed.

12.2 FEEDFORWARD CONTROL

Feedforward control is the most frequently used advanced regulatory control technique. The most common feedforward signal is a feed flow. The result is flow ratio control with feedback correction by means of a feedforward summer. For better operator involvement, the operator must be able to change the desired ratio and see the actual ratio. This operator interface is particularly important for running on just ratio control (no feedback correction) for startup, abnormal operation, and analyzer problems. While the literature leads one to believe a feedforward multiplier is better for changes in ratio, feedforward correction multiplier signal scaling is problematic and the gain nonlinearity introduced into liquid systems with back mixing is undesirable. In well-mixed volumes, effect on PID tuning of the change in process gain with production rate is compensated by the change in the primary time constant (residence time) with production rate. Finally, most of the error in measurements is an offset error that can be corrected by a bias. The fact that MPC and neural networks use a bias correction is a clue to fundamental advantages of a feedforward summer over a feedforward multiplier.

12.2.1 OPPORTUNITIES

The opportunities all stem from a preemptive correction of a disturbance. The feedforward correction needs to arrive at the same time at the same point in the process as the disturbance with an equal magnitude and opposite sign.

1. A feedforward timing and signal accuracy of 10 percent offers a 10:1 improvement in the peak and integrated error for a load disturbance (e.g., feed disturbance).
2. Plantwide feedforward is possible where the plant moves in concert to production rate changes to minimize inventories.
3. PID tuning can be less aggressive and consequently less oscillatory and more robust since most of the work is done by feedforward.
4. Interaction between production and quality control loops can be decoupled, which is particularly important for VPC.

12.2.2 WATCH-OUTS

Most of the watch outs are associated with feedforward measurement or valve problems, feedforward timing mismatch, or unmeasured disturbances.

1. Limit cycles in the feedforward signal from valve stick-slip particularly near the closed position which leads to poor rangeability.
2. Low signal to noise ratio in the feedforward signal at low velocities which leads to poor rangeability.
3. Manipulated variable dead time that is more than twice the disturbance dead time leading to severe undershoot or a secondary disturbance from feedforward correction arriving late by more than half the ultimate period.
4. An unmeasured disturbance (e.g., change in feed composition) coincident with a change in feed flow that causes an upset that is larger in the opposite direction. Here an intelligent feedforward can recognize that a feedforward correction based on the flow change would cause a PV change in the same direction of the current excursion from a different disturbance and should be moderated.

12.3 INTELLIGENT OUTPUT ACTION

The best of operational and process knowledge can be implemented as intelligent output actions to increase the capability of the PID to handle abnormal situations, interactions, and setpoint changes. Anything that an operator thinks he must do by placing the PID in manual can be done automatically providing much greater repeatability and the opportunity for analysis and continual improvement. The PID output tracking and remote output modes are used to schedule the intelligent output actions. The PID is returned to the automatic or higher level mode as soon as the full effect of the intelligent correction is realized in the process.

12.3.1 OPPORTUNITIES

Many operator interventions are due to poor tuning or training. However there are cases where feedback action is too late or too oscillatory due to special circumstances. Here is a list of the biggest opportunities. All of them potentially take advantage of a simple calculation of a future PV value. The calculation is simple. A new PV value is passed through a dead time block to create an old PV value that is subtracted from the new PV value to create a delta PV which when added to the new PV is the future PV. The dead time block parameter is a factor of the total loop dead time. The minimum factor is one so the prediction is at least one dead time into the future. The factor is increased to improve the signal to noise ratio.

1. An open loop backup can prevent a compressor from going into surge by detecting crossing of the surge curve. A large precipitous change in flow indicating the start of surge or preemptively a future PV value of suction flow predicted to cross the anti-surge controller setpoint causes the PID to go to output tracking or remote output mode. The PID output is then conservatively set to a known value that will always prevent surge. The PID output

will be held at this value for at least 10 dead times or 4 seconds, whichever is greater, to ensure any surge oscillation has stopped and recovery is complete. The PID output is then returned to feedback control via the automatic or higher mode (e.g., cascade).

2. An open loop back up can prevent a Resource Conservation and Recovery Act (RCRA) pH violation by keeping the effluent pH between pH 1 and 12. A future PV value predicted to get close to the pH 2 to 12 boundaries triggers an incremental change in the PID output via the output tracking or remote output mode. The incremental change is fast enough to ensure a prevention of an environmental violation in one loop dead time. For inline pH systems (e.g., static mixer pH control systems), the increment may need to be as fast as 2 percent per second. The limit to the total change in PID output is selected to prevent a violation for the worst case scenario. The PID is returned to feedback control when the future PV indicates sufficient recovery for at least 10 dead times. The strategy can be used to prevent other environmental violations.
3. A PID can reach a new setpoint as fast as possible with little to no overshoot by a *full throttle setpoint response* also known as *bang-bang* control as described in Control article "Full Throttle Batch and Setpoint Response" (McMillan 2006). The faster setpoint response can reduce startup, transition, and batch times. The PID is set and held at an output limit until the future PV is predicted to reach setpoint. At this time the PID output is set and held at the final resting value for at least one dead time. Test results show the strategy is most effective for near-integrating, true integrating processes with a dead time factor slightly greater than one. While the strategy can be effective for exothermic reactions, taking a PID even intelligently out of automatic or higher mode that is responsible for preventing a runaway is unsafe.
4. PID outputs can be scheduled to prevent interacting loops from upsetting each other due to a setpoint change. The loop with the setpoint change uses the full throttle logic just described. Affected PID loops on self-regulating processes are held at their final resting values. Affected loops on integrating processes are initially put at an output that compensates for the effect of change in setpoint from interaction before going to the final resting value. The initial output provides an unbalance that drives the loop affected by the interaction back towards the setpoint (the only way to get an integrating process to move is to create a temporary unbalance). The PIDs return to feedback control when all responses of affected loops are complete. The intelligent scheduling can eliminate interactions.

12.3.2 WATCH-OUTS

As with any industrial application there are practical considerations that can make or break the solution. Most of these have to do with noise and changes in dynamics which are the major issues for tuning.

1. Noise can trigger the activation of an open loop backup. Judicious rate limiting of changes in the PV to practical values can help reduce noise without adding a lag. The dead time block parameter can also be increased to reduce noise.
2. The return to feedback control is too soon because of an increase in dead time (e.g., fouling) or settling time (e.g., interaction) or too late making the PID unable to deal with coincident disturbances. The time the PID output is held at the final resting value can be adjusted accordingly.

3. The step change in PID output upsets raw material, reagent, and utility headers. Directional move suppression can be added by the use setpoint rate limits and external reset feedback to minimize the disruption to other systems. See the discussions of external reset feedback and VPC for more details.
4. Operating point nonlinearity or disturbances invalidate final resting value. The final resting value should be captured for the last setpoint change at the same operating point in terms of PV and PID output to account for process and valve nonlinearities, respectively.

12.4 INTELLIGENT INTEGRAL ACTION

Intelligent integral action refers to the enhanced PID documented in Appendix E that suspends integral action until there is a measurement update. Setpoint changes still trigger an immediate reaction of the proportional and derivative modes depending upon PID structure. However, feedback correction is delayed until there is a measurement update. The contribution from the integral and derivative mode is based on the elapsed time since the last update. The contribution from the integral mode is the exponential response of the filter in the positive feedback implementation of the integral mode. Thus, integral action is suspended until there is new information on the PV response. The PID waits out any discontinuity in the valve due to backlash or stiction or in the measurement due to resolution limits, wireless transmitters, or at-line analyzers.

12.4.1 OPPORTUNITIES

Most of the opportunities stem from the ability of the enhanced PID to wait out any discontinuity in the valve due to backlash or stiction or in the measurement due to resolution limits, wireless transmitters, or at-line analyzers. Most benefit from the use of a threshold sensitivity limit to eliminate invalid measurement updates due to noise.

1. Elimination of the need to retune the PID to prevent oscillations from the increase in dead time due to wireless devices and at-line analyzers.
2. Ability to improve the PID response to provide nearly a single correction for a setpoint change or step disturbance by setting the PID gain to be the inverse of the open loop steady state gain when the dead time from the wireless device or at-line analyzer exceed the original 63 percent response time of a self-regulating process.
3. Elimination of limit cycles from valve backlash and stiction and measurement resolution limits by the suspension of integral action till the valve moves and the process responds.
4. Elimination of interaction between VPC and process loops whose valve position is being optimized by the judicious use of a threshold sensitivity limit to reduce overreaction.

12.4.2 WATCH-OUTS

1. Operating point nonlinearities and unknowns eliminate the ability to increase the PID gain to be the inverse of the identified open loop steady state gain. The enhanced PID gain must be decreased.

2. Measurement noise triggers an invalid measurement update. The threshold sensitivity limit must be increased.
3. Too slow a wireless default update rate and at-line analyzer cycle time and multiplex time will appreciably increase the integrated error for an unmeasured disturbance due to the increase in total loop dead time. Too fast an update will reduce the signal to noise ratio due to resolution limits and process noise. Wireless devices and at-line analyzers should provide updates with a frequency that maximizes the signal to noise ratio and minimizes dead time.
4. The actual improvement in process performance may not be visible because trend charts have the measurement delay introduced by the discontinuous updates. Use other measurements that are faster to show the actual process response.

12.5 DEAD TIME COMPENSATION

Dead time compensation in a PID with the positive feedback implementation of the integral mode is extremely simple. A dead time block just needs to be inserted in the external reset feedback signal. An underestimation rather than an overestimation of the total loop dead time should be used as the dead time block parameter. The reset time must be decreased toward a low limit of the PID execution time or dead time error, whichever is larger to see the improvement in the PID performance.

12.5.1 OPPORTUNITIES

Most of the opportunities here are in terms of this method of dead time compensation being easier to implement than a traditional Smith Predictor.

1. No need to model open loop process gain.
2. No need to model open loop time constant.
3. Operator sees true PV with standard PID faceplate.
4. Effect of directional move suppression and slow secondary loops or valves is included in the external reset feedback eliminating oscillations from the PID output trying to change faster than the manipulated variable can respond.

12.5.2 WATCH-OUTS

Most of the problems originate from a variable dead time and an untimely external reset feedback signal.

1. Dead time decreases by more than 10 percent or increases by more than 40 percent from identified value used in dead time block inserted in external reset feedback. Note that contrary to the effect of dead time on the tuning of PID feedback control, a decrease in dead time is more disruptive than an increase in dead time compared to the value used in a dead time compensator.

2. PID was not retuned to be faster principally by a decrease in reset time to take advantage of the dead time compensation. If the PID is not retuned, the performance is actually worse than if there is no dead time compensation.
3. External reset signal from valve being manipulated is too slow. The update by a secondary variable of actual valve position for a Highway Addressable Remote Transducer (HART) protocol device or a wireless device is presently too slow to be used for external reset feedback by most loops.
4. The loop is severely dead time dominant. Contrary to public opinion, the improvement in performance is not as good and the sensitivity to dead time errors is greater for loops where the dead time is much greater than the process time constant. The reset time of a properly tuned PI controller on a dead time dominant process cannot be decreased much by the use of just dead time compensation. However if the dead time is due to an analyzer, the use of an enhanced PID will enable a significant decrease in reset time and increase in PID gain. Contrary to common opinion, large time constant to dead time ratio (near-integrating) processes can have the greatest improvement from dead time compensation.

12.6 VALVE POSITION CONTROL

VPC can achieve optimization by the simple addition of PID controllers in the configuration. The VPC setpoint is the best throttle position of a valve that is constraining the optimization, the VPC PV is the valve input signal, and the VPC output is the PID setpoint being optimized (e.g., feed rate, compressor discharge pressure, or chiller discharge temperature). The use of the external reset feedback and the enhanced PID with directional move suppression is recommended.

12.6.1 OPPORTUNITIES

1. Maximization of valve rangeability and sensitivity by simultaneous stroking of a small and large valve. A large valve for course adjustment is installed in parallel with a small precise valve for fine adjustment. The large valve is manipulated by the VPC output to keep the small valve in the best throttle range (e.g., 50 percent).
2. Maximization of production rate by a VPC maximizing a feed flow setpoint to push a process control variable PID (e.g., reactor pressure or temperature) valve for quality control to a maximum throttle position (e.g., 80 percent for globe valve).
3. Minimization of reagent or fuel cost by a VPC maximizing the flow setpoint of the less expensive reagent or fuel to push a process control variable PID (e.g., neutralizer pH or furnace temperature) valve on the expensive resource to a minimum throttle position (e.g., 20 percent for globe valve).
4. Minimization of utility cost by a VPC lowering a compressor or pump discharge pressure or raising a chiller discharge temperature to push downstream user PID (e.g., reactor feed or temperature) valve to a maximum throttle position (e.g., 80 percent for globe valve).

12.6.2 WATCH-OUTS

The watch-outs mostly have to do the ability of the VPC to deal with interactions, disturbances, and nonideal valve responses.

1. Interaction between the VPC and the process loops that affect the VPC and are affected by the VPC output. The traditional solution of making the VPC closed loop time constant 10 times the slowest process PID causes a slow optimization and reaction to disturbances. The use of an enhanced PID and directional move suppression can enable faster tuning.
2. Fast large disturbances cause the process PID whose valve position is the VPC PV to run out of valve. Feedforward control should be used along with a faster response in the direction of correction for disturbances.
3. A flattening of the installed valve characteristic cause the process PID whose valve position is the VPC PV to run out of valve. Signal characterization should be used to reduce the valve nonlinearity as described in Chapter 9.
4. Limit cycles in the valve whose position is the VPC PV upset the PID whose setpoint is being optimized by the VPC rippling through the process. An enhanced PID should be used for the VPC and the process loops.

12.7 OVERRIDE CONTROL

Override control enables pushing a process to a constraint that can be a PID PV. Multiple constraints can be handled but are singly selected. The override PID controllers go to a signal selector before going to the setpoint of a process PID (e.g., feed flow) for optimization or protection. For optimizations with complex dynamics and objectives that are simultaneously addressing multiple constraints without overshoot, MPC is recommended.

12.7.1 OPPORTUNITIES

1. The inclusion of multiple constraints via multiple VPC to push several valves sequentially by signal selection to their limits for process optimization.
2. The prevention of abnormal conditions by the use of process PID monitoring key PVs (e.g., high reactor pressure or temperature).
3. The addressing of startup or transition conditions by the use of process PID monitoring overshoot (e.g., reactor pressure or temperature overshoot).
4. The protection against measurement failure by the use of inferential measurement as the PV of an override controller (e.g., PID with conservative computed composition as PV overrides the PID whose PV is from the failed analyzer).

12.7.2 WATCH-OUTS

1. Operator gets confused as to which override controller is controlling the process. The faceplate of the selected PID should be highlighted.
2. Incorrect connection of external reset signals from signal selector. Configuration must be thoroughly tested in a virtual plant with sufficient fidelity.
3. Walk-off of PID not selected in PID that does not have the positive feedback implementation of integral action. A filter with a time constant equal to the reset time can be inserted in the integral track signal but the better solution by far is a DCS with the positive feedback implementation of integral action.

4. Overshoot of constraint. The PID structure and tuning must minimize overshoot of the setpoint or the override must be modified to be inside the constraint.

12.8 TEST RESULTS

Test results were generated using a DeltaV virtual plant with the ability to set the process type and dynamics, automation system dynamics, PID options (structure and enhanced PID), PID execution time, setpoint lead-lag, tuning method, and a step change in lower PID load (ΔF_L) or upper PID setpoint (ΔSP). Table 12.1 summarizes the test conditions.

The same terminology is used as was defined for Table 1.2 for test results in Chapter 1.

The disturbance delay was 20 seconds for all the tests. This is the delay between the measurement of the load disturbance and the arrival at the same point in the process as the manipulated flow from the control valve. For these tests, the disturbance entered the process

Table 12.1. Test conditions

Figures	Process type	Open loop gain	Delay (sec)	Lag (sec)	Change	PID tuning
12.1a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 20\%$	Feedforward delay
12.1c, d	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Feedforward delay
12.1e, f	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	Feedforward delay
12.2a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 20\%$	Feedforward lag
12.2c, d	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Feedforward lag
12.2e, f	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	Feedforward lag
12.3a, b	Moderate self-reg.	1 dimensionless	10	20	$\Delta F_L = 20\%$	Feedforward gain
12.3c, d	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Feedforward gain
12.3e, f	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	Feedforward gain
12.4a, b	Near-integrating	1 dimensionless	10	100	$\Delta F_L = 20\%$	Full Throttle delay
12.4c, d	True integrating	0.01 1/sec	10	–	$\Delta F_L = 20\%$	Full Throttle delay
12.4e, f	Runaway	4 dimensionless	10	400	$\Delta F_L = 20\%$	Full Throttle delay

downstream of the process delay and secondary time constant but upstream of the primary process time constant. Consequently, the delay between a change in feedforward measurement and the arrival of the control valve flow is the sum of the feedforward signal delay and the process dead time. In this case the feedforward delay should be equal to the disturbance delay less the process dead time.

The disturbance lag was 10 seconds for all the tests. The 86 percent response time of the control valve was about two seconds. The manipulated flow from the control valve passed through a secondary process time constant of one second before arriving at the same point as the disturbance that is upstream of the primary process time constant. In this case the feedforward lag should be set equal to the disturbance lag (10 seconds) and the feedforward lead should be set equal to the control valve time constant (half the 86 percent valve response time) plus the secondary time constant for a total of two seconds.

The test results in Figures 12.1a, b, c, d, e, f and Figures 12.2a, b, c, d, e, f show that a feedforward signal that arrives too soon or too late due to a feedforward dead time or lag mismatch of 10 seconds will as expected cause an inverse response and an undershoot, respectively. The effect is worse for a moderate self-regulating process. For very fast processes and dead time dominant processes, the consequences of feedforward timing error are even more severe.

The test results in Figures 12.3a, b, c, d, e, f show that the consequences of a feedforward gain being too large are much more disruptive than a gain too low. If there is no initial response in the right direction and the response in the wrong direction is obvious, the feedforward gain is clearly too large.

The test results in Figures 12.4a, b, c, d, e, f show that a full throttle setpoint response offers the greatest improvement for a dead time used in the prediction of the future PV value that is 20 percent larger than the actual dead time.

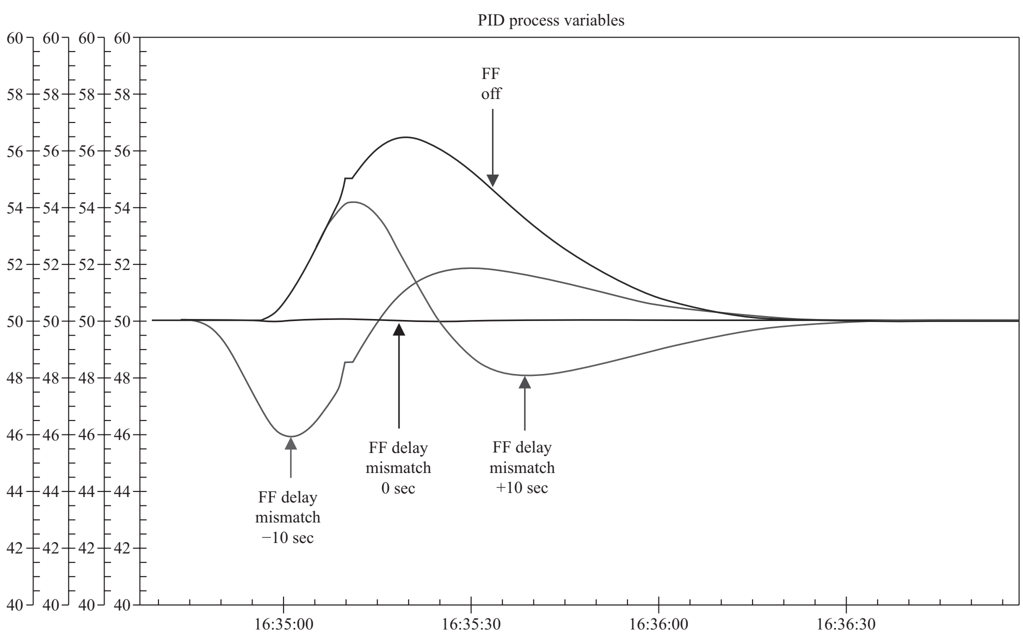


Figure 12.1a. Effect of feedforward delay mismatch on PID PV for moderate self-regulating process.

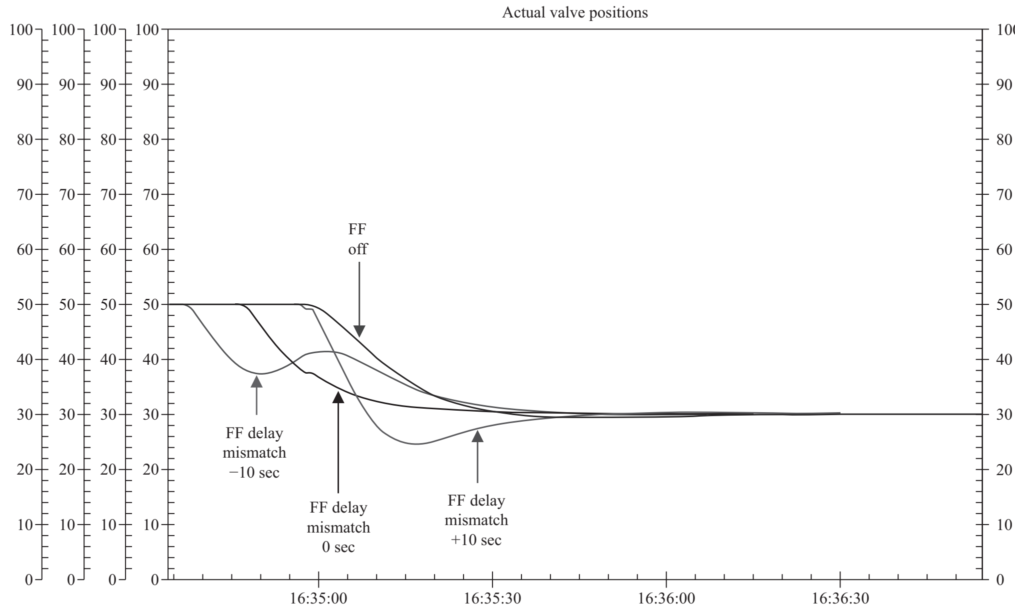


Figure 12.1b. Effect of feedforward delay mismatch on valve for moderate self-regulating process.

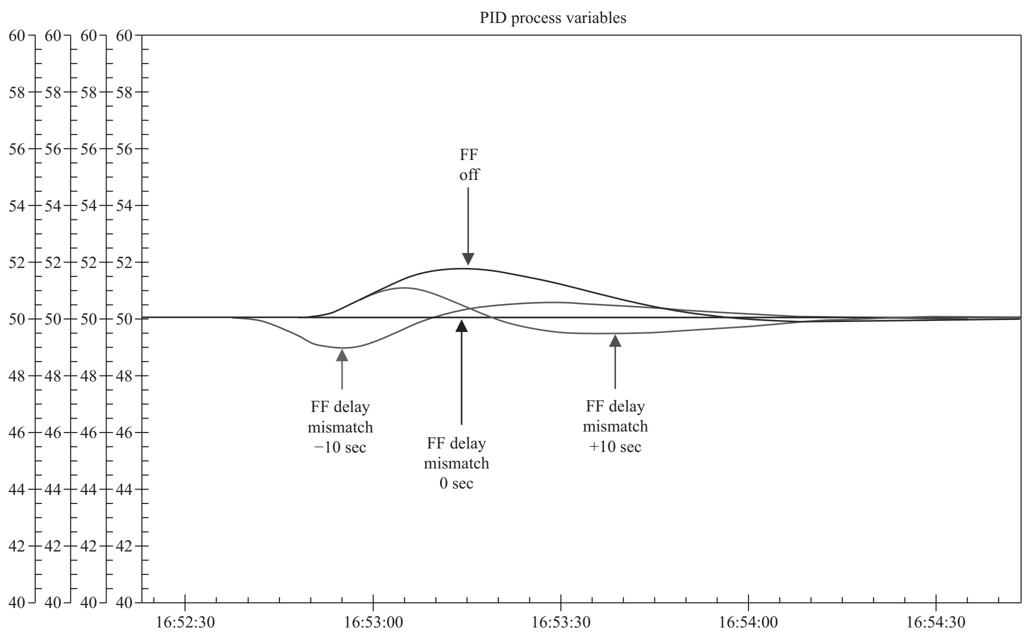


Figure 12.1c. Effect of feedforward delay mismatch on PID PV for near-integrating process.

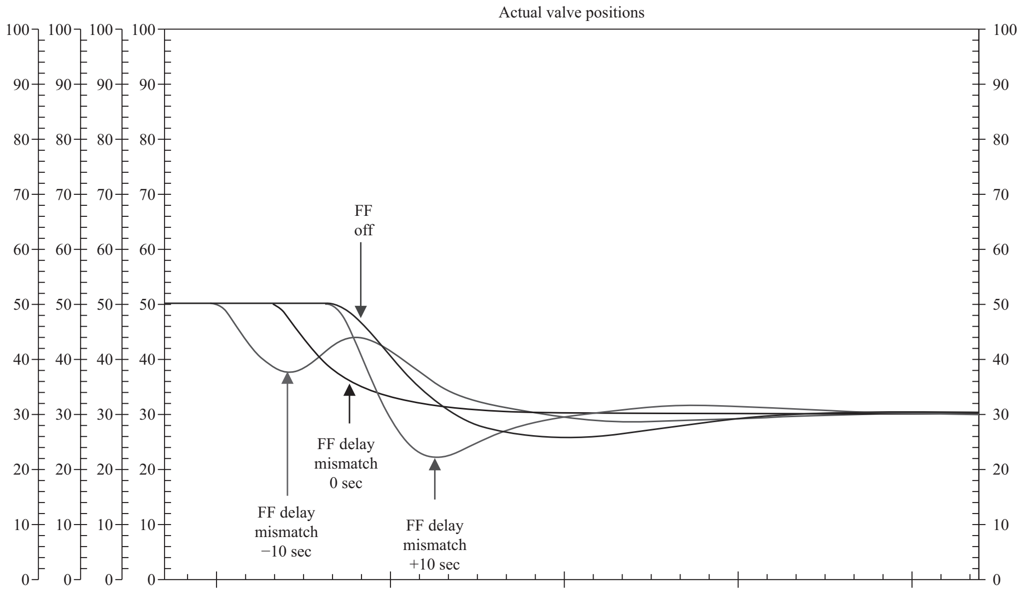


Figure 12.1d. Effect of feedforward delay mismatch on valve for near-integrating process.

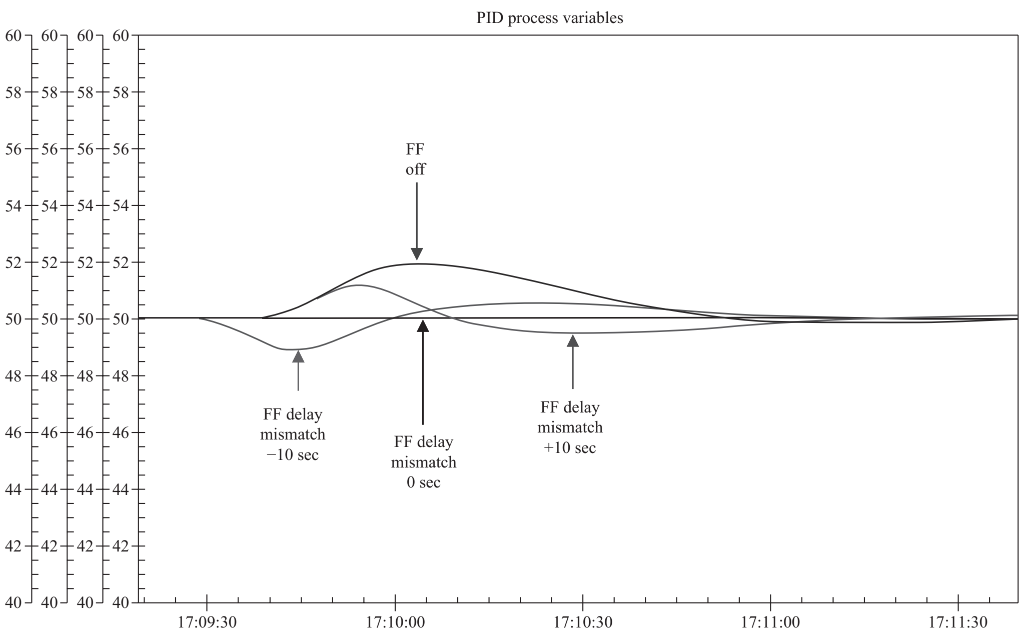


Figure 12.1e. Effect of feedforward delay mismatch on PID PV for true integrating process.

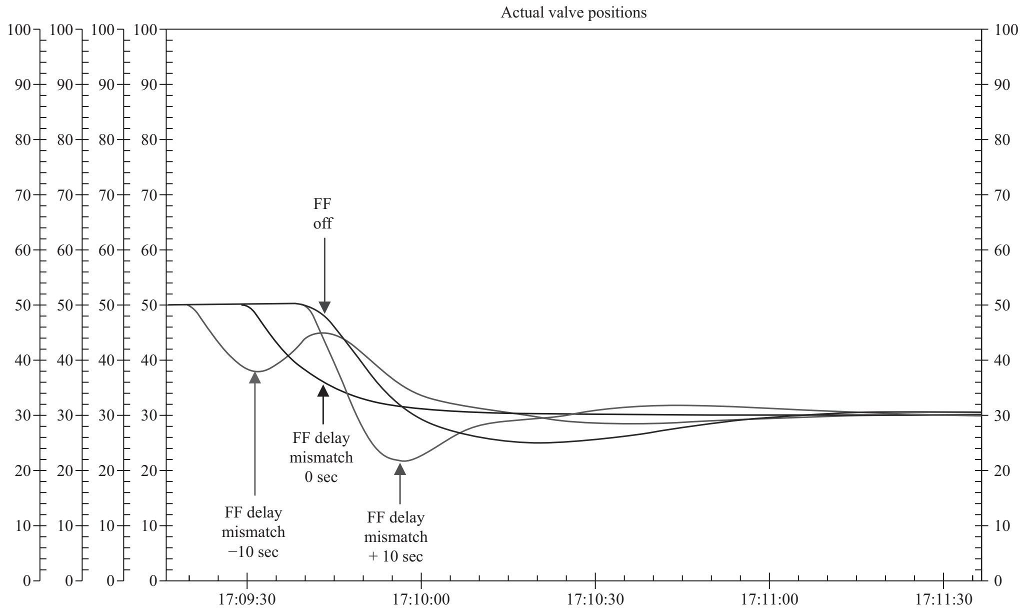


Figure 12.1f. Effect of feedforward delay mismatch on valve for true integrating process.

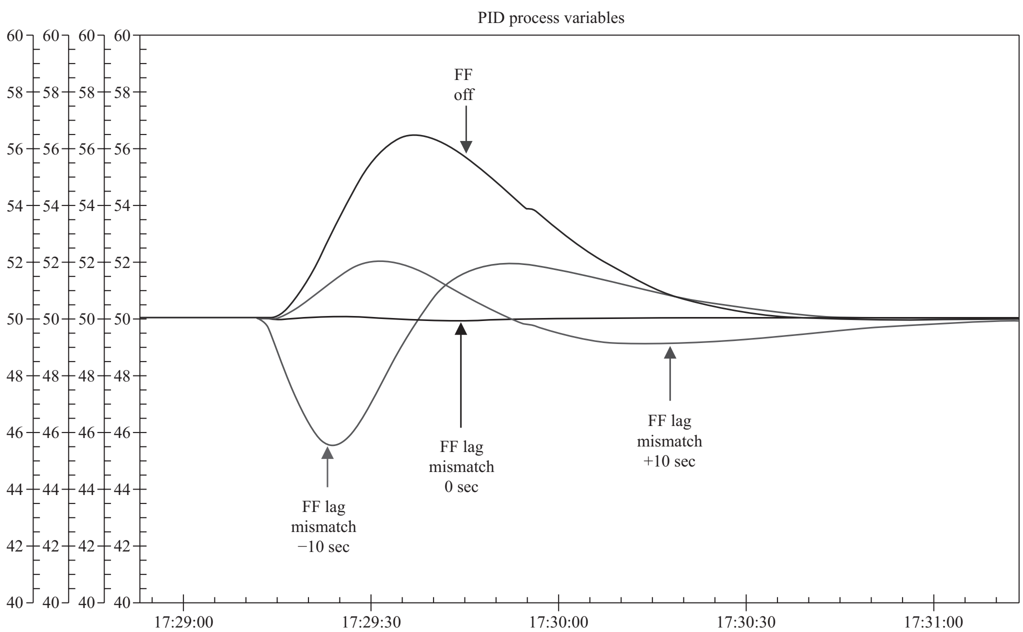


Figure 12.2a. Effect of feedforward lag mismatch on PID PV for moderate self-regulating process.

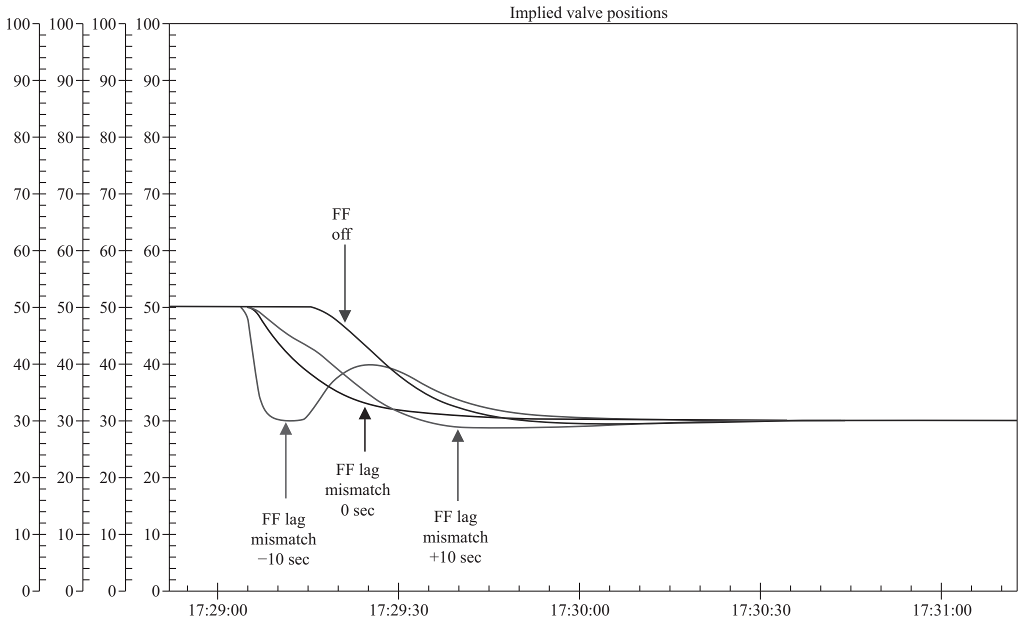


Figure 12.2b. Effect of feedforward lag mismatch on valve for moderate self-regulating process.

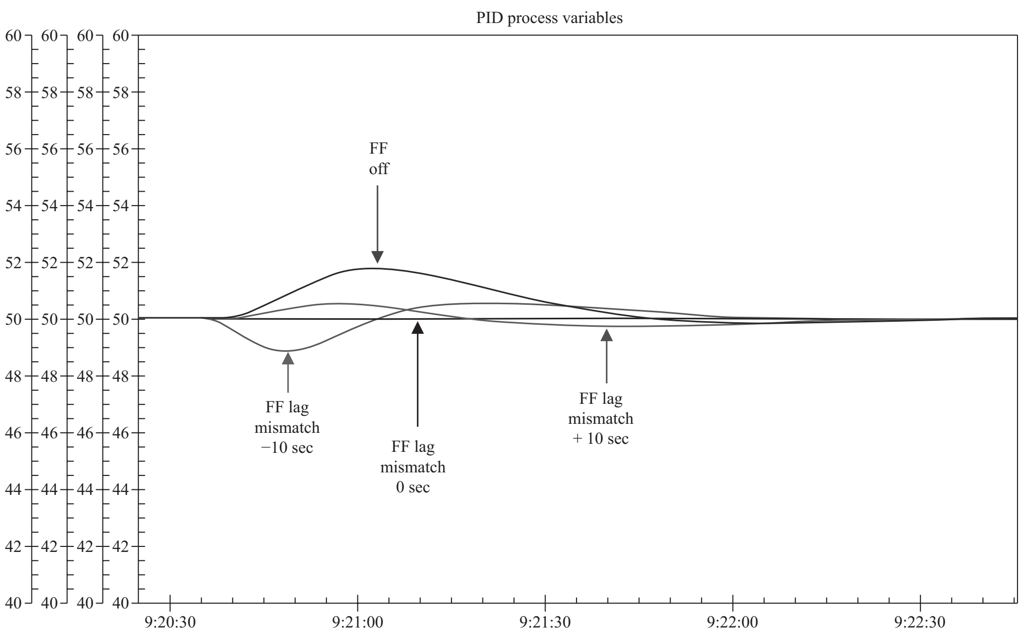


Figure 12.2c. Effect of feedforward lag mismatch on PID PV for near-integrating process.

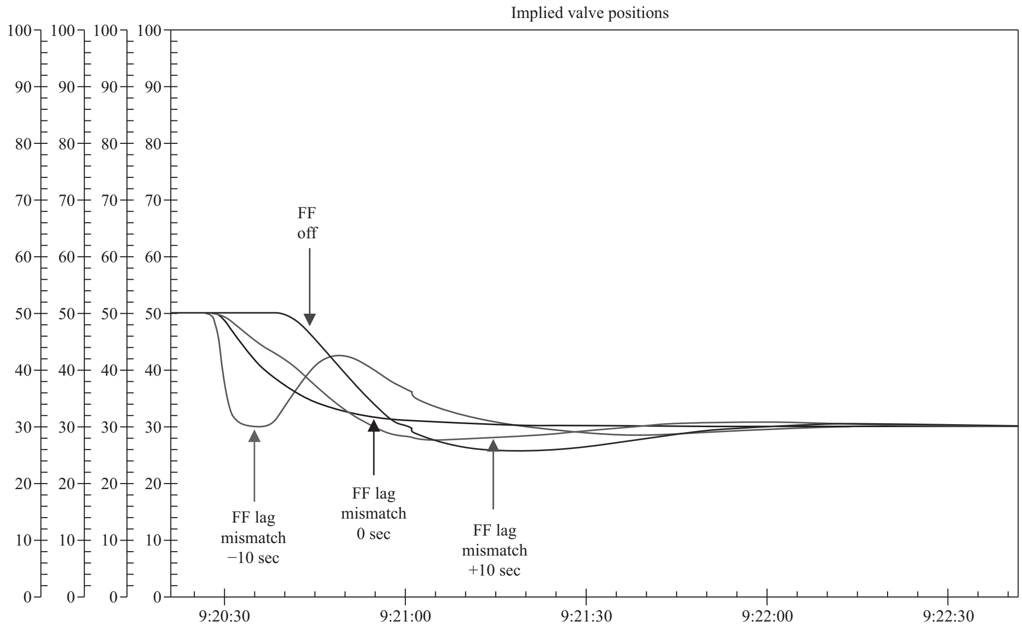


Figure 12.2d. Effect of feedforward lag mismatch on valve for near-integrating process.

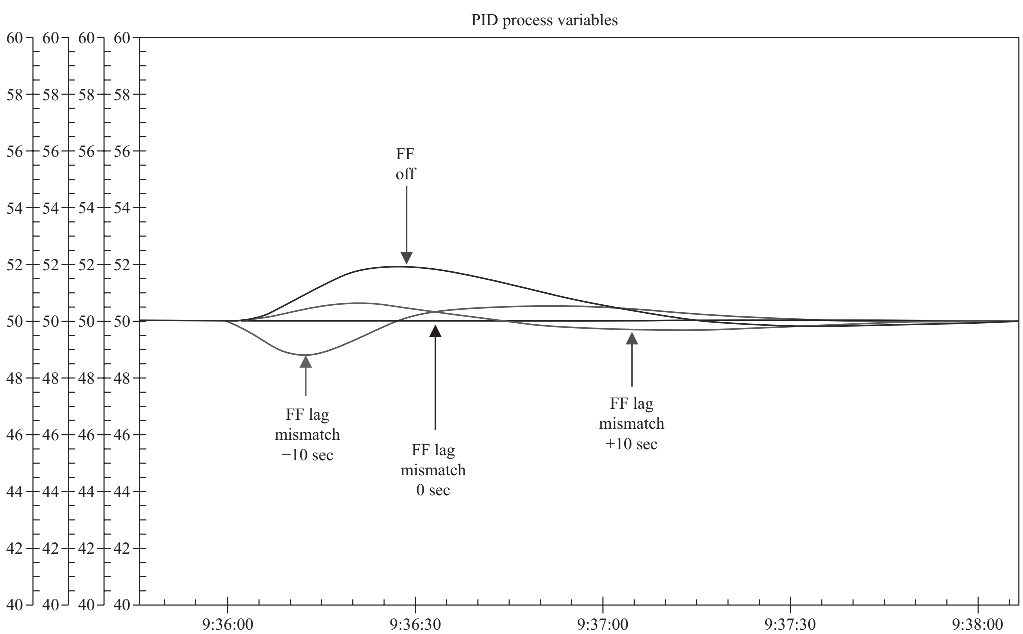


Figure 12.2e. Effect of feedforward lag mismatch on PID PV for true integrating process.

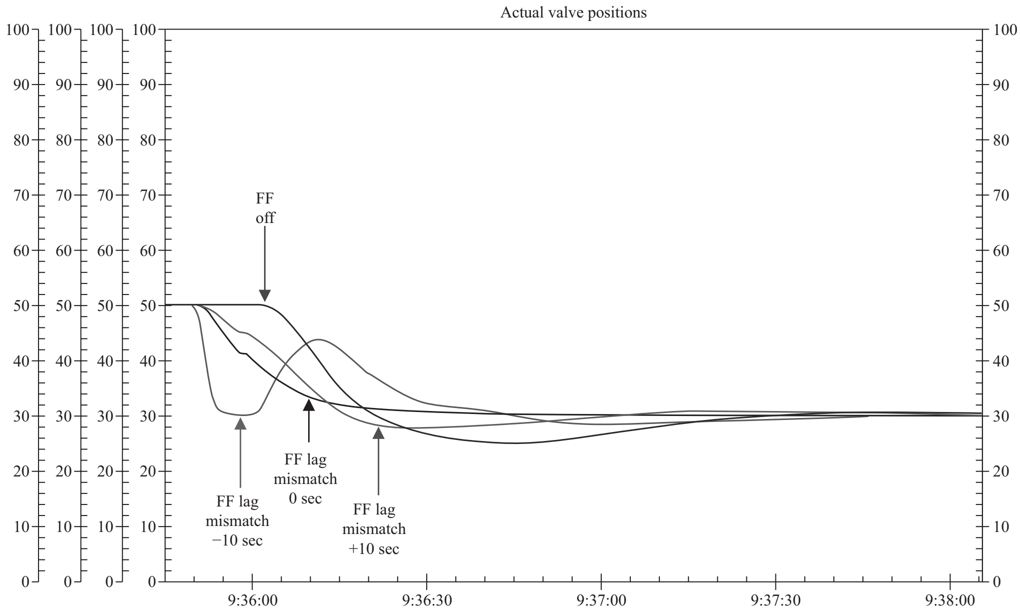


Figure 12.2f. Effect of feedforward lag mismatch on valve for true integrating process.

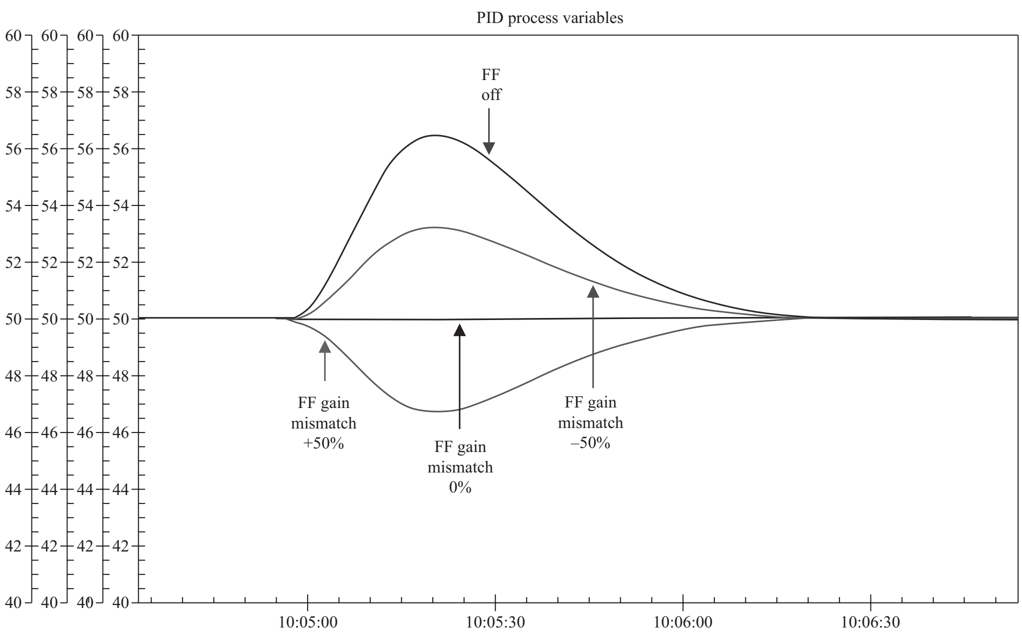


Figure 12.3a. Effect of feedforward gain mismatch on PID PV for moderate self-regulating process.

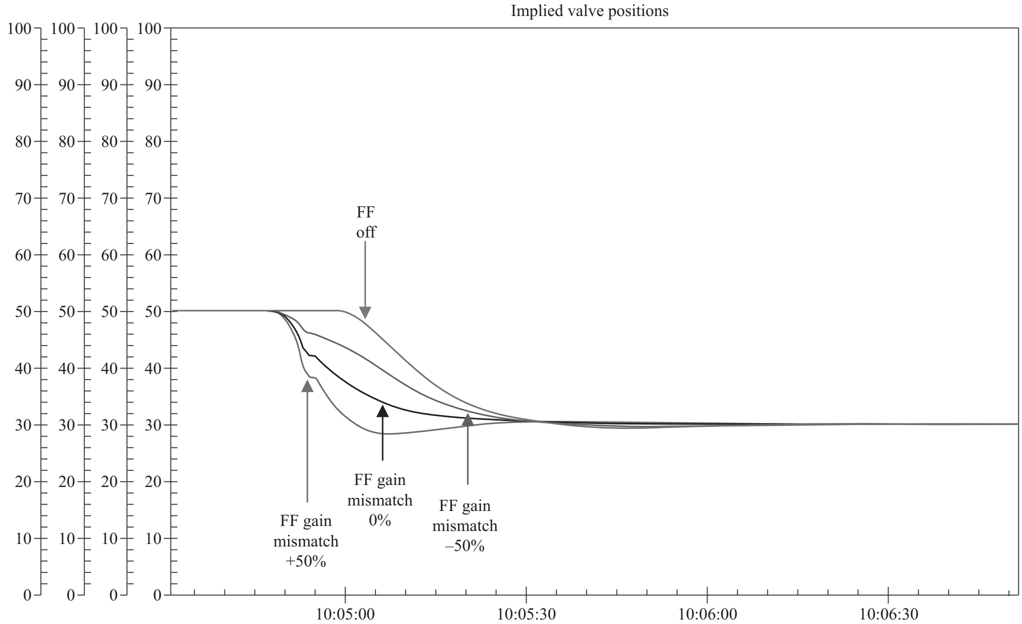


Figure 12.3b. Effect of feedforward gain mismatch on valve for moderate self-regulating process.

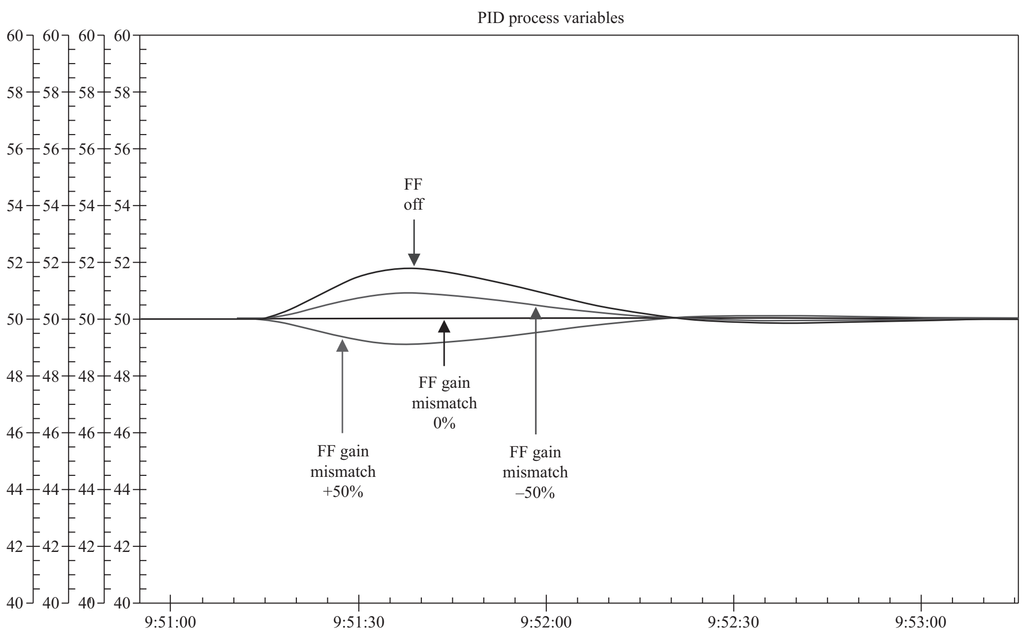


Figure 12.3c. Effect of feedforward gain mismatch on PID PV for near-integrating process.

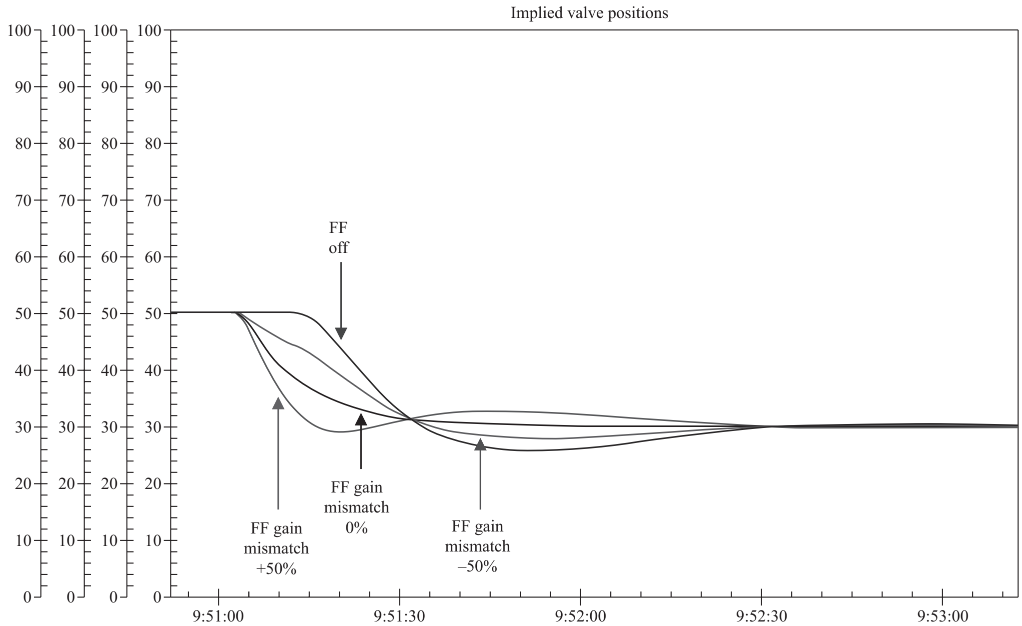


Figure 12.3d. Effect of feedforward gain mismatch on valve for near-integrating process.

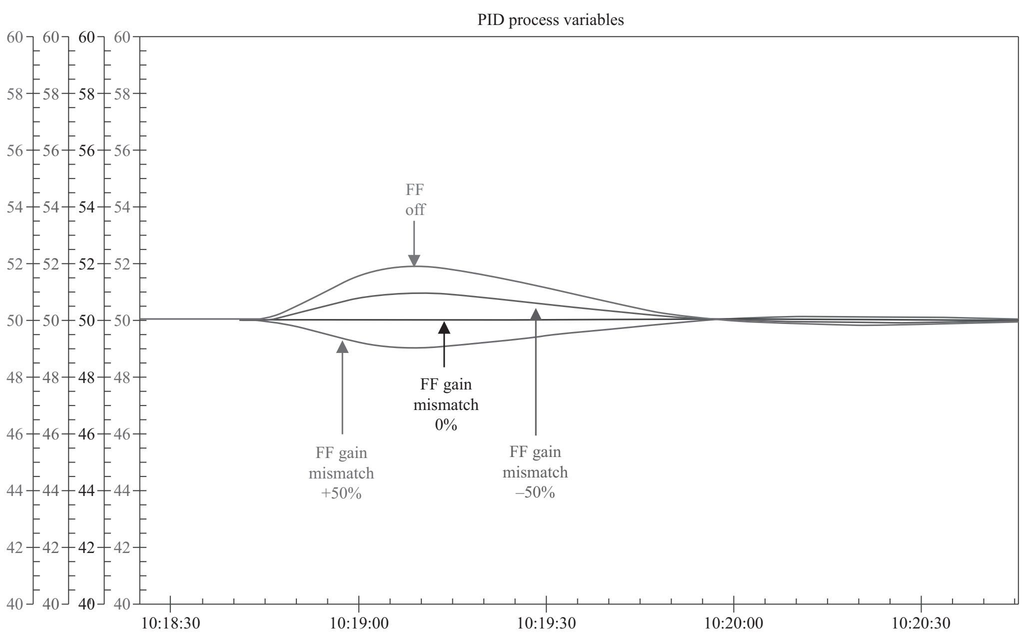


Figure 12.3e. Effect of feedforward gain mismatch on PID PV for true integrating process.

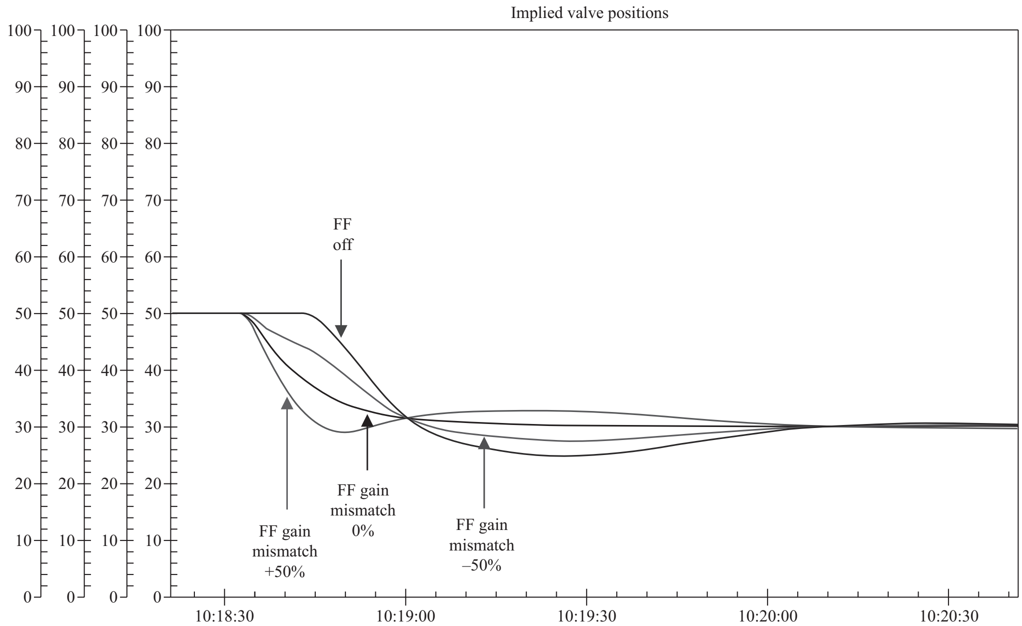


Figure 12.3f. Effect of feedforward gain mismatch on valve for true integrating process.

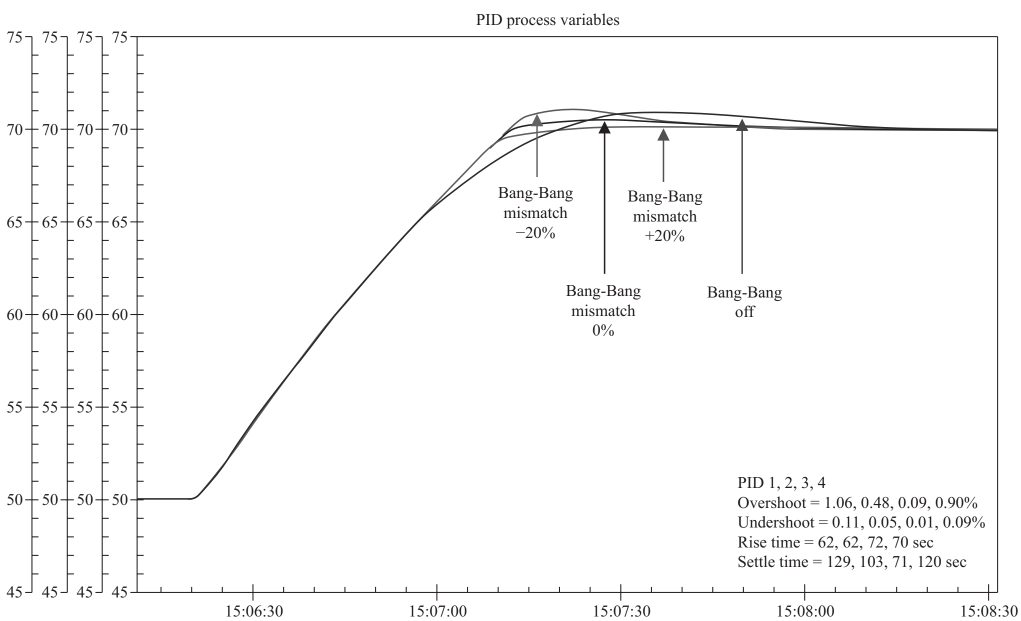


Figure 12.4a. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of PID PV for near-integrating process.

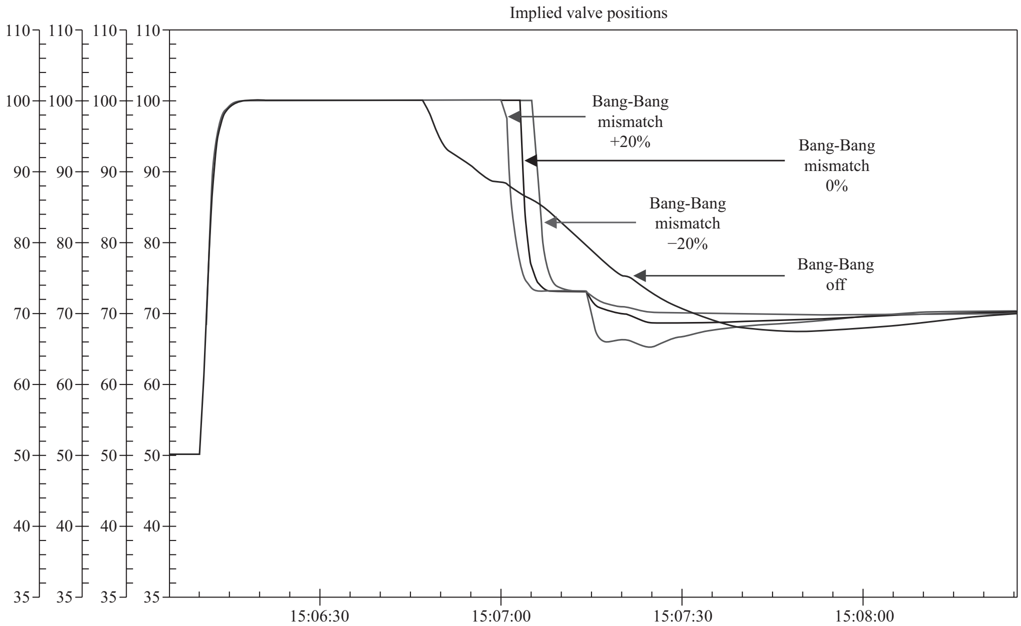


Figure 12.4b. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of valve for near-integrating process.

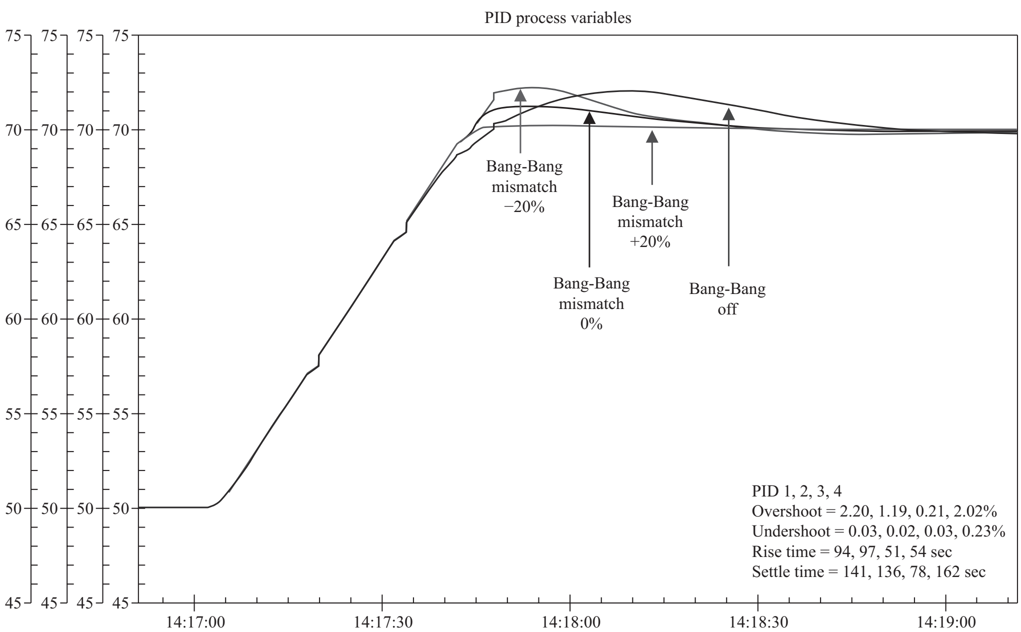


Figure 12.4c. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of PID PV for true integrating process.

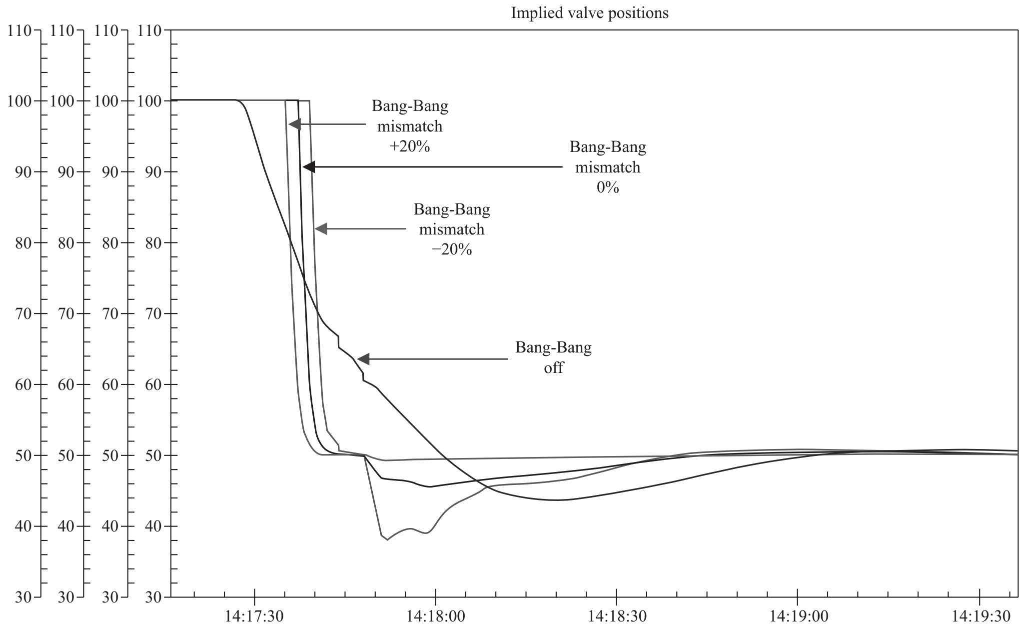


Figure 12.4d. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of valve for true integrating process.

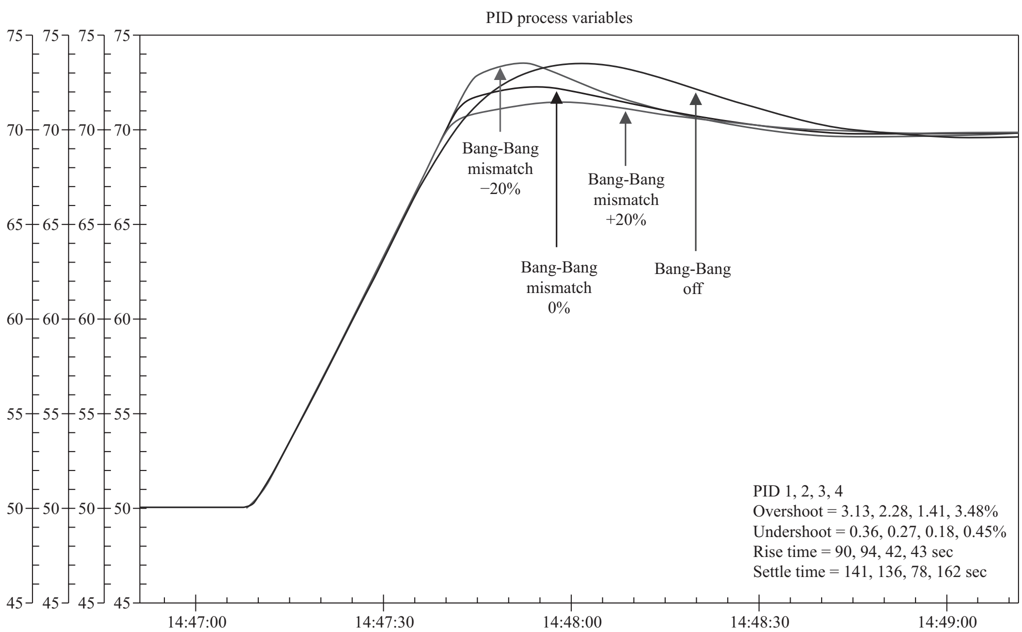


Figure 12.4e. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of PID PV for runaway process.

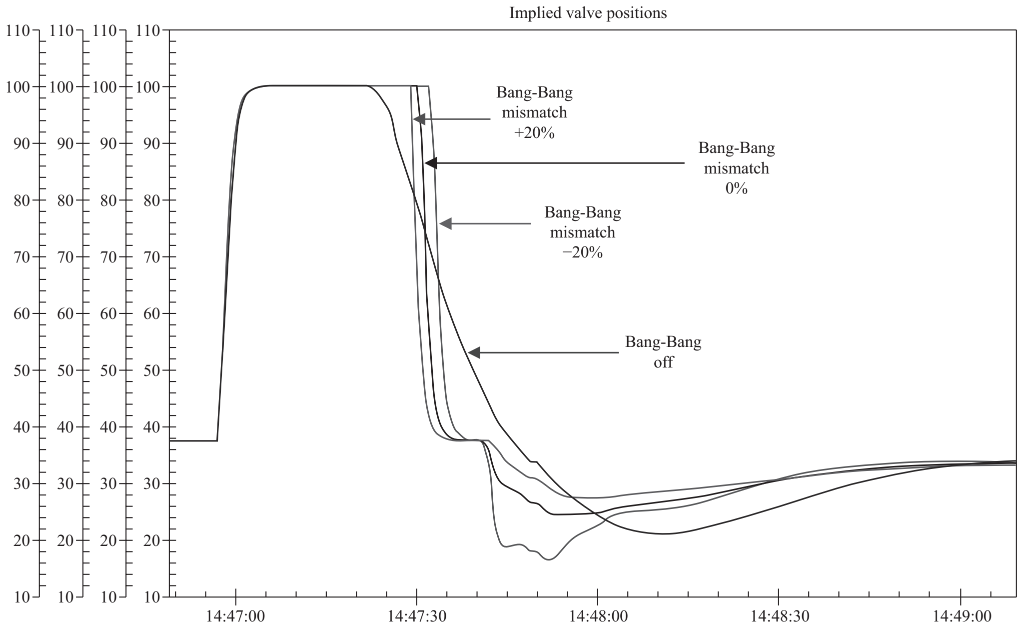


Figure 12.4f. Effect of process dead time mismatch on full throttle (bang-bang) setpoint response of valve for runaway process.

KEY POINTS

1. Feedforward control depends upon good feedforward measurements and dynamic compensation and particularly ensuring the correction does not arrive too soon.
2. Intelligent output action is critical for disturbances that come at a PID too fast or are too confusing (e.g., compressor surge).
3. Smart integral action via an enhanced PID opens the door for better wireless device, analyzer, and VPC.
4. VPC can achieve simple optimizations quickly without advanced control software costs or expertise.
5. Dead time compensation only provides a significant benefit if the dead time is accurately updated and PID the reset time is decreased. The sensitivity to dead time error is highest and the possible decrease in reset time is least for a dead time dominant process. A more effective solution for a dead time dominant process is an enhanced PID if most of the dead time is due to an analyzer.
6. Override control can achieve simple optimizations with simple dynamics and dealing with unusual conditions by honoring constraints singly selected.

CHAPTER 13

Process Control Improvement

13.1 INTRODUCTION

The benefits from process control improvement (PCI) originate from increases in process efficiency, flexibility, and capacity. Often, there is a tradeoff where an increase in flexibility or capacity is accompanied by a decrease in efficiency. Better measurements, process knowledge, online metrics, and an enhanced proportional-integral-derivative (PID) can provide a more intelligent and effective optimization that minimizes the tradeoff between benefits.

13.1.1 PERSPECTIVE

Process knowledge improves as measurement accuracy is increased. Simulations are better verified, material and energy balances closed, better setpoints found, and more accurate metrics computed. The door is then opened to more effective PCI. The variability can be decreased as much as necessary by better basic and advanced regulatory control. Operator biases (comfort zones) can be eliminated by automated protection to enable the setting of setpoints closer to the optimum. The setpoint proximity to the optimum can be optimized by override and valve position control. The benefits can be ball parked first as an opportunity sizing and then estimated by an opportunity assessment. The documenting and reporting of the actual benefits gained enables support for future opportunities.

You cannot control what you do not measure. Online metrics are essential for developing, maintaining, and justifying PCI. Nothing speaks as powerfully as money. Online process metrics are essential for focus, understanding, implementation, achievement, and recognition by the automation engineer and the profession.

Efficiency is the largest factor in the variable costs for the cost of goods sold (COGS) in manufacturing processes. The efficiency in the use of each raw material and utility stream for each unit operation should be computed online and made available to the operator real time. Normally this is done on the basis of the mass of each raw material and the energy of each utility stream used per unit mass of product. For batch processes, mass and energy use are totaled for each batch. Batch energy use is normally much less of a concern than capacity as indicated by batch cycle time. For continuous processes, the use is computed on a flow rate basis often requiring some intelligent filtering and synchronization of input and output flows to eliminate noise and a confusing irregular or inverse response.

Attaining and maintaining optimum operating conditions requires knowing the optimum setpoints and pushing the actual setpoints as close as possible to the optimum based on operability and variability. The optimum can change with production rate, raw materials and impurity, feed stock, product type and grade, recycle stream, ambient conditions, and the operating conditions for the process and utility systems. The efficiency during startup and transitions (changes in feedstock, product type, and product grade) should be computed and included in hourly, shift, daily, weekly, and monthly averages.

Flexibility corresponds to the ability to quickly and efficiently change production rate, feed stock, product type, and product grade based on market demand. The goal to minimize inventory makes flexibility more important and more difficult.

Capacity is affected by production rate and on-stream time. The production rate for batch processes depends upon batch cycle time and wait time. For continuous processes, the final product flow is the production rate. Start-up time, downtime, and transition time undermine on-stream time. The production rate for all types of operation should be computed and included in hourly, shift, daily, weekly, and monthly averages.

A valve position controller (VPC) can maximize production rate or minimize energy use by maximizing a feed rate or coolant supply temperature or minimizing a compressor discharge pressure. The main problem is the complexity of the tuning to eliminate interaction between the VPC and process loops and the resulting slow VPC response to upsets. The key here is an enhanced PID for the VPC to suppress oscillations, make tuning simpler, and enable move suppression via setpoint rate limits on the PID being manipulated by the VPC for smooth optimization and fast disturbance rejection.

To enable greater flexibility in meeting different production rates, feedforward flow control should be used. Part of this realization is the recognition that nearly all loops responsible for maintaining stream conditions (e.g., composition and temperature) are a function of the ratio of manipulated flow to feed flow and all loops responsible for inventory (e.g., liquid level and gas pressure) are a function of the difference between the manipulated flow (e.g., exit flow) and the counteracting flow (e.g., entrance flow) for the given volume. Keeping the flow ratio constant for stream conditions and keeping the entrance and exit flows equal for inventories are the keys to tight control. Flow feedforward is the most underrated advanced regulatory control technique. Also, the need to startup on flow ratio control and provide the operator the ability to see the actual ratio and correct the ratio setpoint has somehow been lost. The interfaces and functionality developed to provide this functionality in the 1980s need to be standard Distributed Control System (DCS) features.

The frequency and amplitude of oscillations are signatures of different problems. Statistical methods can provide the distribution of amplitudes and power spectrum analyzers can provide the relative contribution at particular frequencies. These tools should be used to analyze violations of constraints and the results included in shift, daily, weekly, and monthly averages with the corresponding frequency noted.

The sources of oscillations can often be tracked down based on oscillation frequency (oscillation period) compared to the subject loop's natural frequency (ultimate period). Often a surge tank level controller tuned too aggressively is the culprit. The other major sources of oscillations are loops with too much reset action (too small a reset time) and valves with excessive backlash or stiction. With insight as to candidates for sources of the oscillations based on an oscillation frequency, straight forward troubleshooting techniques should be used such as putting individual loops in manual until the oscillation stops. If oscillations persist with loops in manual then the environment, equipment, or process are the usual suspects.

The attenuation of oscillations by the filtering effect of intervening volumes should be estimated. The ratio of the output to input amplitude is inversely proportional to a fraction of the volume residence time (volume/flow). For well mixed volumes the fraction can be taken as one. The effect of blending of multiple production lines in surge and storage tanks is also significant in terms of reducing variability.

13.1.2 OVERVIEW

An improvement in measurement accuracy enables more accurate material and energy balances, statistical analysis, and metrics. Routine verification of the calibration of sensors on-line and off-line can ensure a more reliable and effective measurement. The resolution, threshold sensitivity, accuracy, and precision should be determined by measurement system analysis. The use of the latest transmitter and sensor technology can significantly reduce the need for recalibration. Some sensors offer much greater threshold sensitivity and sustained accuracy (e.g., less drift and installation effects) such as resistance temperature detectors instead of thermocouples and Coriolis meters instead of vortex meters.

The optimum setpoint can be found from historical data, test results, data analytics, and simulation confirmed by process analysis through a comprehensive design of experiments (DOE) in a virtual plant and if necessary a focused DOE in the actual plant. Unnecessary setpoint biases by operations are revealed in the differences in the performance of various shifts. The reasons for the perceived need for the comfort zone should be identified and addressed by better automation. Finding the optimums via historical data is more difficult for processes that run consistently because there is not enough change in the data.

Process and automation system performance can get better or worse after maintenance or changes in measurements, final control elements, process, piping design, configuration, and tuning. Periods of maintenance and changes should be noted and performance analyzed before and after.

Determine if you need to improve efficiency, flexibility, or capacity. Efficiency and capacity could be both affected by flexibility if you consider that the rework produced during a change-over will affect your performance metrics and that during transitions you could be forced to decrease your current production rate. Although you might be tempted to work on all of them at once, consider establishing priorities or get overwhelmed otherwise. If different teams are formed to work in parallel, they might find each other affecting or interfering with others' objective.

Opportunity sizing consists of finding the best periods of operation in terms of efficiency and capacity from simulations, historical data, online metrics, and cost sheets. These best periods represent the entitlement of the process or the desired target of operation. When compared with the current baseline, the resulting gap represents an initial and potential window for improvement. The revenue minus the cost of goods can be computed for these periods for better analysis of efficiency, flexibility, and capacity tradeoffs. The gaps between the best and average plant performance is the sizing.

Once the objective and the opportunity sizing have been established, make sure the team is all onboard. In a meeting with key people from instrument and electrical, analyzer technology, quality assurance lab, process technology, research and development (R&D), operations, maintenance, configuration, and process control, identify possible solutions for each gap in the opportunity sizing. Build process control diagrams with relative locations shown in piping and equipment of all the control loops, measurements, and final control elements (valves and variable speed drives). These diagrams should be done for utility systems as well. Provide online access to trend charts, troubleshooting results, online metrics, data analytics, statistical measures, power spectrum analyzers, and auto tuner or adaptive tuner results. For each possible solution, have the key process technology person estimate the portion of the gap that can be eliminated. A Cause and Effect Matrix could be used as an initial guide to each process variable (PV) that could affect the efficiency, flexibility, or capacity. An exploratory graphical analysis using trend plots, main effect plots, and matrix plots can help evaluate the relationships.

Have the automation system people estimate the rough time and order of magnitude installed costs (total of hardware, design, and installation costs). Assign priorities and people based on the type of improvement and the possible benefits, time, and cost.

Oscillations are the principal causes of variability. The first thing to do is to put the subject loop in manual and see if the oscillations are significantly reduced. If the oscillations decay, the problem is in the loop tuning, measurement of final control element. If the oscillation amplitude decreases, the subject loop is amplifying them.

When the culprit loop is found, an order of magnitude or more increase in the reset time can reveal whether the solution is tuning, less deadband, better resolution, or better threshold sensitivity. If the oscillations do not originate in a loop, then process instabilities, mechanical deficiencies, changes in phase and poor mixing are sources of fast oscillations. Recycle streams, batch operations, on-off control, safety instrumented system (SIS), relief devices, cyclic unit operations (defrosting and regeneration), tank car deliveries, and ambient conditions are likely causes of slow oscillations.

Performance must be analyzed preferably by online metrics immediately before and after PCI so that benefits can be properly assigned. There are many people making plant improvements anxious to take credit for benefits.

13.1.3 RECOMMENDATIONS

1. Work with the automation system supplier (e.g., representative or local business partner) to estimate and improve installed accuracy of measurements.
2. Use internal resources such as the Quality Assurance Lab to evaluate critical on-line and at-line measurements particularly those for composition and pH.
3. Use historical data, data analytics, simulation, design of experiments, and process analysis to find optimum setpoints.
4. Install online process metrics.
5. Conduct an opportunity sizing and assessment.
6. Use Cause and Effect Matrix to identify all inputs to the process that could be affecting efficiency, flexibility, or capacity.
7. Initially filter out those variables with minimum impact using exploratory graphical analysis.
8. Use principal component analysis (PCA) and projection to latent structures (PLS) as part of a data analytics tool to find important correlations.

9. Track down and eliminate or mitigate the sources of oscillations.
10. Find and address reasons for offsets from the optimum introduced by operations.
11. Implement basic and advanced regulatory control improvements.
12. Improve setpoints and document and report benefits.

13.2 UNIT OPERATION METRICS

You cannot control what you do not measure. Sometimes we lose track of this principle for the most important indication that is plant performance. To achieve this capability, we need a building block approach where unit operation performance metrics lead to a better understanding and more accurate production unit metrics that ultimately lead to plant performance metrics. Here we focus on the starting point of this incredibly important approach by discussing how to generate unit operation metrics. In principle, these dynamic metrics could become optimization variables for model predictive control (MPC).

The use of wireless instruments offers the possibility to have a portable set of instrumentation (e.g., annubar, differential pressure transmitter, resistance temperature detector and transmitter, and pH electrode) for finding metrics if the process connections (e.g., nozzles) exist. Future piping designs should take this opportunity into consideration.

Processes are analyzed and debottlenecked on a unit operation basis. Cause and effect relationships are best deciphered by being able to focus on the biological and chemical engineering principles for key unit operations. For stirred reactor unit operation considered here, there is a liquid and vapor phase exiting streams (bottoms liquid stream and a top distillate stream that starts out as a vapor from the top of the unit operation).

The running mass totals are the totalized mass flow at the end of a batch operation or the totalized flow for the last X hours of a continuous operation. The running total might be for the last number of hours corresponding to a shift to allow comparison of shift performance and the detection of changes at the beginning of a shift (common occurrence). While operations may balk, the competitive nature of people should eventually lead to a desire to do better.

For batch operations, everything is computed at the end of the batch. There is no time synchronization of changes in input streams with the resulting effects on the output stream needed in the computation of metrics. Consequently, the metrics for batch operations can be more much more consistent with less noise than the metrics for continuous operations.

For continuous operations, inputs must pass through a dead block to provide the time delay corresponding to a transportation delay for plug flow volumes and then a filter block with a time constant corresponding to the primary process time constant for well mixed volumes between the input stream and the output stream. The dead time is the plug flow residence time and the time constant is the back mixed volume residence time. This synchronization of the changes in input streams with the changes in output streams for continuous operations is essential to prevent inverse response in metrics. The time synchronization is a function of production rate since the residence time to the corresponding intervening volume is divided by the flow rate. Since the synchronization is never perfect, extensive filtering of the metrics for continuous operations is necessary to avoid confusion. A filter time constant about equal to one-fifth the residence time should be sufficient if the residence time is updated based on level and production rate.

The need for synchronization makes the use of data analytics and neural networks much more difficult for continuous process especially when they span several unit operations. While the temptation is to try and predict changes in the final product from changes upstream in a single overall calculation, the more practical approach is to do the predictions on a unit operation basis with a building block approach similar to what is proposed here for metrics. As an example we consider a stirred reactor.

To simplify the problem we will consider a batch or continuous reactor with multiple feed streams but with single streams exiting the bottom as liquid and the top as distillate. The batch or running totals of each product P_i in a feed flow (F_F) from upstream operations and recycle flow (F_R) as result of recovery operations, exiting via the bottom flow (F_B) and exiting via a top distillate stream (F_D) are computed via Equations 13.1a through 13.1d. The net mass of product P_i produced (M_{P_i}) is computed as per Equation 13.1e as the total mass of the product exiting the bottom and top less the mass of product entering via the feed and recycle streams. The production rate as per Equation 13.1f is this total mass of net product produced divided by the time interval (Δt) used for the running totals (e.g., batch cycle time or continuous shift time). Note that the product exiting the reactor may be an intermediate that requires further processing before becoming saleable. Recovery and purification operations are normally needed to remove excess raw materials, impurities, and solvent. The impurities can be inert components that do not enter into the reaction or waste products created by undesirable reactions or product degradation.

$$M_{FP_i} = \int_0^{\Delta t} F_F * X_{FP_i} * dt \tag{13.1a}$$

$$M_{RP_i} = \int_0^{\Delta t} F_R * X_{RP_i} * dt \tag{13.1b}$$

$$M_{BP_i} = \int_0^{\Delta t} F_B * X_{BP_i} * dt \tag{13.1c}$$

$$M_{DP_i} = \int_0^{\Delta t} F_D * X_{DP_i} * dt \tag{13.1d}$$

$$M_{P_i} = M_{BP_i} + M_{DP_i} - M_{FP_i} - M_{RP_i} \tag{13.1e}$$

$$F_{P_i} = \frac{M_{P_i}}{\Delta t} \tag{13.1f}$$

Equations 13.2a through 13.2c can be used to determine the amount of each reactant R_i consumed in the reaction. Reactants appear in the bottom or top stream due to incomplete reactions or reactant feed added in excess of what is needed as per reaction stoichiometry. Almost all reactors operate with an excess of at least one reactant. Also only in batch reactions with sufficient cycle time are reactants reacted. In continuous unit operations with stirred reactors, some reactants exit quickly because these reactants go from the reactor inlet to outlet connections in about one turn over time due to agitation or even sooner due to vaporization. The fact that there is a spectrum of possible residence times for any given molecule due to agitation patterns implies a significant amount of reactants will be in the bottom stream unless the residence time is much larger than the reaction time. While the amount of reactants recovered via recycle streams proportionally reduce the amount of reactants that must be provided in the feed streams, the overall efficiency of the plant is degraded because of the energy costs of the recovery operations. Raw materials not recovered increase the running total of raw material used since the raw material feed must make up for the excess raw material exiting the reactor that is lost.

$$M_{FR_i} = \int_0^{\Delta t} F_F * X_{FR_i} * dt \tag{13.2a}$$

$$M_{RR_i} = \int_0^{\Delta t} F_R * X_{RR_i} * dt \tag{13.2b}$$

$$M_{R_i} = M_{FR_i} + M_{RR_i} \tag{13.2c}$$

The kilogram of each raw material used per kilogram of all products from the reactor is computed by Equation 13.3a as the ratio of the running total of mass of raw material in the feed and recycle streams divided by total mass of all products in the bottoms and distillate. The raw material efficiency is computed in Equation 13.3b from current ratio (R_{R_i}) and the minimum ratio ($R_{R\min_i}$) for raw material Ri based on the reaction stoichiometry.

$$R_{R_i} = \frac{M_{R_i}}{\sum_{i=1}^{n} M_{P_i}} \tag{13.3a}$$

$$E_{R_i} = 100 * \left[1 - \frac{R_{R_i} - R_{R\min_i}}{R_{R\min_i}}\right] \tag{13.3b}$$

The total energy is the sum of the running total of the energy to heat or cool each feed flow and recycle flow and utility flow to heat or cool the reactor contents. The running total of energy use for each stream is computed by Equations 13.4a through 13.4c as the totalized product of the flow and the change in enthalpy of the respective stream. Equation 13.4d computes the running total energy required for the reaction (Q_X) as the sum of these energy totals and Equation 13.4e computes the kilojoule of energy required per kilogram of all products from the reactor. The energy efficiency is computed in Equation 13.3b from current ratio (R_Q) and the minimum ratio ($R_{Q\min}$) of energy use based on the reaction stoichiometry and an energy balance. For highly exothermic reactors that are generating steam (e.g., fluidized bed monomer reactors) the metric changes to take into account how much energy is produced. In this case the goal is to maximize the running total of net energy produced (Q_X) and as seen in Equation 13.4f maximizing the current ratio (R_Q) relative to the maximum ratio ($R_{Q\max}$). Higher energy efficiency computed by either Equation 13.4f or 13.4g can generally be achieved by running at the optimum stoichiometric and temperature conditions.

$$Q_F = \int_0^{\Delta t} F_F * \Delta H_F * dt \tag{13.4a}$$

$$Q_R = \int_0^{\Delta t} F_R * \Delta H_R * dt \tag{13.4b}$$

$$Q_U = \sum_{i=1}^{n} \int_0^{\Delta t} F_{U_i} * \Delta H_{U_i} * dt \tag{13.4c}$$

$$Q_X = Q_F + Q_R + Q_U \tag{13.4d}$$

$$R_Q = \frac{Q_X}{\sum_{i=1}^{n} M_{P_i}} \tag{13.4e}$$

For water heated or steam condensed in reactor temperature control:

$$E_Q = 100 * \left[1 - \frac{R_Q - R_{Q\min}}{R_{Q\min}}\right] \tag{13.4f}$$

For steam generated in reactor temperature control:

$$E_Q = 100 * \left[1 - \frac{R_{Q\max} - R_Q}{R_{Q\max}}\right] \tag{13.4g}$$

Note that there are considerable costs in downstream unit operations for the separation crystallization, separation, and drying of product. There are also downstream costs of separation and recovery or waste treatment of solvents, impurities, and off-spec product.

Off-spec product can sometimes be sold at a lower price. In the case of sheet lines, the off-spec product is recovered and recycled as a raw material to the extruder that feeds the sheet line. A decrease in off-spec product here can be taken either as a capacity increase by keeping the same raw material feed rate or as an efficiency increase in raw material use from decreasing the raw material required by keeping the same production rate.

Nomenclature:

E_{R_i} = efficiency of raw material i use (%)
E_Q = efficiency of energy use or production (%)
F_B = bottoms flow exiting reactor (kg/sec)
F_D = distillate flow exiting reactor (kg/sec)
F_F = feed flow entering reactor (kg/sec)
F_R = recycle flow entering reactor (kg/sec)
F_{P_i} = production rate of product i by reactor (kg/sec)
F_{U_i} = utility i flow (e.g., steam, cooling water) used by reactor (kg/sec)
M_{BP_i} = running total of product i mass in bottoms flow exiting reactor (kg)
M_{DP_i} = running total of product i mass in distillate flow exiting reactor (kg)
M_{FP_i} = running total of product i mass in feed entering reactor (kg)
M_{RP_i} = running total of product i mass in recycle entering reactor (kg)
M_{P_i} = running total of product i produced by reactor (kg)
M_{R_i} = running total of raw material i used by reactor (kg)
Q_F = running total of energy used to heat or cool feed flow (kJ)
Q_R = running total of energy used to heat or cool recycle flow (kJ)
Q_U = running total of energy gained or lost by utility flow (kJ)
Q_X = running total of net energy used or produced by reactor (kJ)

R_{R_i} = current ratio of running total of raw material i used to product (kg/kg)
$R_{R\min_i}$ = minimum ratio of running total of raw material i used to product (kg/kg)
R_Q = current ratio of running total of energy used or produced to product (kJ/kg)
$R_{Q\max}$ = maximum ratio of running total of energy produced to product (kJ/kg)
$R_{Q\min}$ = minimum ratio of running total of energy used to product (kJ/kg)
X_{BP_i} = mass fraction of product i in bottoms flow exiting reactor
X_{DP_i} = mass fraction of product i in distillate flow exiting reactor
X_{FP_i} = mass fraction of product i in feed flow entering reactor
X_{RP_i} = mass fraction of product i in recycle flow entering reactor
X_{FR_i} = mass fraction of raw material i in feed flow entering reactor
X_{RP_i} = mass fraction of raw material i in recycle flow entering reactor
ΔH_F = change in enthalpy of feed flow (kJ)
ΔH_R = change in enthalpy of recycle flow (kJ)
ΔH_{U_i} = change in enthalpy of utility i flow (kJ)
Δt = batch cycle time or continuous run time used to compute running total (sec)

13.3 OPPORTUNITIES

The challenge is to be able to identify the opportunities and estimate the benefits to provide the focus and resources needed and document and report the benefits achieved to provide the recognition and creditability for future endeavors. The most discussed approach in the literature is the reduction in variability. There are also many other types of opportunities and ways of getting at the benefits.

13.3.1 VARIABILITY

The classic approach to analyzing variability is shown in Figure 13.1a where an approximate normal distribution (bell shaped curve) is depicted in a plot of the number of data points versus the value of the data point. A common metric is two standard deviations termed "2-sigma", the plus and minus deviation around the mean value defining the region with 95 percent of the data points. Considerable expertise and methodology has been developed in helping processes minimize this metric. The principle application discussed in the process control literature is on a PV that impacts product quality but the metric can be applied to any data including improving plant practices, meeting market demands and customer requirements.

In process control, the objective is to improve the control systems that have the greatest impact on the data values to tighten the distribution. Sometimes it is forgotten that after the 2-sigma metric is reduced, the setpoint of the associated control loops must be moved closer to the operating limit to realize the possible benefits. Even less recognized is the opportunity shown in Figure 13.1b where a safety margin imposed by operators can be eliminated by better operator training via higher fidelity dynamic simulation, a more comprehensive evaluation of actual deviations gained by the use of data analytics, and the automatic adjustment of the setpoint by an higher loop (e.g., composition control).

Often the improvement comes in terms of reducing energy and increasing capacity by an impurity approaching but not exceeding a limit. In distillation, the higher boiling point being

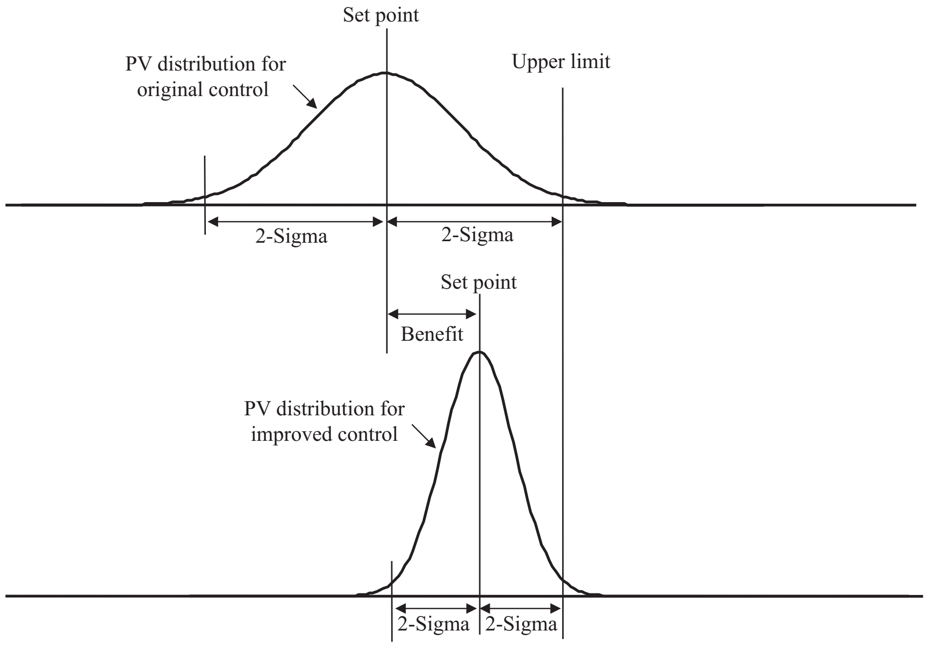


Figure 13.1a. Benefit from reduced variability.

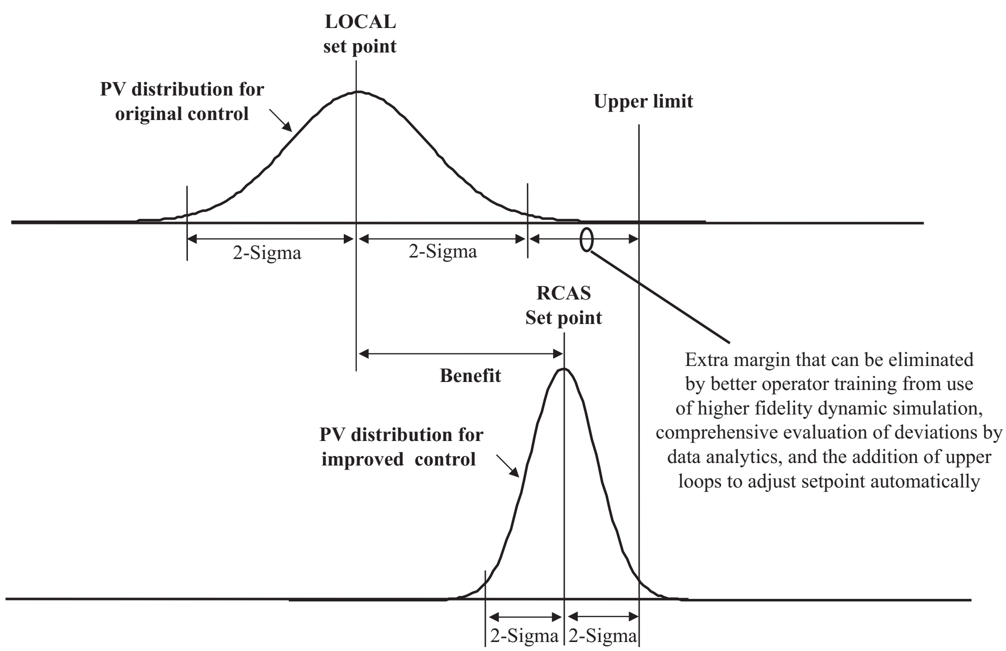


Figure 13.1b. Benefit from models and advanced control.

removed in the column bottoms is maximized in the overhead distillate reducing steam to the reboiler. In the drying of solids or a sheet, moisture is maximized reducing the fuel to the burner(s). In these cases, the maximization of a heavy component or moisture in the product can be taken as a capacity increase or raw material decrease because the product is typically sold on a total weight or volume basis.

Best practices to provide guidance on how to reduce variability are as follows:

1. *Tune level controllers to eliminate oscillations.* The most common source of variability is a level control controller that is creating oscillations in the manipulated flow that ripples through downstream operations. Most level controllers are tuned with too small a reset time resulting in slow rolling oscillations from product of the PID gain and reset being too low as per Equation 1.5c for an integrating process. In an attempt to reduce sudden changes in the manipulated flow, the PID gain is reduced making the problem worse. The Lambda tuning method for integrating processes automatically prevents the violation of the limit for the product of PID gain and reset time. Additionally, the Lambda (arrest time) can be maximized as per Equation 1.22f to maximize the absorption of variability to spread out changes in flow going into a vessel from being transferred to the manipulated flow. The full available volume is used to in dealing with the maximum unbalance between the flows entering and exiting the volume. This technique has enormous benefit to the level control of any volume where a constant residence time is not important and where the manipulated flow is feeding a downstream unit operation. Typical examples are the level controllers on surge tanks, column sumps, and distillate receivers. Note that when distillate receiver level is manipulating a reflux flow back to the column rather than a distillate flow to a downstream unit operation, the distillate receiver level control must be tight to attain internal reflux control and column material balance control.
2. *Use intelligently scaled trend charts to find initiator.* Trend charts that focus on the start of an increase in variability can be used to find the first change in an on-off valve, PID output, setpoint, and PV that started an oscillation. The time scale needs to be coordinated and the variable scale set to show significant changes. In some cases the flipping of the trend by the reversal of the upper and lower scale values can lead to better pattern recognition. The most common manipulated process input is a flow. Fast disturbances are most disruptive. By far the fastest disturbance is a pressure followed by an on-off status or change in the position of a damper, guide vane, or valve or speed of a compressor, fan or pump. Flow measurements can reveal the source of the most disruptive disturbances. For once through continuous operations, the initiator is upstream. For continuous operations with recycle streams, the source can originate from the downstream manipulation of a flow that propagates back upstream. The use of a flow controller somewhere in the path of recycle back to the point of origin can reduce such problems. For example, a flow controller somewhere in the path of a recycled excess reactant going from the reactor discharge to a recovery column and finally back as a reactor feed can reduce variability in the reactor from downstream recovery and recycle unit operations.
3. *Use data analytics software to find correlations.* Drill down to find the largest contributors in a PCA and the relative effect on the PV whose 2 sigma is being minimized via a PLS. Review the results with a process engineer to verify a correlation is an actual cause and effect based on chemical and mechanical engineering first principles. Often overlooked are the contributors of operator shifts and the effect of weather (e.g., temperature,

wind, sun, and rain) on equipment operation and sensor signals. Frequently, the greatest disruption comes at the start of a shift as operators move the process to the point perceived as a *sweet spot* for them. There are many examples of the effects of weather. A *blue northerner* rain storm will cause a sudden increase in internal reflux flow in distillation columns and change in the suction conditions of air compressors. Wind can cause measurement noise from movement in poorly supported load cell installations and unsecured capillary in filled systems for d/p transmitters. Changes in sun and wind can also cause shifts in the sensor outputs.

4. *Use a power spectrum analyzer to find source of an oscillation.* Find the loop with the same frequency as the frequency with the greatest power in the PV whose 2 sigma is being minimized.
5. *Estimate filtering effect of intervening volumes.* To determine whether reducing variability in an upstream PV will have an impact, estimate the attenuation of amplitude using Equation 8.3c where the time constant is the residence time (volume/flow) of the back mixed volumes.
6. *Estimate the blending effect of surge and storage tanks.* To determine whether the process variability from a particular production unit will affect the product that goes to the customer, compute the composition of the blend from a simple material balance on the surge and storage tank. Changes in product from one production unit may be not be seen by the customer when combined with product from many other units in a large tank.
7. *Determine and use the actual shape of distribution plot.* Do not assume a normal distribution (bell-shaped curve). Often the distribution is less known and not as tight on the side closest to the limit. Moving the mean value based on the distribution on the opposite side from the limit can lead to violations of the limit.
8. *Detail the adverse effects of violation of limit.* A greater intelligence as to the final effect on the bottom line and the customer can lead to better focus on what is most important. For example, in sheets or webs, variability in thickness along the edges can be trimmed enabling nearly the entire sheet to be used whereas thickness variability in the middle of the sheet can result in a whole section of sheet being scrapped. The length being scrapped depends upon how long it takes to see and correct the problem. For plastics where clarity is important, the optical density can be improved by minimizing the rate of change of the thickness across the sheet. The consequences are greater when the variability results in a shutdown. Paper sheet lines run at high speeds. The trip of a paper sheet line due to a paper break can result in a huge amount of paper scrap dumped on the floor.
9. *Selectively put controllers in manual to find culprit.* Most short term variability results from control loops reacting to automation system discontinuities, noise, disturbances, interactions, and setpoint changes. Limit cycles from backlash and resolution or threshold sensitivity limits and oscillations from feedback action stop when the PID is put in manual. The controllers that are likely the source of the variability based on intelligent trend charts, PCA and PLS results, and power spectrum analyzers are individually put in manual one at a time until the variability is significantly reduced. The observation of a reduction in variability of an intermediate PV that is shown to be a major contributor to the final PV of interest can shorten the test time.
10. *Determine and set the degree of transfer of variability to the PID output.* Controllers do not magically make variability disappear. Controllers transfer variability from the controlled variable and setpoint to the manipulated variable. Automation system

discontinuities (e.g., backlash and stick-slip) and improper use of PID tuning and options can create more variability in the manipulated variable than what is transferred. Precision valves, tuning to meet objectives and the use of directional move suppression with external reset feedback can go a long way to setting the proper degree of transfer of variability to the PID output.

13.3.2 INCREASING CAPACITY AND EFFICIENCY

The benefits from process control improvement have been stated in many ways. For a better perspective nearly all benefits can be classified as either an increase in capacity or efficiency. Flexibility eventually translates to capacity and efficiency for a given product. An increase in capacity provides more saleable product and can originate from an increase in on-stream time and production rate. An increase in efficiency offers a decrease in the cost of goods from a decrease in raw materials and utilities used. Often there is a choice where an improvement can be taken as a capacity or efficiency increase. For example, a decrease in recoverable excess reactants or recyclable off-spec material can be taken as increase in production rate by keeping the makeup raw material feed flow the same or as a decrease in raw material cost by cutting back on raw material makeup feed flow and a reduction in utility cost of recovery and recycle unit operations for the same production rate. An increase in ethanol yield of corn can be taken as a future production rate increase by shortening batch cycle time or as an immediate efficiency increase by decreasing corn feed rate. Longer batch times once the target endpoint is reached do not increase ethanol concentration due to inhibition effects.

Capacity increases can be gained from increases in production rate by increasing feed rates via valve position control and by reducing batch cycle time by better batch logic and more optimum profile slope control and accurate endpoint prediction and detection as detailed in Chapter 15. Capacity increases can be achieved by increases in on-stream time from reducing startup and transition times, eliminating activations of the SIS, optimal scheduling of downtime for cleaning and defrosting, and minimizing the maintenance interruption frequency and duration.

Increases in efficiency can be achieved by tighter temperature and composition control, more optimum setpoints, minimizing prime mover pressure and maximizing chiller exit temperature via valve position control, eliminating unnecessary crossings of the split range point, optimal scheduling of production rate to take advantage of times when electrical power cost is lowest, and maximizing the use of waste fuels and reagents.

13.3.3 EFFECTIVE USE OF MODELS

Models are increasingly the source of most of the deep knowledge about the production unit and hence the ways to increase capacity and efficiency. Figure 13.2 shows the major classes of models. Stochastic models provide correlations using a data driven approach. Deterministic models provide causes and effects using either a mechanistic (first principle) or experimental (black-box) approach. Figure 13.2 gives examples of stochastic models as being batch or continuous data analytics and mechanistic models as being steady state or dynamic models and experimental models as being step response models (used in MPC) and artificial neural network models (used for inferential measurements). In general, these models provide complementary information. An understanding of the relative capability can help apply the best model for the task.

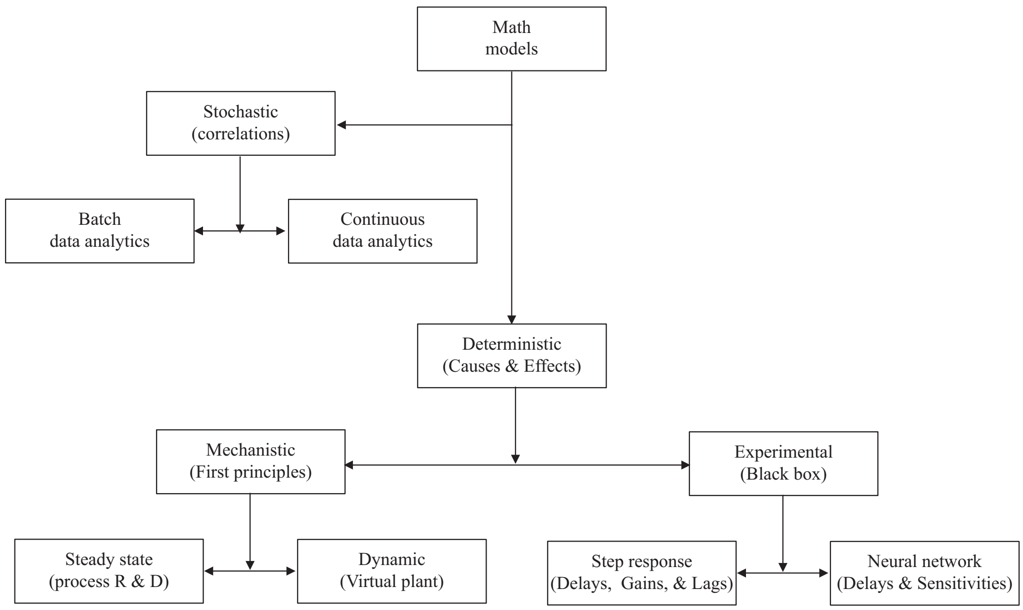


Figure 13.2. Process control improvement models.

The appeal of stochastic models stems from the ability to include a large number and wide variety of process inputs and be purely driven by data. Preconceptions that are prevalent in the mechanistic models do not enter into the results of data analytics. The number of inputs and hidden correlations can make the complexity overwhelming for analysis by other means. Data analytics is able to reduce a large number of inputs with cross correlations to a small set of orthogonal independent inputs called principal components by PCA. The principal components are stated to be *orthogonal* because a multidimensional plot of the process output being modeled has principal components axes at right angles to each other. The user is able uncover previously unrecognized relationships by drilling down into the contributions to each principal component. The correlations between inputs and the output are not necessarily cause and effect and must be reviewed by the automation engineer and process engineer. Some of the correlations may be due to measurement or valve problems. The results enable the user to focus on potential problem areas that were lost in the maze and overload of data dealing with the dilemma "Drowning in Data, Starving for Information."

Unfortunately, you cannot just dump all your data into a data analytics program and expect wonderful results. The use of data analytics requires some screening of extraneous batches or operating periods and time synchronization and addressing of nonlinearities for continuous operations.

In batch data analytics, there is no time synchronization issues because the relationship sought is focused on the result at a single point time that is the batch endpoint. Also, the nonlinearity of the batch profile for a composition is addressed by a piece wise linear fit attained by time slicing of the data.

For batch data analytics, batches are chosen to provide a good reference of an average batch rather than a best batch. This practice is representative of the first step in optimization that is to attain repeatability. Once the variability in batches has been reduced, the second step in optimization can begin, that is, to move to operating conditions with better metrics. After data analytics enables repeatable batches, the search for better batches can begin (e.g., higher endpoint or shorter cycle time). The identification of the largest contributors in the principle components for the best batches offers the exploration of potential opportunities.

For continuous data analytics, changes in process inputs must be coordinated in time with changes in the downstream process output being predicted by the PLS model. How long a change in process input shows up as a change in process output is generally large and variable for continuous liquid volumes. This time synchronization is not trivial particularly if there are a significant number and different types of unit operations between the process input and the PLS model output. The best guidance to date is to use time constants that is the residence time of back mixed volumes and dead times that are the transportation delays between the inputs and PLS output. Additional difficulties arise from operating point nonlinearities. There is no corresponding piece-wise linear fit that addressed nonlinearities of batch profiles.

Mechanistic models are only as good as the software and user expertise. Steady state models are principally used for process R&D and for detailed process design including the process flow diagram (PFD) for continuous unit operations. Steady state models offer the greatest information on process cause and effect for steady state relationships but do not include integrating or runaway process gains or dynamics, automation system dynamics, process control strategies, PID features and tuning, or batch unit operations. Steady state models can find the self-regulating process gain (steady state process gain) and the advantages of measurement locations and paring of controlled and manipulated variables in terms of providing the greatest process gain often stated as process sensitivity. For example, steady state models can show which tray in a distillation column should be used for temperature control by noting the tray that provides the largest temperature change in both directions for a change in reflux to feed ratio and steam to feed ratio. The pairing of temperature as a controlled variable with reflux to feed or steam to feed ratio can be done based on the relative sizes of the corresponding process gains.

Note that steady state models provide the process gains and not the open loop gains that depend on the automation system. Also, the process sensitivity commonly stated (change in a PV for a change in a manipulated variable or operating condition) is really a process gain. Process sensitivity is not to be confused with the threshold sensitivity of sensors and control valves that is the smallest change in input to that will cause response in the output of the automation system component (e.g., valve stick-slip).

In order to get the open loop gain and the dynamics of the process, a dynamic model is needed with valve and pump sizes and installed flow characteristics and PID controllers with correct scale ranges. To get the automation system dynamics right, blocks are inserted to simulate PID execution time and signal filter time, sensor and transmitter time constants, dead times and threshold sensitivity, and resolution limits, discontinuous update time intervals by wireless devices and at-line analyzers, and finally the final control element (control valve or variable frequency drive) with one or more time constants with rate limiting (e.g., for valves stroke is a velocity limited exponential based on slewing rate), deadband, resolution limit, and threshold sensitivity limit. The use of precise and fast valves (e.g., sliding stem valves with digital valve controllers and if necessary volume boosters) can eliminate the need to simulate the dynamic behavior of valves except for applications that are extremely sensitive to valve precision and speed (e.g., pH and surge control). To get at the actual PID response, the extensive features and tuning of the PID must be included. The best way to get all these details is by way of a virtual plant where the actual DCS configuration is downloaded and with blocks added to simulate the field automation system dynamics. PID features in the DCS such as the positive feedback implementation of integral mode with external reset feedback are critical to developing and testing PCI ideas. Trying to emulate a PID is difficult at best because most PID algorithms are proprietary and the PID algorithms depicted in the control theory text books don't have the PID forms, structures, or options used in a DCS. A virtual plant where the actual DCS configuration is downloaded provides the best assurance that the correct PID features and tuning are included.

To summarize, steady state models can find more optimum operating points and the resulting gaps in the key performance indicator for steady states in continuous processes. A comprehensive DOE can be conducted and what if scenarios explored. Dynamic models deal with processes without a steady state and when used in a virtual plant can determine whether a control strategy can consistently achieve the more optimum setpoints. High fidelity dynamic models can also take the place of steady state models in finding better steady states. The future is high fidelity virtual plants.

It is difficult for mechanistic models to include all the process parameters accurately such as heat transfer and mass transfer coefficients and the threshold sensitivity and parameters for a first order plus dead time approximation of sensor and valve dynamics. Experimental models can provide the missing details needed to quantify the dynamic response of the process and automation system.

Step response models provide the most accurate model of the dynamic response for a particular set of operating conditions and step size and direction. However these models assume a linear response and are thus vulnerable to changes in setpoint and load. The model also depends upon step size and direction as noted in Chapter 9 on nonlinearities. The operating point nonlinearity from the installed valve characteristic can be eliminated by the use of flow rather than valve position as the manipulated variable via a secondary flow loop. Signal characterization can also be used to provide new controlled variables that are much more linear. The classic example here is the use of signal characterization to convert a controlled variable from pH to per cent reagent demand by doing a piece wise linear fit that computes the X axis from the Y axis of the titration curve.

The forte of artificial neural networks (ANN) models is the inclusion of nonlinearities. ANN models have been particularly successful for predicting the end points of batch operations. In batch unit operations there is no time synchronization requirement for the ANN prediction of a batch end point for same reason as for a PLS prediction. ANN models have also been useful in the prediction of composition at the outlet of continuous unit operation where just a transportation delay exists between the process inputs and predicted process output (e.g., dryer moisture, pulp stock consistency, extruder product).

For continuous unit operations, a fixed dead time is inserted for each input that corresponds to the 63 percent response time of the process output to that input. The result is some time mismatch when the dynamic response changes with load (normally the case) and when there is a large time constant between the process inputs back mixed liquid volumes. Also the data used to train and test the ANN models must cover the complete operating range and changes must be made to the process inputs to capture the nonlinearities. Unlike PLS models that would use linear extrapolation for predictions outside of the data range, the predictions of ANN models outside the data range can get quite bizarre including reversals of slope sign (process action). A DOE in the actual plant is the best way to ensure adequate richness of data for the development of an ANN. Additionally, a plug flow process with ANN with dead times on the process input set as a function of the production rate makes the time coordination sufficiently accurate.

The use of step response models to find the missing dynamics in virtual plants holds the greatest promise. Just finding the complete dynamics of one operating point to trim the dynamic parameters for mixing and automation system components may be sufficient since the major process and automation nonlinearities are addressed based on first principles. If this is not the case, MPC can be used in an innovative way to automatically adapt these parameters online to match manipulated flow ratios. The MPC setpoint (target) is the actual plant flow ratio, the MPC controlled variable is the virtual plant flow ratio, and the MPC manipulated variable is the process parameter. This setup has been demonstrated to effectively adapt distillation column efficiencies to match the manipulated reflux to feed flow ratios in the virtual plant to the actual plant and to adapt a pH neutralization system influent acid pKa to match the base reagent to influent flow ratio in the virtual plant to the actual plant.

A virtual plant offers many advantages including the conduction of DOE that can be used to find and implement solutions for interactions, sources of variability, and constraints for capacity and efficiency. Process control improvements can be prototyped and benefits quantified. The solution can be confirmed by a focused DOE in the actual plant.

13.3.4 SIZING AND ASSESSMENT

An opportunity sizing and assessment described here has been used by a major monomer, polymer, specialty chemicals, and agricultural chemicals manufacturer over a period of five years to attain an average 2 percent reduction in the COGS. Increases in capacity were taken as equivalent benefits in terms of COGS to provide one overall metric. Figure 13.3 shows the overall procedure and documentation in the PCI process using key performance indicators (KPI). The PCI process progresses from a sizing sheet to an opportunity summary, work plan, "before" and "after" results recognition, and finally benefits achieved reporting.

The opportunity sizing process outlined in Figure 13.4 is a critical first step where the potential financial gains for the KPI of the candidates for process control improvement are itemized. Traditionally this has been done by an analysis of the effect and source of variability. A more comprehensive approach takes advantage of production unit cost sheets and models to find and confirm more optimum operating points.

The best week or month of operation from cost sheets is used to find best KPI after extensive discussion with process engineers and operations to understand what data is relevant or extraneous and whether the best KPI is real. Many plants are reluctant to open their cost sheets to outsiders. The human factor that is at play in all aspects of a successful sizing and assessment is greatest here. The person leading the analysis must be viewed as part of the plant team with a focus on future achievement rather than past. Management can be protective of previous decisions and efforts. Egos need to take a back seat. Fortunately, engineers tend to be objective and receptive to changes when presented with the data and the logic.

Online metrics and models can open minds to possibilities not seen in the cost sheets. Ideally models can be used to conduct a comprehensive DOE leading to a focused DOE in the actual plant. The results and conclusions must also be reviewed for practicality before being placed as an achievable KPI in an opportunity sizing. The gaps between the best KPI possible and the average KPI experienced is quantified preferably with dollars assigned as the benefit of the improvement in capacity or efficiency. Business people may be reluctant to confirm how much more capacity is needed or the expected product price. Contrary to engineering practices,

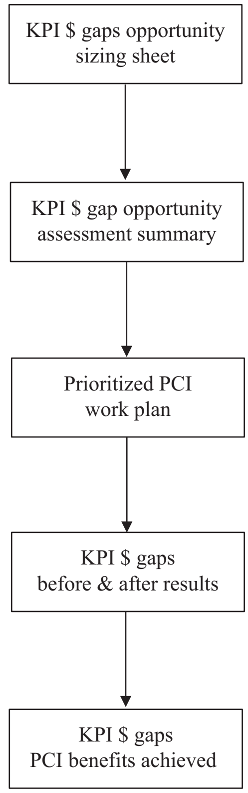


Figure 13.3. PCI documentation using KPI.

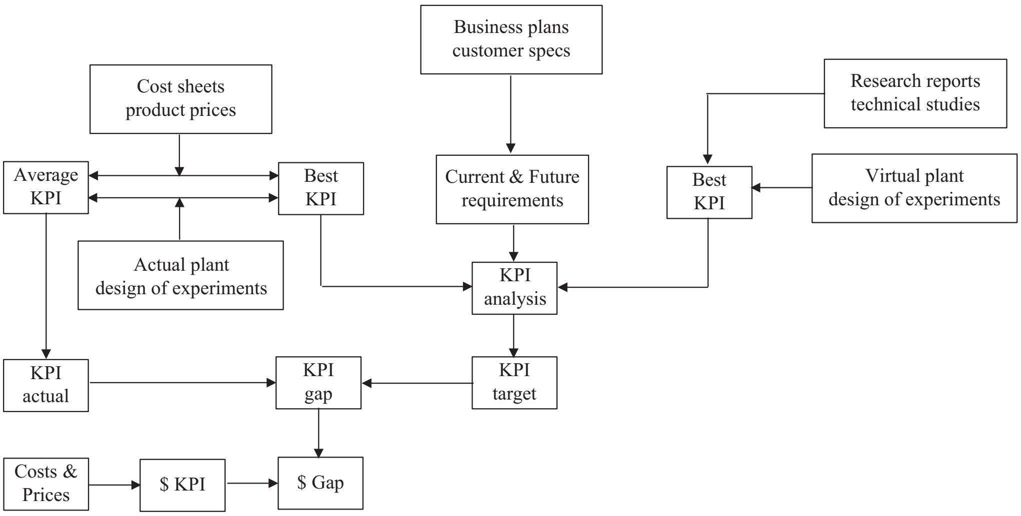


Figure 13.4. Opportunity sizing process.

exact numbers are not needed. Relative numbers from operations can determine the importance of the opportunity.

The next step of the PCI process is to have a two day meeting to do the assessment and the work plan. Key people in operations, R&D, process engineering, process control, maintenance, analyzer support and instrument design, and maintenance should be there in a round table setup. Business people and upper management may attend the final hour on the last day to see a summary of the assessment and work plan.

The meeting starts with a process engineer providing a functional understanding of the production unit including challenges and possibilities by going through PFDs describing the process relationships and simplified control diagrams to show the control strategies. An agreed upon sizing sheet with gaps quantified is presented by the person who generated the sizing. This same person then leads the assessment. The presentation of solutions before the sizing and PFD review is extremely counterproductive. External people, no matter how proficient, coming into a plant and giving a solution may seem to be right but will often miss the point and turn off the audience. Similarly, jumping to conclusions must be avoided. Questions are important to make the meeting a discussion rather than a presentation. The questions can explore the conditions of a possible opportunity without stating the conclusion until after the process engineers and other plant people fully cover the situations in the production unit.

At the end of the first day and continuing into the second day solutions are offered and the relevance is reviewed without addressing how to do the implementation. The process engineer makes an estimate as to what approximate percentage of the gap can be reduced by the solution. The process control engineer makes an extremely rough estimate as to the time and capital required. Solutions that require just a configuration change, better tuning or take advantage of existing tools and software are termed *quick hits*. Solutions that require the installation of instruments that are in stock can become *quick hits* if process connections and the main cable runs exist. In the future, wireless instruments will make the more extensive use of smart instrumentation a *quick hit*.

Based on the value of the KPI gap potentially eliminated and the time and cost requirement, a work plan is developed. The *quick hits* should be started immediately after the meeting to taken advantage of the enthusiasm, knowledge, and synergy generated in the meeting. Often the *quick hits* can be completed in one to two weeks. If specialists go home and plant people go back to their routines immediately after the meeting, much of the potential opportunity may never be realized.

The benefits must be measured and reported. This is the most important step for enabling a continuation of the PCI process. More resources may be subsequently made available for the improvements that are not *quick hits*. The very existence of a process control effort or group today hinges on plant management seeing a monetary advantage. The justification to do something better than a copy job on the next migration project may also be at stake. "Money talks" and can overcome the lack of technical understanding.

The person who does the opportunity sizing and leads the opportunity assessment needs to determine the before and after case. The installation of online metrics makes the quantification more exact. Wireless instrumentation can make this more of a possibility. The before and after cases should be quantified immediately before and after the PCI completion to avoid the complication of changes in operating conditions and equipment maintenance. Dynamic simulations can fill in the blanks and explore refinements. Ideally, the PCI system should be turned on and off for different plant states but in practice, operations will be reluctant to turn off a system that is producing benefits. The transition times between the on and off states may be significant causing some confusion and loss of capacity or efficiency in the transition. Here again the use of high fidelity virtual plants can quickly and effectively show the differences between the on and off states for various production scenarios to strengthen the recognition of PCI and models.

13.4 KEY QUESTIONS

The following updated questions and answers from *Advanced Control Unleashed* serve as examples of the thought process and open discussion that promotes creativity in the search for benefits from process control improvement.

1. *Increase benefits in all areas*
   a. Do you have loops not running in the highest design mode (e.g., auto, cascade, supervisory) because of oscillations, poor control, upsets? Look for improvement by the use of signal characterization, auto tuners, gain scheduling, adaptive control, fast sensitive and repeatable measurements, better loop designs, and valves with negligible backlash and stick-slip.
   b. Do you have upsets (bumps) to critical process parameters from process interactions? Seek the best pairing of controlled and manipulated variables by relative gain analysis. If the interaction involves just two loops consider a flow feedforward as a half decoupler. For more complex interactions, consider MPC to decouple and anticipate upsets.
   c. Do you have upsets from cycling back and forth across a split range point? Look for ways to eliminate split range or reduce crossings by the use of directional move suppression and external reset feedback.
   d. Do you have a tool to define correlations and interactions in systems? Install a loop performance monitoring system and data analytics software.
   e. Do you have bad measurements that you cannot rely on? Improve measurements by using advanced technologies. Calculate batch averages and alerts based on deviations from averages. (e.g., Do not rely on pH for endpoint detection if curve is different from previous good batches.) Use data analytics for a more complete and accurate monitoring of deviations and the effect on batch endpoints.
   f. Do you have areas you cannot control because they are hard to measure? Calculate online property estimators (e.g., inferential measurements of composition or quality) from common field measurements (e.g., density) corrected by lab measurements and design new control schemes.
   g. For composition, pH, and temperature control, do you have ratios installed on related flow loops instead of flows set individually? Flow feedforward with a viewable and adjustable ratio can help maintain energy and material balances and can address the fundamental relationship where the controlled variable is a function of the manipulated flow to feed flow ratio.
   h. On setpoint changes, does it take a long time for the variable to really start to move and once moving does it go into an overshoot (or undershoot)? Use a full throttle setpoint response by putting the valve fully open or closed in the beginning and release to feedback control when coming close to the desired setpoint.

   i. Are you interested in optimizing the rate of change of a PV (e.g., increase in reaction temperature and hence reaction rate). If so, develop a new controlled variable that is the rate of change of the PV by using a dead time block to provide fast updates with the best signal noise ratio (same technique used for batch profile slope control). The new loop can be used as an override controller where a signal selector chooses the safest controller output.

2. *Improve yields*
   a. In the areas where byproducts are made, can you look for other sequences, other parameters (temperatures, hold times, agitation, etc.)? Compare plant to the lab and pilot plant operation and look for differences. Redesign plant system to match lab and pilot plant conditions as close as possible.
   b. Would more accurate endpoint detection improve yield or quality? Consider rate-of-change indication of one or more measurements as an endpoint indicator after a designated batch time or feed total.
   c. Do you have sequential operations in your batch reactors? Consider parallel feed additions and ratio control to generate fewer byproducts, improve quality and yield, or reduce cycle time.
   d. Do you run your recovery systems just high enough to recover products at the strengths you need to reuse them? Lower purity setpoint can reduce the energy and waste treatment costs.
   e. Do you have manually set purge or vent flows? Look for ways to control these flows and ratio them to product flows. Calculate and control minimum required purge or vent rates.
3. *Reduce energy and utilities*
   a. Are steam, utilities, waste waters, and big motors switched off on temporary shutdowns? Implement systems that will shut off these systems automatically on temporary shutdowns.
   b. Do you run dryer temperatures just high enough with good controls or do you run always well above critical wet bulb temperature? Install better controls (feedforward from feed) and adjust set points to run closer to moisture limit for lower energy usage.
   c. Could you reduce energy sources (e.g., refrigeration unit temperature, compressor pressure, boiler pressure) until user valves reach maximum controllable position? Consider a valve position controller to optimize use.
   d. Can you downsize over-designed pumps and motors? Can you eliminate flow valves with variable frequency drives for motors?
4. *Reduce effluent*
   a. Can you reuse waste streams in your process? Hot water and condensate can be reused for heating purposes or for reactor concentration control. Waste streams can be used as fuel or reagents.
   b. When byproducts are created can you find the root cause (e.g., poor agitation, hot or cold spots on heat transfer surfaces, bad mixing) and adjust controls to reduce byproducts? Improve agitation patterns and use tempered water with constant coil or jacket recirculation flow.
5. *Reduce maintenance cost*
   a. Do you have maintenance costs due to poor operation exceeding equipment design conditions (e.g., cavitation, high temperature or pressure, overload, etc.)? Automate startup procedures and best operating practices for equipment.

   b. Do you blow rupture discs and activate relief valves due to manual disoperation or bad control systems? Use better control and automation to reduce safety risk and possible down time.
   c. Could smoother control (reduced thermal and pressure shocks) or tighter control (less byproduct and contamination) increase time between repairing, replacing or cleaning equipment? Use less aggressive tuning and directional move suppression with external reset feedback.
6. *Reduce lab cost*
   a. Can you install in-line analyzers and control to reduce the number of lab analyses? Reduce batch cycle time by not having to wait for lab results.
   b. Can you reduce the number of lab analyses by improved endpoint detections and automation resulting in more consistency? Analyze lab results using statistical process control and adjust frequency of analysis.
7. *Improve product quality*
   a. Can you measure or calculate constraints from existing measurements? Once measurable, these constraints are controllable.
   b. How close are you running to constraints (critical reaction temperatures, pressures, feed rates, etc.)? Use advanced regulatory control to run closer to or *ride* constraints. Use override controls, feedforward, dead-time compensators, valve constraint controllers, rate-of-change control, and MPC.
   c. Do you automatically capture product parameter settings when quality is good? Use last run data on startups, transitions, and batches to compare and find ideal settings for each product.
8. *Reduce rework*
   a. Do you produce rework because the operator sometimes makes human mistakes? Look to automate all manual actions.
   b. Are all your operating instructions clear, up-to-date and easily accessible? Use unique process parameter descriptions and units to avoid confusion.
9. *Reduce shutdowns and upsets*
   a. Do you have too many or too few alarms? Is the operator alerted in advance and is he not confused by too many alarms? Make alarms smart (e.g., don't want Low Flow alarm when pump is stopped). Use advanced alarm handling and logging.
   b. Do the operations and process support have easy access to information when something goes wrong? Can failures be diagnosed? Provide help displays with dynamic information and instructions on what to do to correct. Use smart instrument diagnostics and data analytics to identify a change from normal operation and the contributing factors
   c. Does automation system define corrective actions to be performed by software when something goes wrong? Do not just stop and give control to the operator.
   d. Do you have a SIS shutting you down? Is it easy to track what exactly happened? Use a reliable configurable automation system for the SIS and create dynamic SIS help displays for all abnormal conditions.
   e. Do you have surge tanks in between batch and continuous processes? Smart level control that maximizes the absorption of flow variability can be used to prevent shutdowns in continuous process due to flow upsets.

   f. Do you have a system set up to track every problem and fault that generates upsets, rework, or shutdowns? Learn from mistakes and set actions to prevent their happening again. Use problem reports.
   g. Can product changeover (transition) time be reduced? Automate changeover operations (emptying, cleaning, and washing). Look for methods to reduce transition times and set best operating parameters for every sequence and product. Use intelligent output actions and PID structure to minimize rise time and overshoot in setpoint response.

10. *Reduce cycle time*
    a. Can you reduce batch cycle times by eliminating unnecessary waits and introducing parallel sequences? For example, start heating before loading, and combine hold periods with cooling.
    b. Can you eliminate concurrent actions that slow down your process? For example blowing cold air in a tank for mixing purposes while heating.
    c. Can you speed up heating times and cooling times by using energy more intelligently? For example, use colder water and better agitation.
    d. Can you reduce feed prep and charge time? For example, prepare premixes in separate tanks in parallel with other sequences and use bigger pumps and valves.
    e. Is the operator informed and warned when a system is waiting for a variable to reach a condition, if the variable does not move toward that condition (anymore)? If not the system would wait forever and the operator will find out too late. Configure multiple dynamic waits.
    f. Could you detect endpoints sooner? Consider rate of change detection and the use of future PV. In some cases a high rate of change is indicative of an endpoint (e.g., conductivity for chlorination). In other cases a future PV can be computed for target value (e.g., future pH for neutralization).
    g. Can you replace weighing systems with mass flow meters? Mass flow can also be calculated for raw materials from flow, temperature, and pressure measurements. Coriolis mass flow meters are more accurate than load cells and can reduce batch load times and cycle times dramatically.
    h. Can you redesign process parameters and SIS settings based on experience from the past? Adapt settings to process evolution.

KEY POINTS

1. Online metrics and models offer the best means of quantifying benefits.
2. Understanding the *Human Factor* is critical to opening up minds.
3. Solutions should not be presented until the opportunity sizing and a description of the process, challenges, and previous PCI endeavors have been discussed.
4. Models can play a big role in finding and understanding PCI opportunities and developing and testing the more effective use of PID features and strategies for basic and advance regulatory control. A high fidelity virtual plant with dynamics improved by step response models offer the most promise.
5. The quantification of the performance of the before and after cases is essential to providing the justification for additional PCI efforts and even the existence of the process control group. Models can help fill in the blanks.

CHAPTER 14

Auto Tuners and Adaptive Control

14.1 INTRODUCTION

Automating any process can yield big improvements by eliminating human error and adding repeatability and predictability. The benefits are greatest when the best technology and practices are automated, the novice is protected against mistakes, and the specialist is enabled to capitalize on creativity and expertise. The same is true for proportional-integral-derivative (PID) tuning.

14.1.1 PERSPECTIVE

The tool must also minimize the disruption to the process. However, process dynamics and tuning settings cannot be identified unless a known change is introduced into the process. I learned early on from Bob Otto (Monsanto Senior Fellow) that software cannot sort out the effect of PID response and the unknown external changes in closed loop operation despite claims to the contrary. Thus, auto tuners and adaptive control tuners are triggered by a change in setpoint or output. The change in output is normally a step made in manual or remote output mode. This simulates a step load disturbance.

The tuning is quite different for a setpoint and load disturbance unless a two degrees of freedom (2DOF) structure or a setpoint lead-lag is used to prevent overshoot. Therefore it is wise to check the load disturbance performance by putting the PID in manual, making a change in the PID output, and then immediately returning to the automatic mode. For fast integrating processes (e.g., phosphorous furnace pressure) or runaway process (e.g., polymerization reactor temperature), the controller can only be in manual for a dead time. If the loop requires a variable speed drive or hydraulic actuator instead of an air actuated control valve to be fast enough, the loop cannot go in manual at all. Loops for equipment, environmental, and personnel protection and safety (e.g., surge and over pressure prevention) should also stay closed.

A pulse can be automatically injected into a PID output for loops that must stay closed. The software required to identify the dynamics from a pulse is more vulnerable to noise and unknown disturbances. Consequently, more pulses than step changes are required and the time to achieve good model quality may be extended.

Pattern recognition techniques using heuristic rules can be used to determine when relative contributions of the proportional, integral, and derivative modes are out of balance. This can be a useful back up but care must be taken that the rules are not fooled by noise, periodic disturbances, limit cycles, and the windows of allowable gain and reset time for integrating and runaway processes. Normally, the safer thing to do is to allow for an automatic increase in reset time but require review and approval for an automatic decrease in reset time. An increase in reset time will always reduce loop oscillations and overshoot. The detrimental effect of a larger than necessary reset time are a slow return to setpoint for a load upset and a slower rise time for a setpoint change. Most of the time the reset time is too small. A notable exception is a highly dead time dominant self-regulating process where the minimum reset time is an order of magnitude smaller. Lambda tuning is suitable if a low limit on the reset time of four tenths the dead time is imposed.

If the auto tuner and adaptive control tuner identify the open loop dynamics and allows the user to choose the tuning rules, the user is empowered. The knowledge of the open loop dynamics enables a better understanding and monitoring of automation and process dynamics. For example an increase in a secondary time constant or dead time would be indicative of slower heat transfer (e.g., surface fouling and frosting) or slower measurement (e.g., electrode aging or coating). The flexibility in tuning rules enables the expertise of the specialist to be effectively used. For example, a transition from Lambda tuning for self-regulating processes to Lambda tuning for integrating processes when the time constant to dead time ratio exceeds four, indicating a near-integrating process, provides a better load disturbance response. Also, the setting of Lambda (closed loop time constant for self-regulating processes and arrest time for integrating processes) relative to the loop dead time, enables the user to optimize the tradeoff between maximizing disturbance rejection, variability absorption, loop decoupling, and loop coordination.

A faster test time and the use of integrating process tuning rules are possible if the tuner allows lag dominant self-regulating and runaway processes to be evaluated as near-integrating processes. The open loop integrating process gain can normally be identified within about five dead times. If the software is capable of identifying the secondary time constant, the test identification time should be extended by at least two secondary time constants.

Plants have standardized on tuning rules for maximum load disturbance rejection (e.g., Shinskey), setpoint response (e.g., Internal Model Control or Simplified Internal Model Control), or some specific process objectives (e.g., Lambda). The flexibility to use alternate tuning rules honors user preference and expertise.

The ability to simulate within seconds the response of tuning settings and identified dynamics for a load disturbance and setpoint response, affords rapid experimentation. The simulation enables the user to evaluate performance metrics such as peak error and integrated error for load disturbances and rise time and overshoot for setpoint response.

Regardless of tuning method, the step size must be chosen to provide a PID output and input signal change larger than the deadband, resolution limit, and threshold sensitivity limit in the final control element (e.g., control valve or variable speed drive) and measurement (e.g., sensor and transmitter), respectively. While this step size is necessary to see the effect of the change on the process variable (PV), the increase in the closed loop dead time is not identified for steps in the PID output or setpoint when the immediate change in the PID output is larger than the deadband or resolution/sensitivity limit. Pattern recognition or a ramp rather than a step is needed to find this additional dead time. The ramp rate of PID output needs to be about the same as experienced in the plant for a typical disturbance. Since this depends upon the tuning, there is a chicken and an egg scenario. The tuning is computed based on step tests and then revaluated based on closed loop responses and by use of a ramp instead of a step for open loop responses.

The step size must also be larger than process and measurement noise. A step size that is as large as possible without excessively upsetting the process shortens test time and makes sure the response is not washed out by the filtering action of well mixed volumes or confused by changes in feed streams and the action of other loops. Trend charts can show typical changes in PID output employed to deal with disturbances as a guide. In general, the higher the PID gain the larger the output change needed to prevent test time out and response wash out. For example, a level loop or batch temperature loop may have a PID gain of 10 or more. In this case the step size is much larger. The PV change may not be barely noticeable unless an output change of 20 percent or more is used.

The use of signal characterization to account for operating point nonlinearities can significantly reduce variation in the process gain identified with step size and direction.

The PID should be tuned with a judiciously minimized signal filter. The signal filter will add dead time or create a secondary time constant. The signal before and after filtering should be collected in a data historian (historized). The PID should be tuned with the analog output (AO) block and lower loop PID setpoint rate limits set if used for directional move suppression and the PID option for external reset feedback (e.g., dynamic reset limit) is enabled.

14.1.2 OVERVIEW

The relay method of auto tuning can provide a quick estimate of the ultimate period and ultimate gain and consequently the PID tuning settings in about three oscillations. The ultimate gain can provide an estimate of the gain margin if taken as the maximum gain for stability for a proportional-only controller. The ultimate period can be extensively used for the analysis of tuning settings, loop interaction, cascade dynamics, dead time versus lag dominance and disturbance dynamics. The ultimate gain and ultimate period can also be used to get initial tuning settings via the Ziegler–Nichols ultimate oscillation method that is modified to provide more resilience to changes in process dynamics and to make the closed loop response less oscillatory. The modification consists of simply dividing the ultimate gain by the desired gain margin. Since most adaptive control tuners identify process models from a closed loop response benefit from a reasonable estimate of tuning settings and a non-oscillatory response, a good practice is to use the relay tuner with a gain margin of four or more to get the initial tuning settings in the ball park.

For dead time dominant loops, the factor applied to the ultimate period from the relay method to get the reset time could be reduced to provide a faster closed loop response test. The factor applied to the ultimate period can be reduced from 1 toward 0.25 as the ultimate period decreases from four to two times the dead time as detailed in Equation 1.17c to provide more aggressive integral action as the degree of dead time dominance increases.

The user is often over concerned about oscillations not realizing the attenuating effect of large process time constants or the slowing effect of very small integrating process gains. When an operator is asked point blank what step size is permissible, the step size cited is typically too small. A joint decision by operations and process engineering after a review of the historical data provide much more useful step sizes. Note what process excursions are acceptable and the associated changes in the PID output. Realize that statements made before analysis are difficult to reverse due to ego, a conservative mindset, and a lack of understanding of how dynamics attenuate and camouflage changes in process inputs.

Find from tuning tests at various setpoints, production rates, and batch or run times whether tuning should be scheduled based on the PV or time for operating point nonlinearities or the PID output for final control element and process nonlinearities. Consider the use of signal characterization for operating point nonlinearities and nonlinear installed characteristics to free adaptive control to concentrate on unknowns and other nonlinearities. For pH, the input signal characterization involves the translation from the ordinate to the abscissa of the titration curve so that the PV of the PID is percent reagent demand instead of pH. For final control elements, the output signal characterization translates the ordinate of the installed flow characteristic to the abscissa so that the manipulated variable is no longer signal but percent flow demand. A percent reagent or flow demand rather than engineering units is preferential. The actual pH and valve signal must be displayed and historized for operations and maintenance to enable diagnostics and analysis of the automation system and process performance.

A goal of less than 10 percent error in tuning settings from data compression and data update is advisable. Wireless transmitter default update rate and trigger level are set accordingly. Similarly, the PID module execution rate and transmitter damping and PID signal filter effect on the PID tuning settings is targeted to be less than 10 percent (Chapters 5 and 6).

14.1.3 RECOMMENDATIONS

1. Ensure data compression is less than the smallest of the final control element and measurement resolution and sensitivity limits and the compression divided by the maximum rate of change of the signal is less than 10 percent of the total loop dead time or reset time setting whichever is smallest. Basically this amounts to turning compression off for most loops.
2. Ensure data update time is less than 10 percent of the total loop dead time or reset time setting whichever is smaller.
3. Set the module execution time to be less than 10 percent of the total loop dead time or reset time whichever is smaller.
4. Set the transmitter damping or filter time setting just large enough to prevent the valve or variable speed drive speed from changing due to measurement noise. The transmitter damping and filter time setting should be less than 5 percent of the total loop dead time or reset time whichever is smaller.
5. Review historical data with operators and process engineers and decide on largest possible change in PID output that will not cause an unacceptable variability.
6. Choose a PID structure of PI on error and D on PV if setpoint response is not important and all disturbances are a process input rather than a process output. Take into account the significant filtering effect of intervening process volumes.
7. If both the load and setpoint response are important, use 2DOF structure (e.g., beta = 0.5 and gamma = 0.25) or a setpoint lead-lag (e.g., lag time = reset time and lead time = one-fourth lag time).
8. Specify the process to be integrating for lag dominant self-regulating and runaway processes and choose test time to be at least four dead times plus two secondary time constants to save on test time and to enable integrator tuning rules.

9. Review the supplier documentation and presentations and attend short courses on the use of auto tuner and adaptive control software.
10. Use an auto tuner or open loop tests to get ball park tuning settings. For near and true integrating processes, you just need to identify the dead time and ramp rate.
11. For a better estimate of the primary time constant, use Equation 4.5a with a visual identification of the dead time and period from the relay auto tuner trend chart.
12. For temperature control of large liquid volumes (e.g., reactor and distillation columns), identify the secondary time constant associated with heat transfer surfaces or the thermowell via the use of special software. The identification package for some auto tuners and most model predictive control offer this ability. Consider the process dynamics to be near or true integrating so that only the initial response needs to be identified. Make sure the PID rate time is equal to the secondary time constant or half the total loop dead time, whichever is largest. For processes with positive feedback (e.g., highly exothermic reactors), the identification may need to be done closed loop so that the temperature control can stay in automatic to prevent a runaway reaction.
13. Do tests at several different setpoints and production rates to check performance and determine operating point (PID input) and feed (PID output) nonlinearities.
14. Use signal characterization on the most severe and relatively fixed nonlinearities to free up adaptive control to concentrate on unknown and variable nonlinearities and to make the results much less dependent upon step size and direction.
15. Use adaptive control to identify and correct for changes in process dynamics. For PID controllers with a severe process nonlinearity (e.g., pH control), schedule the tuning settings based on the PV. Otherwise, schedule the tuning settings based on PID output to take care of production rate nonlinearities (e.g., increase in process gain and dead time as production rate decreases) and the valve gain nonlinearity due to the change in the slope of the installed flow characteristic (e.g., flattening of characteristic at low and high valve positions).
16. Use simulation (e.g., Simulate screen in Auto Tuner) to determine the effect of changes in tuning settings on load response and setpoint response. Note that the simulation may not include the effect of setpoint lead-lag. Decrease gain margin by increasing PID gain to decrease integrated absolute error (IAE) and rise time. Increase phase margin by increasing PID reset time to reduce overshoot.
17. Evaluate load disturbance response by making a step change in the PID output with the PID momentarily in manual if PID loop is allowed to go into manual. Do this is both directions. To reduce peak error, increase the PID gain.
18. Evaluate setpoint response by making a step change in the PID setpoint. Do this in both directions. To reduce overshoot, increase reset time or decrease beta in a 2DOF PID structure or increase the lag time in setpoint lead-lag.
19. Periodically run tuning tests to look for deterioration in process and sensor dynamics with time (e.g., fouling of heat transfer surfaces and electrodes or loss of catalyst activity with continuous run time or batch cycle time). Schedule tuning settings and maintenance procedures based on these changes in dynamics.
20. Use pattern recognition and heuristic rules to evaluate the balance between proportional, integral, and derivative mode contributions to the output. An increase in reset time is stabilizing and thus permissibly done automatically as detailed in Section 5.3 on adaptive reset.

14.2 METHODOLOGY

The material presented here is intentionally sparse because so much depends upon the actual tool used. Guidance should be sought from the supplier of the software. Onsite assistance by the supplier for critical loops is advisable particularly until the user has gained extensive application experience in the use of the tool.

The relay oscillation method developed by Karl Astrom can provide a relatively fast and reliable identification of the ultimate gain and ultimate period. The parameters identified and the equations utilized in Figure 14.1 are relatively simple. The tests should be done at different operating conditions and the worst case used (largest ultimate gain and ultimate period). If the Ziegler Nichols ultimate oscillation tuning rules are employed, the PID gain factor applied to the ultimate gain should be decreased by factor of two or more and the reset time factor applied to the ultimate period increased by a factor of two or more to increase the gain margin and phase margin, respectively. The change in factors reduces the chances of an oscillatory response from changes in the open loop gain and total loop dead time and provides a reasonable good starting point for further testing and tuning. The method enables model based adaptive control to identify the process dynamics much faster because the starting point of the search region is based on the tuning settings.

Most issues as to length of the test and disruption to the process can be mitigated by adhering to recommendations 1 through 8 in Section 14.1.3 and best practices. The following best practices are useful for most types of auto tuners and adaptive control software. Note that best practices 6 and 7 can save an enormous amount of test time.

Best Practices:

1. Set the tuner noise band to be half the peak to peak amplitude of noise.
2. For liquid pressure and flow loops, use a relatively small step size (e.g., 2 percent).

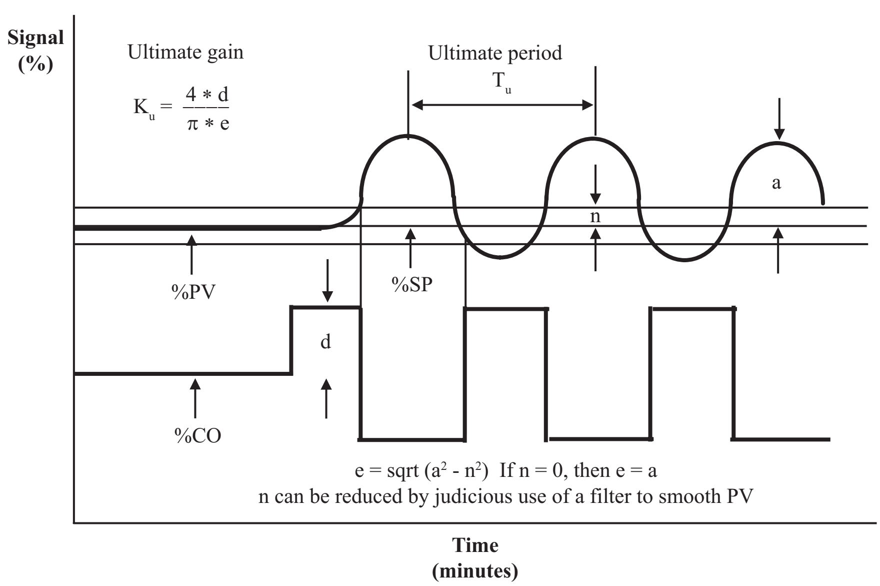


Figure 14.1. Relay oscillation method.

3. For pH loops with setpoints on the steep part of the titration curve, use an extremely small step size (e.g., 0.5 percent).
4. For pH loops with setpoints on the flat part of the titration curve, use a moderate step size (e.g., 5 percent).
5. For temperature and pressure loops on gas flow reactors and furnaces, use a relatively small step size (e.g., 2 percent).
6. For level loops and composition and temperature loops on large well mixed volumes (batch or continuous) use a large step size (e.g., 10 percent). If these large step sizes are going to upset users of the same resource (e.g., utility headers), put a setpoint rate limit (e.g., 1 percent per second) on the AO or secondary flow loop and enable external reset feedback. This directional move suppression may be useful long term especially for startup and setpoint changes.
7. For gas pressure loops, level loops, and composition, pH, and temperature loops on large well mixed liquid volumes (batch or continuous), check the integrator box so that the pretest only needs to see the initial ramp rate and does not wait for the process to decelerate to a final value.
8. If possible do tests at different production rates, setpoints, and run or cycle times and use the most conservative tuning settings until adaptive control can be implemented.

The ability to use simulation to see the effect of changes in PID settings and dynamics can improve both the performance and robustness of the PID. This can be done by a virtual plant or by use of a “Simulate” capability built-into the tuning software. The ability to prevent an excessive oscillatory response for changing dynamics should be checked on a robustness plot, such as phase margin versus gain margin. In general there is a tradeoff between load response performance and robustness.

The gain margin is the ratio of the PID gain where instability starts to the actual PID gain. The gain where instability starts is the ultimate gain for a proportional only controller. A gain margin of five means that the process gain can change by a factor of five before the PID becomes unstable. Increasing the PID gain decreases the gain margin but decreases the IAE and peak error. Low gain margins will not cause an overshoot unless the phase margin is also low.

The phase margin is the amount of additional phase lag in degrees that can occur before the loop becomes unstable. Dead time is the principal source of additional phase lag followed by the secondary time constant and then possibly the primary time constant depending upon the primary time constant to dead time ratio. Decreasing the reset time decreases the phase margin and decreases the IAE but increases overshoot.

Note that severe oscillations will result before the gain margin or phase margin is reached. In nearly all industrial process control applications, the gain margin should not be less than 4 and the phase margin should not be less than 70 degrees for worst case dynamics (highest open loop gain and dead time and lowest primary time constant). PID tuning with a Lambda greater than twice dead time for these worst case dynamics in a moderate self-regulating process gives a gain and phase margin above these limits.

Auto tuning and adaptive control software should be subsequently used that identifies the dynamics of the open loop response and offers the practitioner the choice of using the best tuning rules and factors for the application.

Since model based adaptive control needs to define a search region, when the adaptive control is first applied to a loop, the search region is initialized based on the tuning settings.

The software may require five or more tests depending upon how far away the initial dynamics are from the final dynamics estimated. In many cases, the identified dynamics will continue to change. If the changes become consistently less than 10 percent, the new settings may be used after evaluation based on tuner metrics and reasonableness in terms of a visual identification of the dead time and ramp rate from trend chart traces for the changes in the PID output during the tests. The tests noted in recommendations 17 through 19 in Section 14.1.3 should be done with the new settings. New tuning settings should be closely observed and should not be implemented near the end of the day or workweek. The diagnostics detailed in Section 5.4 should be used to decide if there is a tuning problem and what is the corrective action.

The PID tuning settings to minimize the IAE from load fast disturbances are achieved by a lower gain margin and lower phase margin. The phase margin does not affect the peak error appreciably except for dead time dominant loops. The setpoint overshoot for these tuning settings for minimum IAE and peak error can be minimized by the use of a setpoint lead-lag or 2DOF PID structure.

There can be many tuning objectives other than minimum IAE and peak error. If the load disturbances are very slow as would be the case for pH and temperature control in a bioreactor, minimizing IAE and peak error takes a back seat to eliminating overshoot and minimizing drastic changes in the manipulated reagent or utility flow. These other criteria are particularly important for mammalian cells since the agitation rate is low and the sensitivity to local variations in pH and temperature is high. Regions of high reagent concentrations or hot spots can decrease the cell growth rate and if severe enough can increase the cell death rate.

Other chapters have shown how an increase in gain margin by an increase in Lambda has been able to improve loop coordination, maximize the absorption of variability in surge tanks, and moderate the effects of interaction, nonlinearities, and inverse response.

KEY POINTS

1. The identification of the ultimate period and ultimate gain can enable diagnostics to identify and prevent existing and potential tuning problems.
2. Simulation is extremely valuable for exploring "what if" scenarios for the effect of tuning settings on the setpoint and load response and the gain and phase margins.
3. Nearly all processes are nonlinear and require gain margins greater than five and phase margins greater than 70 degrees unless signal characterization, scheduling of tuning settings, or adaptive control is extensively used.
4. A choice of tuning rules in auto tuners and adaptive control enable the practitioner to apply the rules needed to achieve diverse objectives and make the most of onsite expertise. There are many tuning objectives besides minimum IAE and peak error.
5. Testing the load response and setpoint response of new tuning settings at different setpoints, production rates, batch cycle times, and continuous production run times is essential to verify performance and robustness.

CHAPTER 15

Batch Optimization

15.1 INTRODUCTION

The highest value-added products use batch operations. Batches can take days to complete and be worth millions of dollars. In many cases bad batches cannot be fixed downstream, thus bad batches must be avoided. There are many, techniques for making batches more repeatable and faster by better monitoring and control. Essential are an appreciation of automation and an understanding of the nonself-regulation inherent in a batch process and the implications in terms of control strategy and controller tuning.

15.1.1 PERSPECTIVE

Pharmaceutical and specialty chemical manufacturing tend to use batch unit operations because these mimic the lab bench top operations in product research and development (R&D) offering the shortest time to market. The pharmaceutical industry is starting to change goals. Big driver for batch is minimizing cross contamination risk and bracketing of out of spec material so it can be quarantined. Additionally, the batch can be held until the product meets specifications allowing for time for conversions and separations to complete and for operations to make corrections. Small volume high value added products with patent protection benefit the most from batch operations. The extreme example is biological pharmaceuticals where each batch is worth millions of dollars, total yearly production may be less than 100 kg, and the clock is ticking on patent protection. The other major factor preventing continuous bioprocesses is the buildup of inhibiters and contaminants, viruses, harmful bacteria, and non-viable, dead and mutated cells.

While continuous operations can provide greater capacity, some of the feed or incompletely processed material in well mixed volumes is being discharged, reducing yield and necessitating, in many cases, recovery and recycle. The absence of a liquid discharge flow in batch operations until the end of the batch leads to a lack of self-regulation that is the key to the distinctive features of batch dynamics and the unique control system requirements. While some reactions can be self-limiting in that reactant consumption is functionally similar to a discharge flow, unbalances (e.g., excess reactant as per stoichiometry) can cause a buildup of reactant. Consequently, batch composition, pH, and temperature response join the gas pressure response as integrating or runaway processes. The continual accumulation feeds and products by separation and conversion results in process conditions changing throughout the batch. There is no steady state in the conventional sense. The response may be in one direction only (unidirectional). The dynamics are nonlinear. Since model predictive control (MPC) thrives on steady state linear processes and requires a response in both directions, a creative translation of controlled variables is needed. One such possibility is the use of the slope of batch concentration, pH, or temperature profile as a controlled variable. The profile slope has a steady state, is more linear, and can decrease besides increase in value.

Valve position control (VPC) and override control are commonly used for optimization of batch cycle time. These advanced regulatory control systems are less dependent upon a process model and can be readily configured using existing process variables (PVs) to achieve simple optimizations such as the maximization of feed rate for fed-batch operations.

In pure batch operations, the feeds are sequenced on and off based on total charges and batch logic. Phases of pressurization, heating, and cooling are sequenced as well. Temperature, pH, and pressure control loops are cycled in and out of service.

In fed-batch operations, the feed rates are typically manipulated by flow loops. Often flow ratio control is used. An analyzer can be used to provide a higher level of control to correct the ratio to maintain the stoichiometry. The ratio of utility to feed flow may also be used to provide a desired vaporization, heating, or cooling rate. Fed-batch operations create the opportunity for more feedforward and feedback control and optimization opportunities. Some call fed-batch semi-continuous because there is a throttled rather than a sequenced flow rate. Since there is no liquid discharge flow, the process response is still integrating or runaway. In fed-batch operation, the variability in the composition profile is transferred to variability in the manipulated feeds. Many process engineers are reluctant to turn over the transfer of variability to a control loop. Most want to fix both the feed rates and the composition not realizing that variability does not disappear but is transferred from the controlled variable (composition profile) to the manipulated variable (feed flow). Some process engineers try to duplicate some of the benefits of fed-batch operation by scheduling the flow rate and timing of a sequenced feed at opportune points.

Unlike continuous processes, setpoint response rather than load response is most important. Often the overshoot is more critical than rise time (time to reach setpoint). Some applications might require increasing the batch temperature to help dissolving additives; if the product of the batch needs to be fed at lower temperatures to the next process stage, an overshoot in batch temperature represents an unnecessary and costly waste of energy. For bioreactors with mammalian cell cultures, the overshoot typically must be less than 0.05°C and 0.05 pH, in temperature and pH loops, respectively, for a shift in the setpoint to optimize product formation. The time to reach setpoint in these bioreactors is a small fraction of the total batch cycle time of 10 days or more. Also, achieving specified quality is more important than reducing batch cycle time. In many cases, the batch time is fixed and conservatively set for biologic products. For these applications where rise time is not important, a setpoint filter equal to the reset time or a structure of proportional derivative (PD) on PV and I on error is an effective solution for eliminating overshoot, provided a tuning method such as Lambda tuning for integrating processes is used so that correct balance between integral mode and proportional mode is maintained. The balance is expressed by the product of the proportional-integral-derivative (PID) gain and reset time being greater than twice the inverse of the open loop integrating process gain.

For more mature and higher volume products, batch yield and capacity tend to become more important. To reduce batch cycle time, sequential operations can be replaced with simultaneous operations. Time intervals between actions can be reduced and actions can be intelligently automated, such as the automatic detection of end points and initiation of the next phase. Valve positions of both on-off and throttle valves can be maximized. Setpoint rise time can be minimized by intelligent scheduling of PID outputs. Yield can be increased by optimization of setpoints and feed rates based on raw material analysis and feedback correction by composition profile control or analysis of batch end points.

15.1.2 OVERVIEW

Batch composition, pH, pressure, and temperature control can be approximated as integrating processes and integrating tuning rules used. The setpoint response of control loops (e.g., temperature, pH, and pressure) is important in batch operations. For fed-batch operations, feed rates can be optimized to meet batch performance objectives, such as yield, quality, repeatability and/or capacity. For traditional batch operations, sequences are automated and the timing and charges are optimized to achieve batch objectives.

For high value small volume products, such as biologics, the quality is critical. The product quality that involves complex folding of the protein is difficult to measure and correlate. Repeatability is important for all batches particularly for pharmaceutical and for food and beverage processes. Large blend volumes downstream of the batch operation may make these requirements less stringent on a batch to batch basis. For mature higher volume products, yield and capacity take on a greater emphasis.

Yield, capacity, quality, and repeatability are interrelated. Variability affects all of these metrics and the solutions to improve one metric have benefits that extend to the other metrics. Often batches are held longer than necessary due to the concern about variability. Interruptions in the execution of the batch logic and the dependence upon operations to resolve problems directly affects batch cycle time and repeatability.

If all interruptions and manual actions have been eliminated by automation, most of the remaining sources of variability can be traced to variability in the composition and mass totals of components added to the batch and deficiencies in the control system components, strategy, implementation, and tuning.

For chemical processes, the totals of each component added must be accurately set and achieved. Coriolis meters can be used to measure the concentration of a key component in the feed stream via an extremely precise density measurement and to provide the ultimate in feed totalization by an exceptionally accurate mass flow measurement. For reactions, achieving the right stoichiometric ratio and best temperature are critical. If the concentration of just one component is deficient, the other components remain unreacted.

The presence of trace components in the raw materials can inhibit conversion and cause product quality problems. Raw material storage should have at-line analyzers if feasible. If laboratory analysis must be used, the time from when the sample taken to the result is entered into the control system must be fast enough to compensate for the effect in an existing batch and correct the feed for the next batch.

For pharmaceuticals the principal source of variability are the cells (incredibly complex microscopic reactors) and the magical mixture of nutrients. For fermentation processes, the variability is in the composition and processing of the grains (e.g., fermentability) and in the recycle streams (e.g., backset for ethanol). Analyzers of the feeds can provide an opportunity for an immediate adjustment in feed rate to the unit operations for processing of the feeds (e.g., slurry tanks for ethanol) and the feed totals to the fermenters.

Multivariate statistical process control better known as data analytics can be a particularly effective tool for tracking down the source of variability in batch operations and making mid batch corrections. The current batch is compared to an average batch. Principal Component Analysis (PCA) can tell you if anything in the current batch has changed. Drilling down into the PCA can give the contribution of each batch measurement to the variability observed. PCA along with a process understanding can be used to trace abnormal batches back to equipment and automation system problems. The PCA is also employed in Projection to Latent Structures also known as Partial Least Squares (PLS) to predict key compositions at various points in the batch cycle. Often a correction at the mid batch point provides enough time to achieve a better endpoint.

The PLS predictions of endpoint do a piecewise linearization of the batch process conditions and can thus handle the process nonlinearity inherent in batch processes. Unlike in the use of PLS for continuous processes, delays do not need to be inserted on the PLS inputs to synchronize with a PLS output. In continuous processes the process time constant and transportation delay for each PLS input must be used to provide the coincident effect in the path to a downstream PV that is the PLS output. In batch processes these time delays are not needed since the effect is on a PLS output at a given point in the batch cycle (e.g., end point).

The response of temperature for heating without cooling, pH for a single reagent (acid or base), and product composition are in one direction only. For heating without cooling the temperature, for a base reagent the pH, and for a non-reversible conversion the product concentration, can only increase. Although the control valve is totally closed, temperature and pH still ramp up due to the residual heat in the jacket and reagent in the dip tube, respectively. The effect is more pronounced as the vessel volume decreases. Any integral action will cause overshoot of the setpoint from which there is no recovery. Proportional plus derivative control is needed. A smart bang-bang control to stop the PV just short of the setpoint can minimize rise time. A translation of the temperature, pH, or composition to a batch profile slope can give a decrease as well as increase in the controlled variable enabling the use of PID or MPC for profile optimization.

Inferential measurements enable more effective monitoring and automation of batch processes. The measurements can be representative of the rate of product formation or in the case of biological processes, cell growth and death rate.

The cooling rate in chemical processes offers an inferential measurement of an exothermic reaction rate or crystallization rate. The totalized cooling rate can be an indicator of total conversion. Cooling rate can be computed by the temperature difference between jacket or coil inlet and output temperatures multiplied by the flow through the jacket or coil. The inlet temperature is passed through a dead time block to mimic the transportation delay to synchronize the inlet with the outlet temperature.

The oxygen uptake rate (OUR) in bioreactor can be computed as the oxygen flow from air or oxygen spargers minus the overhead vapor oxygen concentration measured by a mass spectrometer multiplied by the vapor flow. For tight dissolved oxygen control, the oxygen flow may be representative enough of changes in oxygen demand. The OUR offers an inferential measurement of cell growth rate assuming oxygen uptake for cell maintenance and product formation are negligible.

In ethanol and wine production, the carbon dioxide production rate (CPR) in a fermenter can be computed as the overhead vapor oxygen concentration measured by a mass spectrometer multiplied by the vapor flow. The CPR in fermenters can also be computed as the loss in weight measured by load cells. The CPR gives an inferential measurement of yeast growth rate and alcohol production rate.

A dielectric probe and spectrum analyzer can provide an inferential measurement of cell volume, size, and membrane integrity (cell viability). The rate of change of this measurement can be used as an inferential measurement of cell growth or death rate (e.g., cell membrane rupture).

A sudden increase in conductivity may signal the end of chlorination processes. Similarly, a sudden change in pH may indicate the completion of a reaction where a reagent is being consumed.

When a batch composition or temperature profile is measured but not controlled, a decrease in the slope of the PV is indicative of the end of conversion processes. If the profile is controlled, a decrease in the slope of the manipulated variable (e.g., feed or cooling rate) is indicative of an end point being reached.

The computation of the rate of change of a process or manipulated variable is useful in providing batch profile slopes, inferential measurements, and future values for the intelligent scheduling of PID outputs and detection of batch end points. If the variable is passed through a dead time block equal to the process dead time, the signal to noise ratio is improved and the computation of the future value and rate of change is simple and immediate (new results for every update of the process or manipulated variable). The block input (new value) minus the block output (old variable) gives the change in variable in one dead time (delta variable). If you add the block input to this delta variable, you have a value of the variable one dead time into the future. If you divide the delta variable by the dead time, you have the rate of change of the variable. If you want to compute a future value more than one dead time into the future, you can multiply this result by the desire time interval into the future.

VPC and override control has been effectively used to push fed-batch feed rates to the maximum controllable position for jacket and overhead condenser coolant valves and overhead vapor valves. The lowest output of the VPC for each valve position pushed to the limit is selected as the setpoint of the lead feed flow rate controller. Override controllers on PVs can prevent violation of process constraints such as high level from bubbles, density changes, and higher batch volumes.

15.1.3 RECOMMENDATIONS

1. Use Coriolis meters on liquid feeds for precise concentration control and extremely accurate component totalization.
2. Analyze raw materials for trace components that could affect conversion time or product quality and take compensatory actions.
3. Make unit operations such as filling, heating, and pressurization as simultaneous as possible to reduce batch cycle time.
4. Automate all manual actions by operations.
5. Eliminate wait and hold times by proceeding without manual operator or lab data entry or approval by using inferential measurements and online data analytics.
6. Reduce failure expression activation and automate failure expression recovery in batch manager software.
7. Use data analytics to monitor batch repeatability and predict endpoints.

8. Use integrating process tuning rules for primary batch composition, pH, pressure, and temperature PID controllers.
9. For unidirectional (single ended) process responses where PV can only go in one direction (e.g., heating with no cooling), use a PD controller (no integral) or translate PV to be the slope of the desired profile.
10. To control the slope of the batch profile and eliminate the one direction response setpoint overshoot, use the rate of change of the key batch PV as the controlled variable.
11. Optimize the setpoint response for overshoot and rise time by the choice of structure and the use of setpoint lead-lag and the scheduling of PID output and PID tuning (Chapters 2, 12, and 14).
12. Compute the rate of change of a PV or manipulated variable for indicators and controllers, respectively, to predict and detect the completion of batch phases.
13. Compute inferential measurements of conversion such as cooling rate and total for chemical reactors and crystallizers, OUR and total for bioreactors, and CPR and total for fermenters.
14. To compute the rate of change or future value of a process or manipulated variable, use a dead time block to improve signal to noise ratio and immediacy.
15. Optimize feed rates based on raw material and batch composition analysis.
16. Use the enhanced PID for batch composition control with at-line or offline analyzers (Chapters 6 and 12), for stopping limit cycles from deadband, and resolution limits, and to enable direction move suppression for split range point discontinuities (Chapter 7) and VPC (Chapter 12).
17. Use VPC and override control to maximize fed-batch feed rate.
18. Use a virtual plant to explore and develop opportunities.

15.2 CYCLE TIME

Most batch processes are held longer than necessary by about 10 percent unless there is product degradation. There is also another 10 percent capacity that can be gained by reducing the source of variability and providing more optimum operating conditions. The total average opportunity of 20 percent for a batch capacity increase is about 10 times larger than what is possible in mature continuous processes. The opportunity is even greater for batches that are interrupted due to problems with on-off valves or batch logic or activation of Safety Instrumented System (SIS) or batches that are waiting on lab results or operator action.

Poor control valve design and variable frequency drive implementation can increase batch variability. Deadband and resolution limits can cause limit cycles and overshoot in temperature, pressure, and composition loops. Discontinuities in the transition between reagents for pH, between air and oxygen for dissolved oxygen, and between heating and cooling for temperature can cause cycling back and forth across the split range point.

Data analytics can show the degree of deviation of a batch from normal. The normal batch is often an average of representative batches. A bad batch can be terminated or corrected. A drill down into the contributions to the PCA can help identify the sources of variability that should be addressed if not in time for this batch than at least for future batches.

An increase in feed conversion capability often shows up as a shorter batch time to the desired endpoint. Either product inhibition for biological processes or reactant depletion for chemical processes results in no further conversion toward the end of the batch.

The following best practices are useful for most types of batch operations. The details on the implementation of most of the items involving control loops for temperature and fed-batch control have been covered in previous chapters. For example, design and implementation considerations are addressed for the better use of PID features and tuning parameters in Chapters 1, 2, and 5, the proper selection of measurements and final control elements in Chapters 6 and 7, and the effective design of advanced regulatory control systems (e.g., bang-bang, override, and VPC) in Chapter 12.

Best Practices:

1. Eliminate wait times, operator attention requests, excess process hold times, and manual actions by use of automation, at-line analyzers, and data analytics.
2. Reduce lag times by better sensor and valve design and location.
3. Reduce SIS interruptions by tighter control and optimization.
4. Minimize acquire time by improving prioritization of users and releases.
5. Reduce failure expression activation by better instruments, redundancy and signal selection, and more realistic expectations of instrument performance.
6. Improve failure expression recovery by configuration and display methods.
7. Eliminate sequential phases by simultaneous actions (e.g., heat-up and pressurization).
8. Increase feed and heat transfer rate by an increase in pump impeller size.
9. Reduce transition time by full throttle and profile rate of change control.
10. Minimize processing time by constraint control (override control).
11. Minimize non constrained processing time by all out run, cutoff, and coast.
12. Minimize processing time by pairing of loop variables to maximize conversion rate.
13. Minimize processing time by better endpoint prediction and detection.
14. Minimize processing time by capture of last batch's best outputs for next batch.
15. Maximize effectiveness of solutions by the use of a virtual plant in the development and testing and training for maximum utilization of the improvements.

The ability to use higher fidelity real time simulations in virtual plants to see the effect and train the operators on the use of the improvements listed is particularly powerful. In the virtual plant, the actual control system configuration is imported and downloaded and the actual operator graphics are used on the consoles. The batches use various speedup factors to run 5 to 200 times real time depending upon the batch length. For bioreactor, kinetic speedup factors are used to provide a multiplicative effect. Thus, a liquid factor speedup of 10 accompanied by a kinetic speedup factor of 20 can achieve a real time speedup of 200 times real time. The scale-up of feed flows and totals are handled internally by the virtual plant model and are transparent to the user so that results match up with what the control system and the operator would see as a function of batch time.

15.3 PROFILE

Nearly all chemical and biological reaction rates depend upon the concentrations of the reactants and quality depends upon the resulting product and byproduct concentrations. Yet, you would be pressed to find offline let alone at-line or on-line concentration measurements of any components of reactants, products, by-products, or contaminants during any batch. The best kept secret of batch reactors are the concentration profiles.

In product development, the concentration profiles are measured in the lab with bench top analyzers. The chemist or process development engineer knows the values of these profiles but how can the need for this knowledge be realized and carried over to the commercial plant? A virtual plant can open minds and provide the justification for batch concentration profile analysis by prototyping advanced controls to make batch reaction rates more repeatable and faster. The virtual plant can also verify how fast do you need concentration results from the analyzers to do closed loop control.

A computation of the heat removal rate in a cooling system with a sensible heat correction can provide an inferential measurement of conversion rate for exothermic reactors. If there are no significant side reactions, the conversion rate can be taken as being specific to the production rate. If the jacket circulation rate is constant and known, a flow measurement is not necessary. The description here is for a jacket but the same principles and considerations apply to a coil.

The heat removal rate uses temperature sensors on the inlet and outlet of the jacket. The inlet temperature measurement is sent through a dead time block with a dead time setting equal to the transportation delay so the jacket inlet temperature can be synchronized with the jacket outlet temperature.

Resistance temperature detectors (RTDs) provide a more accurate inferential measurement than thermocouples which is critical is to the computation of small differences between the inlet and outlet temperatures. The heat removal rate is the outlet temperature minus the synchronized inlet temperature multiplied by the jacket flow and the heat capacity of water at the operating temperature. If the jacket flow is not constant, the heat removal computation and the transportation delay must be updated based on a jacket flow measurement. For a constant coil or jacket flow and cooling or chilled water temperature, the difference between the reactor and coil or jacket outlet temperature (approach temperature), can be used as a simpler computation. The use of a jacket recirculation system to maintain a constant jacket flow simplifies the computation and offers tighter temperature control by preventing a low jacket flow that would greatly increase the process dead time and decrease the heat transfer coefficient. The integration of the heat removal rate over the course of a batch or for the residence time of a continuous reactor can provide an inferential measurement of total conversion.

Wireless transmitters for temperature and pressure compensated annubar mass flow meters can be inserted in spare nozzles to establish the variability in coil and jacket flow and utility temperature besides inexpensively prototyping inferential measurements. The portability of wireless measurements enables one to demonstrate and quantify the benefits of online metrics and diagnostics in less than the time spent in meetings guessing the value and feasibility.

For batch profile composition control, it is critical to use the slope rather than the actual product concentration. The product concentration should always be increasing. This unidirectional response is a problem for the MPC, which like all feedback controllers is expecting it can decrease a controlled variable. By using product formation rate as the controlled variable, the MPC can see decreases as well as increases during identification of the process model. The use of an MPC can shorten batch cycle times by 25 percent or more. Alternately an MPC can increase the product concentration for the same batch cycle time. Figure 15.1 shows such a case for a bioreactor with an MPC whose controlled variables were growth rate and product formation rate and the manipulated variables were glucose and glutamine concentrations. The MPC penalty on error was reduced for growth rate so that a higher importance was placed on product formation rate. The MPC not only greatly reduced the cycle time but also improved the predictability of the product concentration at the end of the batch by a more repeatable profile

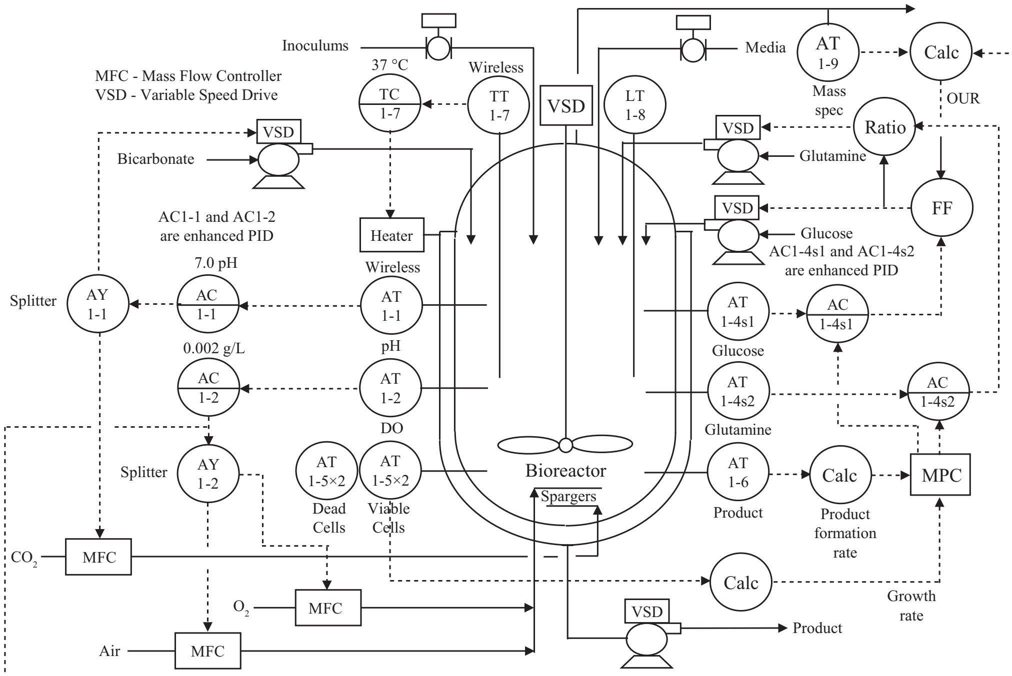


Figure 15.1. Control strategies for batch bioreactor. Cell and product concentration profile control.

and uniform slope, important for real time release. In this case, the MPC set points simply tracked the PVs when the MPC was in manual. The MPC was switched to auto at the peak in the product formation rate which captured the max product formation rate (max slope of the product concentration profile) as the target for the batch and adapted to changes in initial conditions and inoculums.

If an inferential measurement of conversion rate (e.g., cooling rate or OUR) is not available, a periodic composition measurement is needed. In this case, continuous online and relatively frequent at-line analyzer measurements are inputted to a dead time block to create a continuous train of old measurements. For batches that take days to complete, lab analysis results can be used. A new measurement minus an old measurement divided by the dead time is the slope of the concentration profile. The dead time must be set at least as large as the time in between analysis results but is increased to be a multiple of the analysis time to provide a good signal to noise ratio.

15.4 END POINT

From the slope near the end of the batch, the additional product produced in the dead time interval or the analysis time interval for sampled measurements is computed. The slope (e.g., conversion rate, cell growth rate, product formation rate) near the end of the batch is used to make an economic decision whether the batch should be terminated for extra capacity or extended for extra yield. The slope is converted to product mass flow and multiplied by the analysis time interval to get the additional product mass for the given dead time or analyzer time interval.

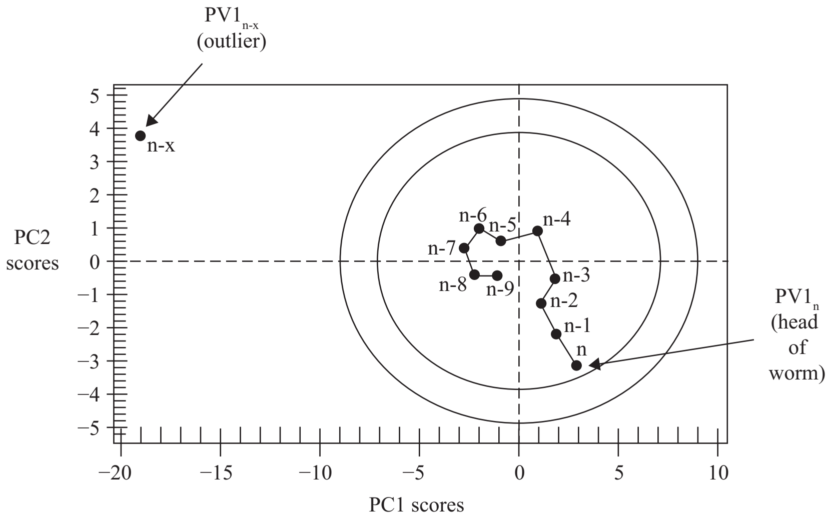


Figure 15.2. Worm plot of batch end points shows where batches are headed.

The current product mass in the batch divided by the mass of each key raw material added to the batch gives the yield in terms of product for each key raw material. The additional product mass per batch is divided by this yield and multiplied by the cost per unit mass of each key raw material. The results are summed to arrive at a dollar value of the additional yield by extending the batch giving additional product for the raw materials used.

The current product mass in the batch divided by the current batch time in hours offers an estimate of the current batch production rate. Alternately the production rate can be used from the flow controller based on predicted feed yield. The production rate multiplied by the profit per unit mass and finally multiplied by the dead time or analysis time interval gives an estimate of the value of additional capacity by terminating the batch.

The analysis time interval should be shortened to be just large enough for a good signal to noise ratio near the end of the batch to make the optimization more accurate. If the analysis of the profile of a key composition or product is not available until after the batch has been transferred, the results can be cautiously used as a fractional correction for the next batch.

A worm plot as shown in Figure 15.2 can be used to predict when a batch is going to be outside the quality specifications. The worm plot consists of batch endpoints plotted as scores versus principal components from multivariate statistical models where outliers are removed. The points are distinguished by numbers or data point size to show the progression of batches as a worm. Analysis includes whether the worm is coiled as desired within the batch specifications shown as a circle or ellipse or are the latest batches uncoiling and is the head approaching the specification limits.

KEY POINTS

1. The average opportunity for a capacity increase in batch operations is ten times larger than what is seen in continuous operations.

2. Batch quality, repeatability, capacity, and yield are interrelated.
3. Batch composition, pH, pressure, and temperature loops have an integrating response that requires more proportional and derivative action and less integral action than continuous processes. The PID output must be driven past the final resting value.
4. Batch data analytics does not require the dynamic compensation of inputs and through a piecewise linear fit deals with the nonlinear profile of batch operations.
5. Batch data analytics can warn operators of deviant batches, enable process engineers to track down the source of batch problems, and provide worm plots to provide pattern recognition of whether batches are headed for a problem.
6. VPC and override control can maximize fed-batch feed rates.
7. Inferential measurements, such as conversion rate from cooling rate and growth rate from OUR for a chemical and biological reactor, respectively, can provide profile control and analysis.
8. The computation of the rate of change of a composition measurement from online, at-line, or offline analytical measurements can be done with a good signal to noise ratio by the simple use of a dead time block where the dead time is a multiple of the time between analyzer results. The resulting composition rate of change can be used for profile control and endpoint prediction.
9. Profile control can provide more repeatable and optimum composition trajectories. For a single optimization variable, VPC using an enhanced PID for directional move suppression is a fast and inexpensive solution. For multivariable control, complex dynamic and optimums, MPC using the slope of composition profiles as the controlled variables and a linear program for the finding and achieving the optimum, is a more effective and complete solution.
10. Endpoint prediction can be used in an economic analysis of whether the batch should be shortened for a capacity increase or lengthened for a yield increase.

APPENDIX A

Automation System Performance Top 10 Concepts

Automation system performance not only depends upon tuning but also upon the performance of each component in each control loop. System performance can be maximized by attention to design and implementation details that affect the ability of the controller, final control element(s), equipment, piping, sensor, and transmitter to do their job. Here is a summary of the Top 10 Concepts to guide the automation engineer in this endeavor. We finish with a perspective on the relative importance of accuracy and precision.

1. *Delay*
    - *Without dead time I would be out of a job*
    - *Fundamentals*
        - A more descriptive name would be *total loop dead time*. The total loop dead time is the amount of time required for the start of a change to completely circle the control loop and end up at the point of origin. For example, an unmeasured disturbance cannot be corrected until the change is seen in the controlled variable and the correction arrives in the process at the same point as the disturbance.
        - Process dead time offers a continuous train of values whereas digital devices and analyzers offer non continuous data values at discrete intervals, the resulting delays add a phase shift and increase the ultimate period (decrease the natural frequency) like process dead time.
    - *Goals*
        - Minimize delay (the loop cannot do anything until it sees and enacts change).
    - *Sources*
        - Pure delay from process dead times and discontinuous updates
            - Piping, duct, plug flow reactor, conveyor, extruder, spin-line, and sheet transportation delays (pure process delays set by mechanical design— remaining pure delays set by automation system design)
            - Digital device scan, update, reporting, and execution times
            - Analyzer sample processing, analysis cycle time, and multiplex time
            - Sensor threshold sensitivity

        - Signal resolution
        - Valve stiction
        - Valve backlash
        - Variable speed drive deadband
        - Wireless trigger level
        - Wireless update rate
    - Equivalent delay from lags
        - Mixing, column trays, dip tube size and location, heat transfer surfaces, and volumes in series (process lags set by mechanical design—remaining lags set by automation system design)
        - Thermowell lag
        - Electrodes
        - Transmitter damping
        - Signal filters

2. *Speed*
    - *Speed kills—high speed processes and disturbances and with low speed control systems can kill performance*
    - *Fundamentals*
        - The rate of change in four dead time intervals is most important. By the end of four dead times, the control loop should have completed most of its correction. Thus, characterizing self-regulating processes with a large time constant to dead time ratio as near-integrating process is consistent with performance objectives.
    - *Goals*
        - Make control systems faster and make processes and disturbances slower.
    - *Sources*
        - Control system
            - Proportional-integral-derivative (PID) tuning settings (gain, reset, and rate)
            - Slewing rate of control valves and velocity limits of variable speed drives
        - Disturbances
            - Steps—batch operations, on-off control, manual actions, Safety Instrumented Systems, startups, and shutdowns
            - Oscillations—limit cycles, interactions, and excessively fast PID tuning
            - Ramps—reset action in PID
        - Process
            - Degree of mixing in volumes due to agitation, boiling, mass transfer, diffusion, and migration

3. *Gains*
    - *All is lost if nothing is gained*
    - *Fundamentals*
        - Gain is the change in output for a change in input to any part of the control system. Thus, there is a gain for the PID, valve, disturbance, process, and measurement. The open loop gain that is the product of the valve, process, and measurement gain is needed for controller tuning. Knowing the disturbance gain (e.g., the change in manipulated flow per change in disturbance) is important for sizing valves and for feedforward control).

- *Goals*
  - Maximize control system and process gains
    - Maximize valve sensitivity
    - Maximize sensor sensitivity
    - Maximize transmitter gain (minimize calibration span)
    - Maximize controller gain
    - Maximize process sensitivity (e.g., find best distillation column tray)
- Minimize disturbance gain
  - Minimize size of upstream disturbances by automation (e.g., elimination of oscillations and transfer or absorption of variability by control).
    - Decrease limit cycles
    - Decrease interactions
    - Replace operator actions and on-off control with throttling control.
    - Increase absorption of flow variability by surge tank level tuning.
    - Use redundancy to minimize failures (e.g., middle signal selection).
  - Minimize size of process disturbances by enabling process self-regulation (e.g., cancel or moderate disturbances).
    - Get internal reflux control by tight distillate level control.
    - Use weak acid and base reagents to buffer pH systems.
- *Sources*
  - Abnormal operation and failures (e.g., pump, measurement, and valve failure)
  - PID controller gain
  - Inferential measurement location (e.g., sensitivity of temperature change as an inference of composition change in distillation column)
  - Sensor type (e.g., threshold sensitivity of radar vs. d/p level measurement and RTD vs. thermocouple)
  - Control valve or variable speed drive installed characteristic (gain is slope that includes effect of inherent characteristic and system pressures)
  - Measurement calibration (gain is 100%/span)
  - Process design
  - Backlash and deadband (limit cycle amplitude)
  - Stiction and resolution limits (limit cycle amplitude)
  - Interaction (fighting loops oscillation amplitudes)
  - Operator actions
  - Safety Instrumented Systems
  - Sequences

4. *Resonance*
   - *Do not make things worse than they already are*
   - *Fundamentals*
     - Oscillation period close to ultimate period can be amplified by feedback control (Figure A.1).
   - *Goals*
     - Make disturbance oscillation period slower or control loop faster.
     - For interacting loops with similar priorities, make fast loop faster.
     - For interacting loops with different priorities, make least important loop slower.

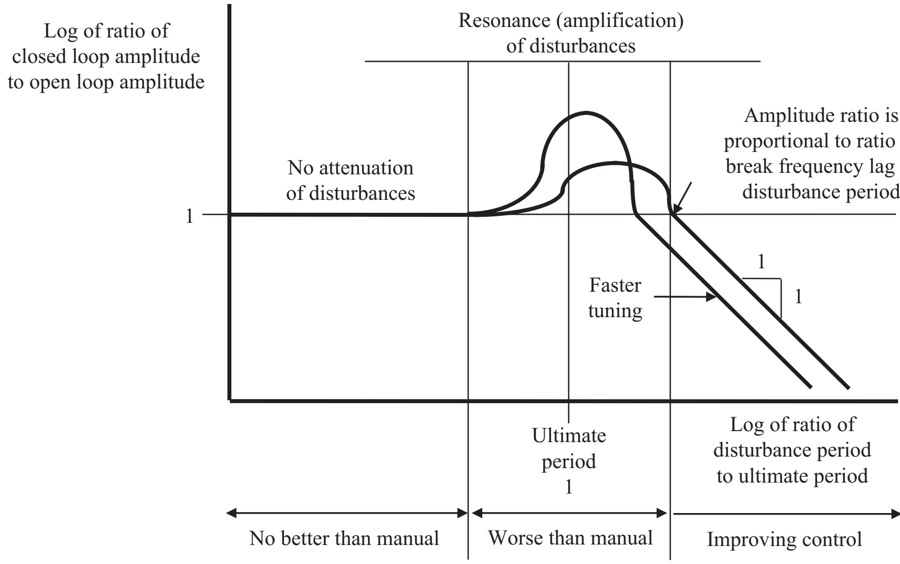


Figure A.1. An oscillation with a period near the ultimate period (frequency near the natural frequency) is amplified by the PID from *resonance*.

- *Sources*
    - Control loops in series with similar dead times (e.g., multiple stage pH control)
    - Control loops in series with similar tuning and valve stiction and backlash
    - Control loops that affect each other with similar dynamics
    - Day to night ambient changes to slow loops (e.g., column temperature control)

5. *Process Attenuation*
    - *If you had a blend tank big enough you would not need control*
    - *Fundamentals*
        - Attenuation increases as the volume of the blend tank increases and the ultimate period of the control loop decreases (filter time constant is back mixed volume residence time) (Figure A.2).
    - *Goal*
        - Maximize attenuation by increasing volume and mixing and making loops faster.
    - *Sources*
        - Mixed volume size and degree of back mixing

$$A_f = \frac{A_o}{\sqrt{1+(\tau_f * \omega)^2}} \tag{A.1a}$$

The attenuated (filtered) amplitude (A_f) for original amplitude (A_o) and attenuation (filter) time constant (τ_f) that is much greater than the oscillation period (T_o) using the relationship between oscillation frequency and period ($\omega = \frac{2*\pi}{T_o}$):

$$A_f = \frac{T_o}{2*\pi*\tau_f} \tag{A.1b}$$

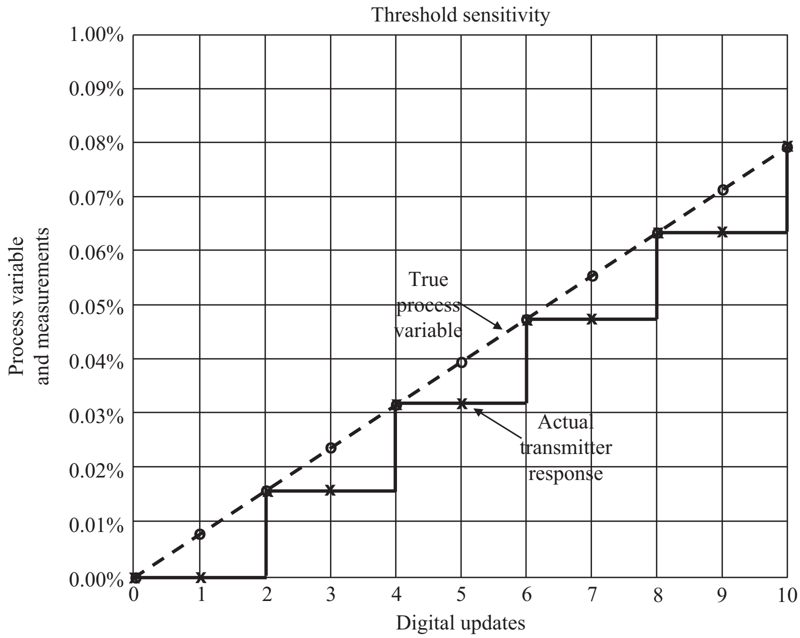


Figure A.2. When the change in process variable exceeds the *threshold sensitivity limit*, the full change in PV is seen.

Equation A.1b is also useful for estimating the filtering effect of thermowell lag, electrode lag, transmitter damping, and signal filtering on process oscillation. The equation can also be used to estimate original process amplitude given measurement amplitude when measurement lag is causing an attenuated version of the real process variability.

6. *Threshold Sensitivity and Resolution*
   - *You cannot control what you cannot see*
   - *Fundamentals*
     - Minimum change measured or manipulated—once past threshold sensitivity limit full change is seen or used (Figure A.2) but resolution limit will quantize the change (stair step where the step size is the resolution limit) (Figure A.3). Both will cause a limit cycle if there is an integrator in the process or control system.
   - *Goal*
     - Improve threshold sensitivity and resolution.
   - *Sources*
     - In measurements, minimum change detected and communicated (e.g., sensor threshold and wireless update trigger level) and quantized change (A/D & D/A)
     - Minimum change manipulated (e.g., valve stick-slip and speed resolution)
7. *Backlash Deadband and Hysteresis*
   - *No problem man—if you do not ever change direction*
   - *Fundamentals*
     - Hysteresis is the bow in a response curve between full scale traverses in both directions. Normally much smaller and less disruptive than backlash deadband (Figure A.5).

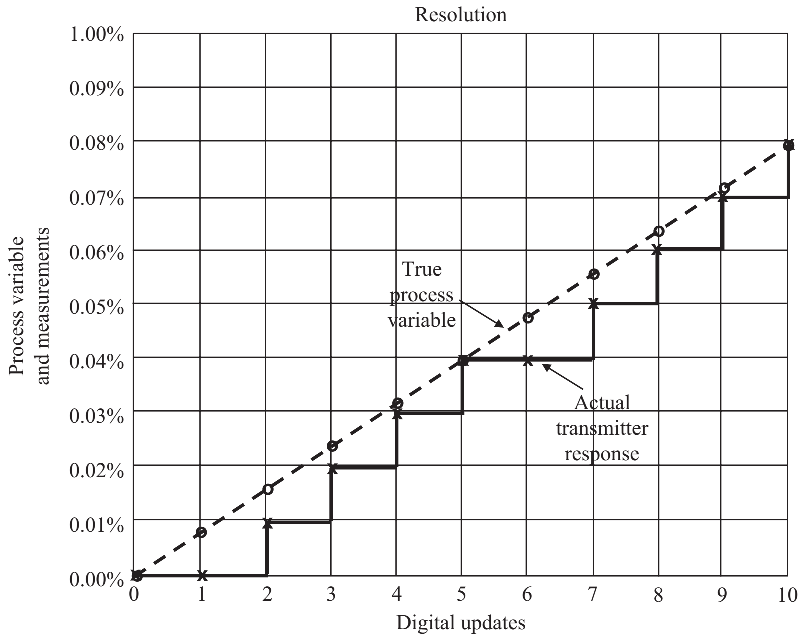


Figure A.3. When the change in process variable exceeds the *resolution limit*, the change in PV is seen as a step the size of the resolution limit.

- Backlash deadband is minimum change measured or manipulated once the direction is changed—once past backlash-deadband limit you get full change (Figure A.4).
- Backlash deadband will cause a limit cycle if there are two or more integrators in the process or control system.
- Deadband can be a parameter in a PID or a variable frequency drive (VFD).

- *Goals*
  - Minimize hysteresis in measurements (eliminate mechanical components).
  - Minimize backlash deadband in final control elements.
  - Minimize deadband settings in PID and VFD.
- *Sources*
  - Pneumatic instrument flappers, links, and levers (hopefully these are long gone)
  - Rotary valve and damper links, connections, and shaft windup
  - VFD setup parameter to eliminate hunting and chasing noise

8. *Repeatability and Noise*
   - *The best thing you can do is not react to deception*
   - *Fundamentals*
     - Noise is extraneous fluctuations in measured or manipulated variables.
     - Repeatability is the difference in readings for same true value in same direction (Figure A.6).
     - Repeatability is often confused with noise.
   - *Goals*
     - Minimize size and frequency of noise.
     - Do not transfer noise to process.

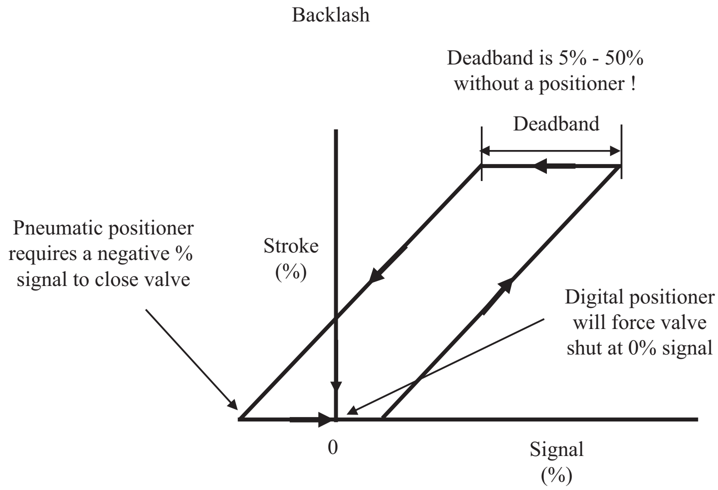


Figure A.4. When the controller output changes direction and the change exceeds the *deadband*, the full change in valve stroke or drive speed is seen.

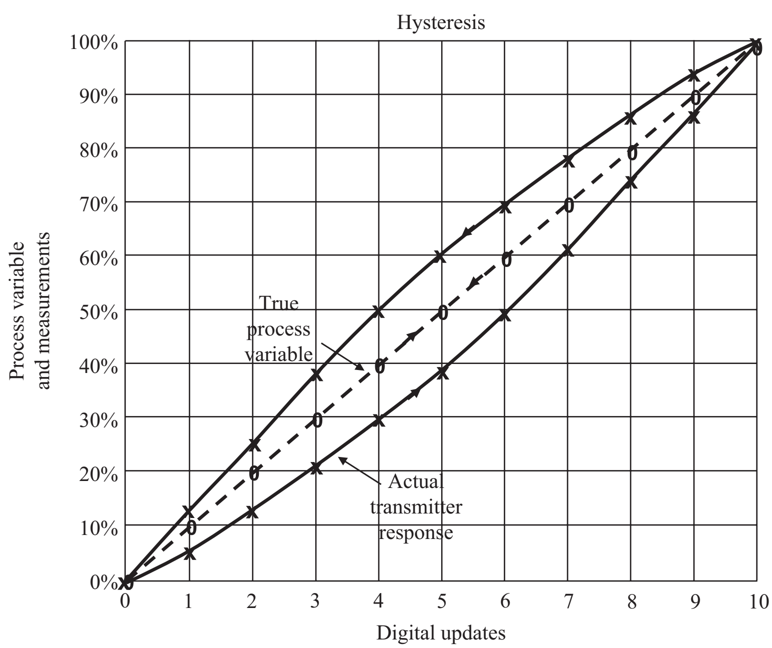


Figure A.5. When the controller output changes direction, the path is bowed by the *hysteresis* but the change is immediately and fully seen.

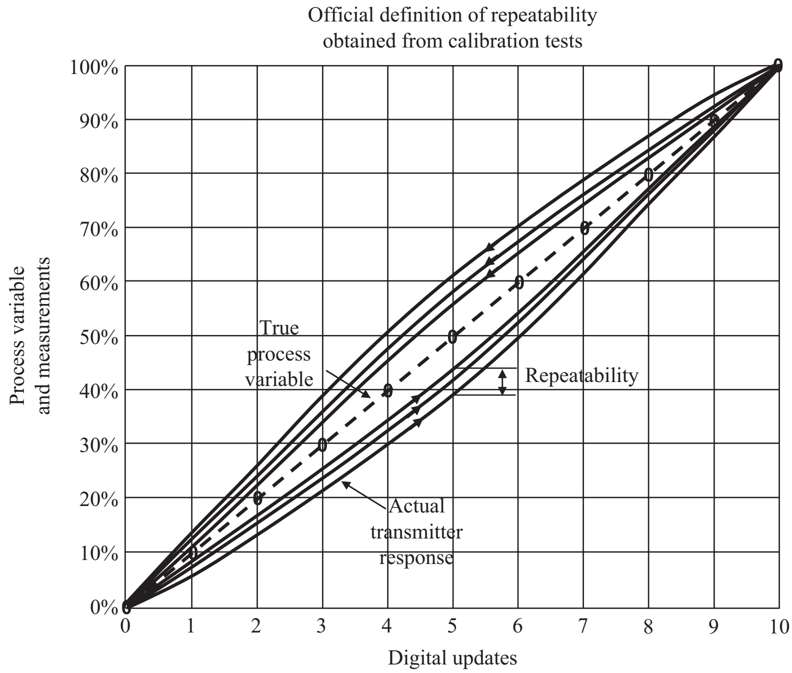


Figure A.6. The maximum separation of the traverses of the scale range in the same direction is the *repeatability*.

- *Sources*
  - Noise
    - Bubbles
    - Concentration and temperature non-uniformity from imperfect mixing
    - Electromagnetic interference
    - Ground loops
    - Interferences (e.g., sodium ion on pH electrode)
    - Velocity profile non-uniformity
    - Velocity impact on pressure sensors
  - Repeatability
    - Threshold sensitivity, resolution, and extraneous effects

9. *Offset and Drift*
   - *There are always an offset and drift; it is matter of their size and consequence.*
   - *Fundamentals*
     - Offset is the deviation of the peak in the distribution of actual values from true value (Figure A.7).
     - Drift shows up as a slowly changing offset (Figure A.8).
   - *Goals*
     - Minimize offset and nonlinearity (see Concept 10) by using smart transmitters and sensor matching and smart tuned digital positioners with accurate internal closure member feedback.

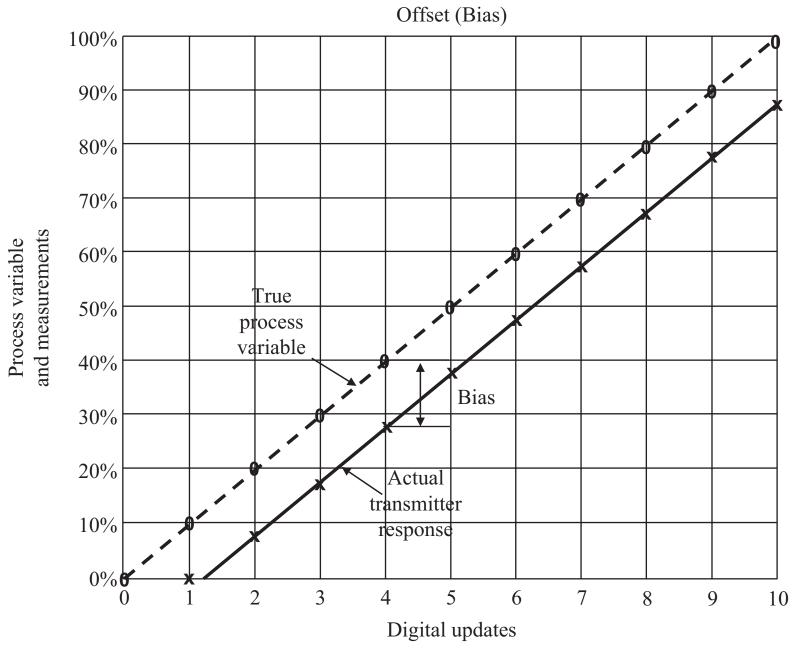


Figure A.7. An *offset* (bias) is a constant difference between the process variable and the measurement in a complete traverse of the scale range.

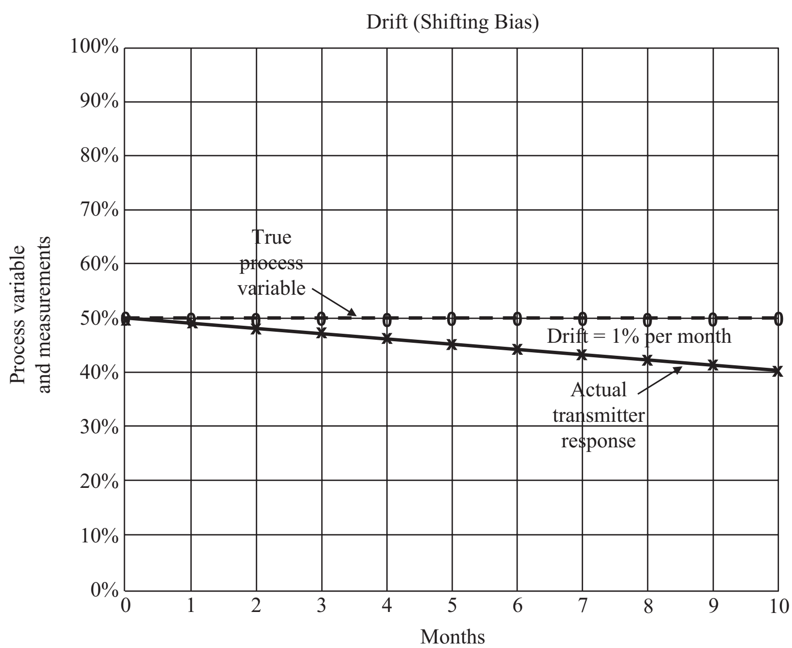


Figure A.8. A shifting difference between the process variable and the measurement with time is *drift*.

- *Sources*
  - Manufacturing tolerances, degradation, de-calibration, and installation effects (process and ambient conditions and installation methods and location)

10. *Nonlinearity*
- *"Not a problem if the process is constant, but then again if the process is constant, you do not need a control system."*
- *Fundamentals*
  - While normally associated with a process gain that is not constant, in a broader view, a nonlinear system result occurs if a gain, time constant, or delay changes anywhere in the loop. In addition, all process control systems are nonlinear to some degree (Figure A.9).
- *Goals*
  - Minimize nonlinearity by process and equipment design (e.g., reagents and heat transfer coefficients), smart transmitters and sensor matching, valve selection, signal characterization, and adaptive control
- *Sources*
  - Control valve and variable speed drive installed characteristics (flat at high flows)
  - Process transportation delays (inversely proportional to flow)
  - Digital and analyzer delays (loop delay depends upon when change arrives in discontinuous data value update interval)
  - Inferred measurement (conductivity or temperature vs. composition plot is a curve)

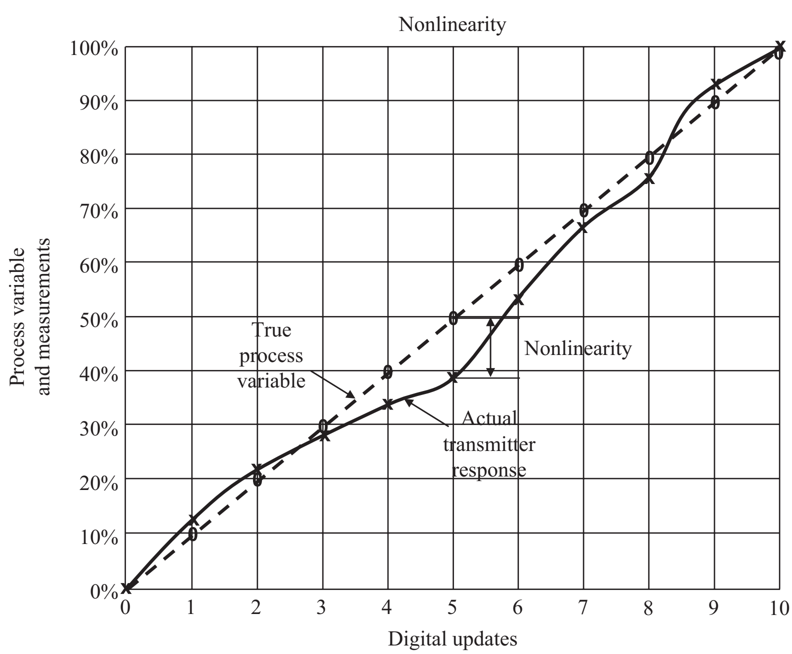


Figure A.9. A difference between the process variable slope and the measurement slope for a traverse of the scale range is *nonlinearity*.

- Logarithmic relationship between sensor output and process concentration (glass pH electrode and zirconium oxide oxygen probe)
- Process time constants (proportional to volume and density)

WHICH IS MORE IMPORTANT, ACCURACY OR PRECISION?

The accuracy of a measurement is determined by as the difference between the true value and the measured value. The precision is how much the measured value changes for repeated measurements. In a plot of measurement frequency versus measured value for the same true process variable value (Figure A.10), a mean measurement value closer to the true process value is indicative of better accuracy. A tighter distribution of measurement values (small standard deviation), is indicative of better precision.

Inaccuracy can be corrected by changes in setpoint, either manually or by upper level loops, so that the effect is less on performance and more on process calculations and analysis. In contrast, a lack of precision adds an uncertainty for which there are few remedies for the resulting poor loop performance. For example, control valve stroke inaccuracy will introduce errors in system pressure drop and installed flow characteristic calculations but the effect on loop performance will be minimal because the process control loop throttling the valve will adjust the signal to the valve as needed. Similarly, secondary flow measurement inaccuracy will create errors in material balance calculations but the effect on process performance will be minimal because the primary controller in a cascade control system will eliminate the bias by adjusting the secondary controller setpoint.

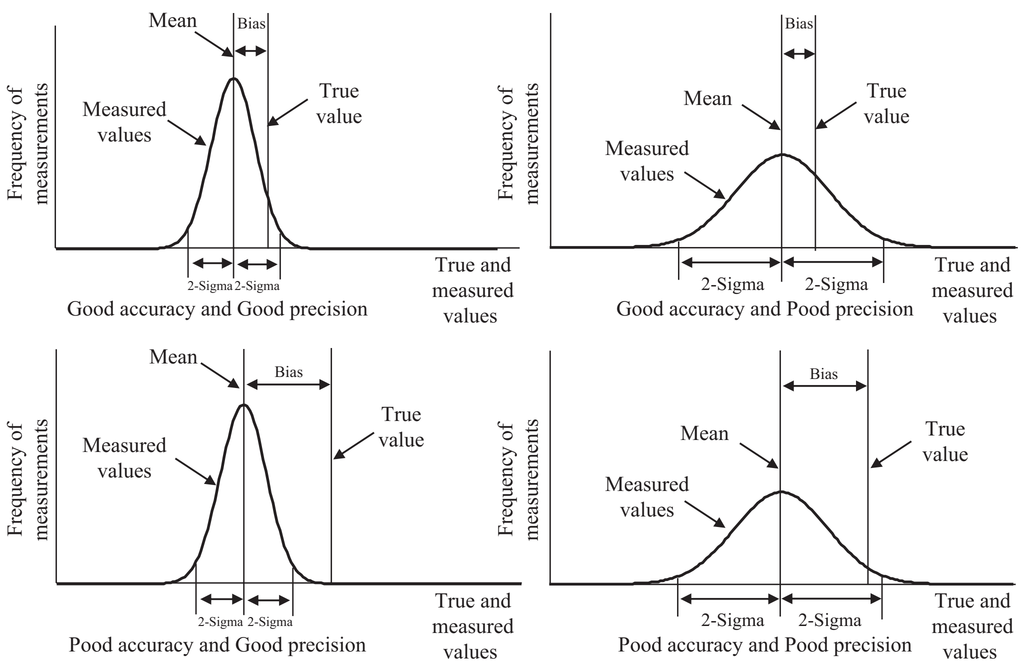


Figure A.10. *Accuracy* can be corrected by a setpoint but *precision* is prevalent uncertainty.

Consequently, improving factors that affect precision, such as threshold sensitivity, repeatability, and resolution and minimizing backlash, deadband and noise, are more important than reducing factors that affect accuracy, such as offset, drift, and nonlinearity.

Besides the precision of measurements and final control elements, minimizing dead time and resonance, slowing down disturbances and making them smaller, optimizing process gains, and maximizing process attenuation are important for loop performance. What is left is controller tuning, the primary subject of this book.

APPENDIX B

Basics of PID Controllers

Each of the three Proportional-integral-derivative (PID) controller modes have distinct advantages and disadvantages. Here is a review of the basic functionality of each mode.

Figure B.1 shows the contribution of each mode to the output for a setpoint change. In this example, the proportional mode and the derivative mode are acting on the error (the difference between the setpoint and the process variable [PV]) rather than just on the PV. In the time period shown, there is no update from the process measurement so we can see more clearly the function of each mode without any feedback from the process.

PROPORTIONAL MODE BASICS

The proportional mode provides an immediate reaction to a change in the PV. If the PID structure on proportional action on error ($\beta = 1$), there is an immediate reaction to a setpoint change as well as shown in Figure B.1. The change in percent output is the change in percent setpoint multiplied by the controller gain. For a step change in setpoint, there is a step in the controller output from the proportional mode. This step will help get the PV to setpoint sooner and in particular will eliminate the dead time from valve stiction and backlash. However, this step can be too large or disruptive to other loops for some applications. In addition, operators get concerned about rapid changes. The use of a setpoint filter or setpoint rate limit will slow down the change in the PID output and the approach to setpoint, reducing overshoot. The tradeoff is an increase in rise time (the time to reach setpoint). If the setpoint filter time constant is set equal to the reset time, the PID effectively has a structure of proportional action on PV rather than on error.

The contribution from the proportional mode will decrease for a PV that is increasing and approaching the setpoint (reverse acting). This change, counteracting the approach provides some anticipatory action and a sense of direction. If the proportional mode contribution is greater than the integral mode contribution, overshoot is minimized for controller gain settings that provide a fast nonoscillatory (critically damped) response. If the proportional action is too great, there will be a faltering in the approach to setpoint and oscillations that could lead to overshoot. Thus, we have the interesting situation where too low a gain can cause an overshoot by not counteracting the integral mode and too high a gain can cause an overshoot from oscillations.

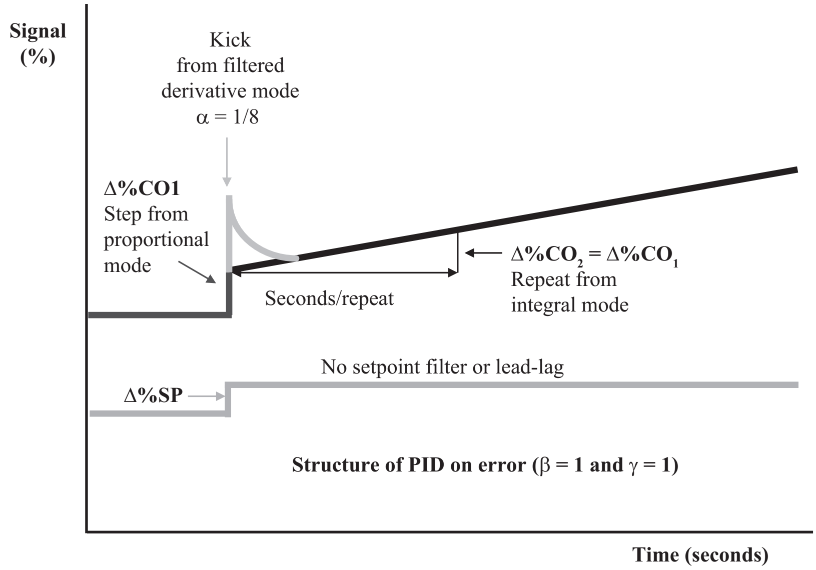


Figure B.1. Response of each PID mode to a change in setpoint for PID action on error.

If there is no change in the PV or setpoint, the proportional mode does not change the output. Thus, a proportional only or a proportional plus derivative controller will not cause a limit cycle from stiction or backlash in a self-regulating process.

Note that many analog controllers and one major Distributed Control System (DCS) use proportional band instead of gain for the proportional mode tuning setting. Proportional band is the % change in the PV (Δ%PV) needed to cause a 100% change in controller output (Δ%CO). A 100% proportional band means that a 100% Δ%PV would cause a 100% Δ%CO, which is a gain of 1. It is critical that users know the units of their controller gain setting and convert accordingly.

$$\text{Gain} = 100\%/\text{Proportional Band}$$

PROPORTIONAL MODE ADVANTAGES

- Minimize dead time from stiction and backlash
- Minimize rise time
- Minimize peak error
- Minimize integrated error

PROPORTIONAL MODE DISADVANTAGES

- Steps in output upset operators
- Steps in output upset other loops
- Amplification of noise

INTEGRAL MODE BASICS

The integral mode integrates the error with respect to time. For a constant error as shown in Figure B.1, the result is a constant ramp rate. The integral mode will continue to increase the controller output even though the PV is increasing and approaching setpoint (reverse acting) because the error is negative. Integral action will not change sign until the PV crosses setpoint. Consequently, there is no anticipatory action or sense of direction of change. Integral action that exceeds proportional and derivative action will delay an output in getting off of an output limit and will cause overshoot. Excessive integral action can also cause a runaway reaction by accelerating a temperature that is increasing and approaching a higher setpoint.

If there is no change in the PV or setpoint, the integral mode will continue to ramp since the error is never exactly zero. Thus, a controller with an active integral mode will cause a limit cycle from stiction in a self-regulating process and from backlash in an integrating process. The "ideadband" (integral deadband) parameter will suspend integral action when the PV is within the specified deadband. The enhanced PID developed for wireless will inherently suspend integral action if there is no PV update. In this case a small threshold sensitivity setting is used to prevent a reaction to noise.

Note that many analog controllers use reset settings in repeats per minute instead of reset time for the integral mode tuning setting. Repeats per minute indicate the number of repeats of the proportional mode contribution in a minute. Today's reset time settings are minutes per repeat or seconds per repeat which gives the time to repeat the proportional mode contribution. The *per repeat* term is often dropped, giving a reset time setting in minutes or seconds.

Seconds per repeat = 60/repeats per minute

INTEGRAL MODE ADVANTAGES

- Eliminate offset
- Minimize integrated error
- Smooth movement of output

INTEGRAL MODE DISADVANTAGES

- Limit cycles
- Overshoot
- Runaway of open loop unstable reactors

DERIVATIVE MODE BASICS

The derivative mode provides an output that is proportional to the rate of change of the error with respect to time. For a step change in error, the result is a spike. A built-in derivative filter with a time constant that is fraction α of the rate time, changes the spike to a kick as shown in Figure B.1. Unless the trend chart update time is fast and time scale is short, the kick will look like a spike. When derivative action is used on the PV instead of error, the spike is eliminated since there is rarely a step change in the PV of loops where derivative is used. Derivative is not used in dead time dominant loops because the changes in the PV are abrupt due to the lack of a significant process time constant. If the full response occurs within one dead time, derivative action can do more harm than good. The most effective use of derivative action is the compensation (cancelation) of a secondary lag, particularly in an integrating or runaway process.

Too much derivative action will cause an oscillatory approach to setpoint with a frequency faster than the natural frequency of the loop. These oscillations can persist after the setpoint is reached.

Nearly all derivative tuning settings are given as a rate time in seconds or minutes. The effective rate time setting must never be greater than the effective reset time setting. The effective settings are for an ISA Standard Form. For a PID with Series Form, an interaction factor prevents the effective rate time from becoming larger than the reset time.

The tuning methods developed prior to the 1990s were for the Series Form predominantly in use at the time. These tuning methods may set the rate time equal to the reset time. The use of these settings in an ISA Standard Form as predominantly used today will result in severe oscillations.

The advantages and disadvantages of the derivative mode are similar to those of the proportional mode except the relative advantages are less and the relative disadvantages are greater for the derivative mode.

DERIVATIVE MODE ADVANTAGES

- Minimize dead time from stiction and backlash
- Minimize rise time
- Minimize peak error
- Minimize integrated error

DERIVATIVE MODE DISADVANTAGES

- Kicks in output upset operators
- Kicks in output upset other loops
- Amplification of noise

SPLIT RANGE EXAMPLE

I have had several reports from control rooms stating that a PID controller was doing the wrong thing. Consider the temperature control loop where the PV is 48 degrees, the SP is 52 degrees, and the output is split ranged between a water valve and a steam valve with the split range point at 50 percent. An operator looking at digital values on the operator graphics or the faceplate in Figure B.2 wants the steam valve to be open when in fact the water valve is open. He will ask for the controller tuning to be corrected to make this happen, which results in a technician or engineer adding more integral action to get the steam valve open. Integral action has no sense of

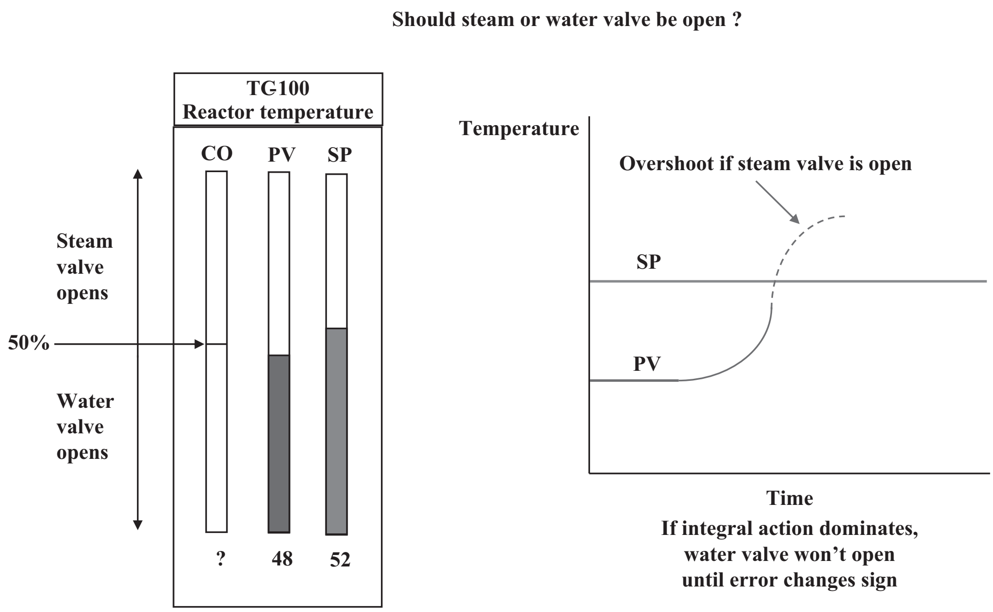


Figure B.2. An operator looking at digital values thinks steam valve should be open whereas operator seeing trend trajectory realizes water valve should be open.

direction and will help meet our expectations and deal with impatience when looking at digital displays.

An operator looking at the trend chart and understanding dead time will realize that the temperature was on a trajectory to overshoot the setpoint and any change now will take one dead time to have an effect. The water valve should be open. Improvements in operator displays will continually show relative rate and direction of change and values one dead time into the future.

OPEN LOOP GAINS

The contribution of each mode in the predominant PID Forms used in industry is proportional to the controller gain since this gain is a multiplier in each mode. The controller gain is inversely proportional to a gain that is the change in the percent input for a change in percent output. The process control literature typically specifies this as a process gain without consideration of the contribution of individual components such as installed flow characteristics, feed rate, PV as a function of feed rate, and measurement span.

To analyze the effects of each contributing component and to account for fundamental differences in the process response, the terms open loop self-regulating process gain or open loop integrating process gain are used instead of simply process gain. The open loop gain is the change in percent PV for a change in percent PID controller output for the controller in manual hence the term *open loop*. Equation B.1a shows that the open loop gain is the product of the gains for each component in the loop. This breakdown of gains allows one to understand the implications of a change to final control elements, system pressures, production rates, process nonlinearities, and transmitter calibration.

PID SCALES

While the user sets and sees the PID PV and output scales in engineering units, all calculations in the PID algorithm for the contribution of each mode are done in percent. Thus, all tuning calculations for controller gain have to include the effect of input and output scale span as shown in Equation B.1a for the open loop gain.

Except for the primary controller in a cascade control system, the PID output scale is 0 to 100 percent. When there is a primary controller output scale, it is usually set equal to the secondary controller setpoint scale, effectively cancelling out any effect of the scale ($K_v * K_r = 1$). Consequently, we can focus on the effect of the PV scale span on the measurement gain to account for the effect of PID scale choice.

We see that the open loop gain is inversely proportional to the PV scale span. Since the controller gain is inversely proportional to this open loop gain, the controller gain is proportional to the measurement span. Thus, if we decrease transmitter span to improve measurement accuracy, which is often expressed in percent of calibration span, we need to proportionally decrease the controller gain for the PID to achieve the same speed of response. If the controller was tuned for maximum disturbance rejection, the controller gain must be decreased to maintain stability. If the PID was tuned with too small a PID gain, the smaller transmitter span will provide tighter control. Thus, one can understand the confusion about the effects of a transmitter span since changing the span can make a loop do better or worse.

INSTALLED FLOW CHARACTERISTICS

The final control element gain is the change in flow for a change in controller output (Equation B.1b). This gain is the slope of the installed flow characteristic for the control valve or variable speed drive. For a control valve, the installed flow characteristic is the inherent trim characteristic for a constant pressure drop. For a variable speed drive the installed flow characteristic is linear for zero static head.

$$K_o = K_v * K_r * K_p * K_m \tag{B.1a}$$

$$K_v = \frac{\Delta F_v}{\Delta \% CO} \tag{B.1b}$$

For composition, temperature, and pH loops:

$$K_r = \frac{\Delta R}{\Delta F_v} = \frac{\Delta F_v / F_p}{\Delta F_v} = \frac{1}{F_p} \tag{B.1c}$$

$$K_p = \frac{\Delta PV}{\Delta R} \tag{B.1d}$$

For flow, level, and pressure loops:

$$K_r * K_p = \frac{\Delta PV}{\Delta F_v} \tag{B.1e}$$

$$K_m = \frac{100\%}{\left(PV_{100\%} - PV_{0\%}\right)} \tag{B.1f}$$

PROCESSES

For composition, temperature, and pH loops in continuous operations, the process gain is the change in slope in a plot of the PV versus a ratio of the manipulated flow to the process flow. Consequently, the process gain is inversely proportional to the process flow (Equation B.1c).

For flow, level, and pressure loops, the PV is not a function of this flow ratio. As a result, the process gain simplifies to being just the change in the PV with the manipulated flow (Equation B.1e).

For integrating processes, such as level, the open loop self-regulating process gain (K_o) is replaced with an open loop integrating process gain (K_i) with inverse time units (1/sec) that results from the process gain term units not cancelling out the flow time units. For mass flow, the level process gain is simply the inverse of the product of the vessel cross sectional area and the liquid density. The PV is a level in length units (e.g., meters).

Nomenclature for System Gains:

K_m = measurement gain (%/PV e.u.)
K_i = open loop integrating process gain (1/sec)
K_o = open loop self-regulating process gain (%/%) (dimensionless)
K_p = process gain (PV e.u.)
K_r = flow ratio gain often embedded in process gain (1/flow e.u.)
K_v = valve or variable speed drive gain (manipulated flow per % output)
$\Delta\%CO$ = change in controller output converted to % of controller output scale (%)
ΔF_v = change in valve or variable speed drive flow (flow e.u.)
F_p = process flow at current production rate (flow e.u.)
$\Delta\%PV$ = change in PV converted to percent of controller input scale (%)
$\%PV_{100\%}$ = PV at 100 percent scale value (PV e.u.)
$\%PV_{0\%}$ = PV at 0 percent scale value (PV e.u.)
ΔR = change in ratio of manipulated flow to process flow (dimensionless)

APPENDIX C

Controller Performance

The following is based on an excerpt from Chapter 14 that I contributed to *PID Control in the Third Millennium: Lessons Learned and New Approaches* (Vilanova and Visioli 2012).

Special algorithms can be designed to deal with measured load disturbances at the process input, setpoint changes, and disturbances at the process output (e.g., noise). Often neglected is the overriding requirement that controllers in industrial applications must be able to deal with unmeasured and unknown load disturbances at the process input. Fortunately, the proportional-integral-derivative (PID) controller excels at this load disturbance rejection. An estimate of the current and best possible load rejection as a function of the process and automation system dynamics and controller tuning provides the information on what can be done to improve plant design and tuning. A simple set of equations can be developed that estimates the integrated error and peak error for a step change in a load disturbance. The value is more in helping guide decisions on improvements rather than predicting actual errors because of the uncertainty of the size and speed of load disturbances and the nonlinear and nonstationary nature of industrial processes. The equations are simple enough to provide key insights as the relative effects of the controller gain and integral time and the first order plus dead time (FOPDT) approximation of the process and automation system dynamics. In FOPDT model, a fraction of each of the time constants smaller than the largest time constant is taken as equivalent dead time and summed with the pure dead times to become the total loop dead time (θ_o) termed a process dead time (θ_p) in the literature. The fraction of the small time constants not taken as dead time is summed with the largest time constant to become the open loop time constant (τ_o). While the equations for tuning and estimation of errors is based on the open loop time constant, we will assume the largest time constant is in the process so we have the more common term of process time constant (τ_p) seen in the literature. In reality, fast loops, such as liquid flow and pressure, have a time constant in the FOPDT model much larger than the flow response dead time due to a transmitter damping setting and signal filter time constant. Similarly, the equations seen in the literature use a process gain (K_p) rather than the open loop gain (K_o) that is the product of the final control element, process, and measurement gain. For improving dynamics, a distinction of the location of nonlinearities, dead time, and the largest time constant are important. By avoiding the categorization of dynamics as being solely in the process, a better understanding of the effect of the final control element size, installed characteristic, stick-slip, and backlash, the effect of measurement noise, lag, delay, calibration span, and the effect of PID filter and execution time is possible. The nomenclature used in the quantification of these effects is defined at the end of the chapter.

Since a controller cannot compensate for an unmeasured load disturbance before the loop dead time, the peak error (E_x) (maximum error for a disturbance) is the excursion of the first order response to the step disturbance (E_o) based on the open loop time constant for a time duration of the loop dead time (Equation C.1). The open loop error is the final error seen at the PID from an unmeasured load disturbance if the PID was in manual. The terms *open loop* and *closed loop* are used for a response without and with feedback correction, respectively.

$$E_x = [1 - e^{-\theta_o/\tau_o}] * E_o \tag{C.1}$$

If the total loop dead time is much larger than the open loop time constant, then the peak error is basically the open loop error. If the dead time was less than the time constant, then Equation C.1 can be simplified to Equation C.2 eliminating the exponential term. The ultimate limit to the peak error can then be visualized as the ratio of the dead time to 63 percent response time (dead time plus time constant), which is the altitude of the triangle shown in Figure C.1.

$$E_x = \frac{\theta_o}{(\theta_o + \tau_o)} * E_o \tag{C.2}$$

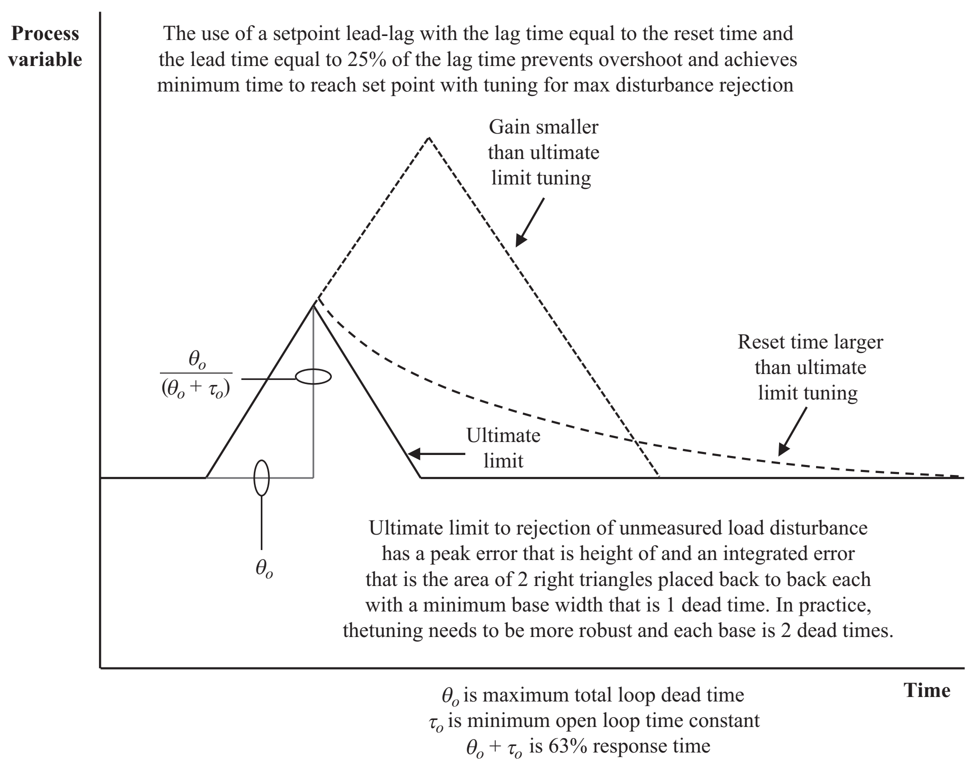


Figure C.1. Slower than optimum tuning increases the minimum peak error and integrated errors that are the height and area of an equilateral triangle.

The minimum integrated error (E_i) can be approximated as the area of two right triangles with the altitude equal to the peak error and the base equal to the dead time as depicted in Figure C.1. The two right triangles form an equilateral triangle. Taking the area of each triangle as half the base multiplied by the altitude, we obtain Equation C.3 where the integrated error is simply the peak error multiplied by the dead time and consequently proportional to the dead time squared.

$$E_i = \frac{\theta_o^2}{(\theta_o + \tau_o)} * E_o \tag{C.3}$$

Equations C.2 and C.3 are for the minimum possible errors determined by the open loop process and system automation system dynamics. It is not possible to do better than what is permitted by the dynamics of the process and automation system. Thus, Equations C.2 and C.3 provide a benchmark for the ultimate limits to loop performance for unmeasured load disturbances. What is achieved in feedback control depends upon the tuning. In practice, controllers are not tuned aggressively enough to achieve the ultimate limit because the response tends to be too oscillatory especially for large setpoint changes and the controller lacks robustness. A 25 percent increase in loop dead time or open loop gain or 25 percent decreases in the open loop time constant can result in oscillations that do not sufficiently decay. Consequently, a more practical benchmark would be right triangles each with a base of two dead times and a time to return to setpoint of four dead times.

We can develop the equations that set the practical limit in terms of controller tuning settings from the equations for the ultimate limit based on open loop dynamics. We will also see that we can independently arrive at the same equation for the integrated error from the response of the PI algorithm to a step disturbance.

If we divide through by the dead time term in Equation C.2, we have Equation C.4 where the peak error depends upon the ratio of the open loop time constant to total loop dead time.

$$E_x = \frac{1}{(1 + \frac{\tau_o}{\theta_o})} * E_o \tag{C.4}$$

Most tuning methods for maximum disturbance rejection use a controller gain (K_c) that is proportional to the ratio of the open loop time constant to total loop dead time and inversely proportional to the open loop gain (Equation C.5).

$$K_c = \frac{\tau_o}{\theta_o * K_o} \tag{C.5}$$

If we solve for the open loop time constant to total dead time ratio, we see this ratio is simply the product of the controller gain and open loop gain ($K_c * K_o$). If we substitute the product for the ratio in Equation C.3, we have Equation C.6, which is the practical limit to the peak error. In practice an aggressively tuned PID controller with substantial derivative action can achieve this limit with a numerator of 1.1 for a process time constant much greater than the loop dead time. Peter Harriott developed the same form of the equation but with a numerator of 1.5 for the peak error from a proportional only controller tuned for quarter amplitude decaying response.

$$E_x = \frac{1.5}{(1 + K_c * K_o)} * E_o \tag{C.6}$$

For time constant to dead time ratios that are much larger than one, which is the case for pressure and temperature control of vessels and columns, the product of the controller gain and open loop gain is much greater than one leading to the peak error being simply inversely proportional to the product. However, due to nonideal effects, the numerator is increased for industrial applications and we end up with Equation C.7 for the peak error.

$$E_x = \frac{2}{K_c * K_o} * E_o \tag{C.7}$$

Equation C.7 corresponds to a peak error reached in about two dead times. If we approximate the integrated error as the area of two right triangles each with a base equal to two dead times and consider the integral time (T_i) setting as being four dead times, we end up with Equation C.8 for the integrated error.

$$E_i = \frac{T_i}{K_c * K_o} * E_o \tag{C.8}$$

We can derive Equation C.8 from the equation for a PI controller's response to an unmeasured load disturbance. The change in controller output from time *t1* to time *t2* is the sum of the contribution from the proportional mode and the integral mode (Equation C.9a). The module execution time (Δt_x) is added to the reset or integral time (T_i) to show the effect of how the integral mode is implemented in digital controllers. An integral time of zero ends up as a minimum integral time equal to the execution time so there is no zero in the denominator for the integral mode. For analog controllers, the execution time is effectively zero.

$$\%CO_{t2} - \%CO_{t1} = K_c * (\%E_{t2} - \%E_{t1}) + \left[\frac{K_c}{T_i + \Delta t_x}\right] * \int_{t1}^{t2} \%E_t * \Delta t_x \tag{C.9a}$$

The errors before the disturbance ($\%E_{t1}$) and after the controller has completely compensated for the disturbance ($\%E_{t2}$) converted to percent of controller input scale are zero ($\%E_{t1} = \%E_{t1} = 0$). Therefore, the long-term effect of the proportional mode, which is first term in Equation C.9a, is zero. Equation C.9a reduces to Equation C.9b.

$$\Delta\%CO = \left[\frac{K_c}{T_i + \Delta t_x}\right] * \int_{t1}^{t2} \%E_t * \Delta t_x \tag{C.9b}$$

The integrated error is the integral term in Equation C.9b giving Equation C.9c. For overdamped response the integrated error and the integrated absolute error are identical.

$$\%E_i = \int_{t1}^{t2} \%E_t * \Delta t_x \tag{C.9c}$$

If we substitute Equation C.9c into Equation C.9b, we have Equation C.9d.

$$\Delta\%CO = \left[\frac{K_c}{T_i + \Delta t_x}\right] * \%E_i \tag{C.9d}$$

The change in controller ($\Delta\%CO$) multiplied by the open loop gain (K_o) must equal the open loop error (E_o) for the effect of the disturbance to be eliminated. We can express this requirement as the change in output being equal to the open loop error divided by the open loop gain (Equation C.9e).

$$\Delta\%CO = \frac{E_o}{K_o} \tag{C.9e}$$

If we substitute Equation C.9e into Equation C.9d and solve for the integrated error, we end up with Equation C.9f, which is the same as Equation C.8 except for the addition of the execution time interval for the digital implementation of the PI algorithm.

$$\%E_i = \left[\frac{(T_i + \Delta t_x)}{K_o * K_c}\right] * E_o \tag{C.9f}$$

Recently, Greg Shinskey added a term to the numerator to include the effect of a signal filter time constant on the integrated error (Equation C.10). In Shinskey's presentation of the equation, the change in controller output rather than the open loop error is used, which eliminates the open loop gain in the denominator. Equation C.10 is applicable regardless of tuning settings. The additional equivalent dead time from the filter time and execution time interval may necessitate a decrease in controller gain and increase in integral time further degrading performance.

$$\%E_i = \left[\frac{(T_i + \Delta t_x + \tau_f)}{K_o * K_c}\right] * E_o \tag{C.10}$$

To summarize, in the process industry, automation system and process dynamics, and in particular the loop dead time, set the ultimate limit to loop performance but controller tuning sets the practical limit for unmeasured disturbances. For example, a loop with a small dead time will perform as badly as a loop with a large dead time if the PID has sluggish tuning. On the other hand, a PID with fast tuning may have an excessive oscillatory response for increases in the loop dead time or process gain. Equation C.6 shows the practical limit to the peak error (E_x) is inversely proportional to 1 plus the product of the PID gain (K_c) and the process gain (K_p). Equation C.9f indicates the integrated error (E_i) is proportional to the ratio of the PID integral time to gain (T_i / K_c). For small filters (τ_f) and PID execution time (Δt_x), the maximum controller gain is decreased, and the minimum integral time is increased based on the increase in loop dead time. The filter and execution time should also be added to the integral time for the integrated error to show the increase in the practical limit (Equation C.10). For a PID tuned for maximum disturbance rejection, Equation C.3 reveals the ultimate limit to the peak error depends upon the ratio of the total loop dead time (θ_o) to the open loop time constant (τ_o). Equations C.8 and C.10 indicates the integrated error depends upon the ratio of the loop dead time squared to open loop time constant. A PID controller tuned for maximum disturbance rejection has a controller gain proportional to the ratio of the largest open loop time constant to loop dead time (τ_o / θ_o), and an integral time proportional to the loop dead time. Note the controller tuning depends upon the largest open loop time constant and not the process time constant. If the largest time constant is in the measurement path, the observed peak error in the measurement predicted by Equation C.2 will be smaller than the actual peak error in the process because of the signal filtering effect of the measurement time constant.

The peak error is important for preventing: shutdowns from reaching trip settings of safety instrumentation systems, environmental emissions and process losses from reaching the relief settings of rupture discs and relief valves, off-spec paper sheet and plastic web from exceeding permissible variation in thickness and clarity, compressor shutdowns from crossing surge curve, and recordable incidents by exceeding environmental limits.

The integrated error is a good indicator of the quantity of liquid product off-spec in equipment with back mixing. In these volumes, positive and negative fluctuations in concentration are averaged out unless irreversible reactions are occurring.

An important emerging consideration is the realization that initial open loop response in the FOPDT approximation of a self-regulating process is a ramp seen in the response of an integrating process such as level and batch temperature. The ramp is more persistent in a self-regulating process with a large open loop time constant. The process is termed *near-integrating* or *pseudo integrating*. An equivalent integrating process gain (K_i) can be approximated as the open time constant divided by the open loop gain (Equation C.11). For processes where the open loop time constant is more than 10 times larger than the dead time, the identification of this near integrator gain in three dead times can reduce the time required for process identification by more than 90 percent compared to those techniques that go to the 98 percent response time. Since the PID algorithm works with percent signals, the process variables are converted from engineering units to a percent of PID input scale ($\%PV$). The process variable is passed through a dead time block to create an old $\%PV$ that is subtracted from the new $\%PV$ to create a $\Delta\%PV$ and then an integrating gain by dividing by the dead time and the change in controller output $\Delta\%CO$. The maximum of a continuous train of these *near-integrating* process gains updated every execution of the PID module can be used for tuning controllers on all types of processes.

$$K_i = \frac{K_o}{\tau_o} \tag{C.11}$$

If we substitute the near-integrating gain for the time constant to dead time ratio in Equation C.5, we have Equation C.12. Recently, this method was found to even work on processes where the dead time was greater than the time constant. To provide a smoother response, less setpoint overshoot and more robust settings, the controller gain in both Equation C.5 and C.12 is cut in half.

$$K_c = \frac{1}{\theta_o * K_i} \tag{C.12}$$

The optimum integral time depends upon the type of process. The integral time ranges from about four times the dead time to integrating and *near-integrating* processes to half of the dead time for severely dead time dominant process ($\theta_o >> \tau_o$). Equation C.13 provides a reasonable curve fit to the required relationship for self-regulating processes. For a dead time less much less than the time constant ($\theta_o << \tau_o$), the ultimate period is about four times the dead time, and the denominator is about one giving an integral time that is about four times the dead time. For a dead time much greater than the time constant ($\theta_o >> \tau_o$), the ultimate period is about two times the dead time, and the denominator is about four, giving an integral time that is half of the dead time.

For self-regulating processes:

$$T_i = \frac{T_u}{Min\left(4, 10 * \left(\frac{4 * \theta_o}{T_u} - 1\right)^2 + 1\right)} \tag{C.13}$$

For a dead time dominant process, the combination of Equation C.13 for integral time and Equation C.5 for controller gain results in almost an integral-only controller. Since the controller gain is so low, this process is a candidate for setpoint feedforward to reduce the setpoint response rise time.

For an integrating process, the product of the controller gain and integral time must be greater than twice the inverse of the integrating process gain to prevent slowly decaying oscillations from the integral mode dominating the proportional mode. If the user is confident in the knowledge of the integrating process gain, this relationship can be used to find the integral time (Equation C.14a). Since the maximum controller gain allowable on many level and batch temperature loops is greater than 100 and the actual controller gain used is often less than 10, the integral time must be increased to prevent the slow rolling oscillations as per Equation C.14a. Consequently, while an integral time of four dead times is possible for an integrating process, in practice an integral time of 40 dead times is more appropriate because the maximum controller gain is beyond the user's comfort level (Equation C.14b).

To prevent large amplitude slowly decaying oscillations for processes with no appreciable deceleration in the PID time frame (four dead times), the integral action must be greater than one-fourth the inverse of the product of the controller gain and integrating process gain.

$$T_i \geq \frac{1}{4 * K_c * K_i} \tag{C.14a}$$

The positive feedback in the runaway processes necessitates an integral time 10 times larger than the integral time for a *near-integrating* self-regulating process. The integral time should be 40 dead times or larger for a runaway process (Equation C.14b). Some highly exothermic polymerization batch reactors have gone to proportional plus derivative control to avoid the problem of a user setting too small of an integral time.

For integrating processes with controller gains less than 10 times, the maximum permissible controller gain and for runaways (processes with positive feedback):

$$T_i = 40 * \theta_o \tag{C.14b}$$

Too small a controller gain besides too large a controller gain can cause a runaway. There is a window of allowable controller gains for positive feedback processes. Any changes in tuning settings particularly for runaway reactions must be closely monitored.

Common metrics for a setpoint response as shown in Figure C.2 are rise time (time to reach setpoint), overshoot (maximum error after first crossing of setpoint), undershoot (maximum error after second crossing of setpoint) and settling time (time settle out within a specified band around the setpoint). The ultimate limit for rise time is proportional to the loop dead time. The ultimate limit for overshoot and settling time is theoretically zero. The practical limit to rise time is similar to the practical limit for peak error for fast tuning settings but degrades to the relationship for the integrated error for sluggish tuning settings. Fortunately, there are many features that can be used to readily help achieve the ultimate limit to the rise time. The practical limits for overshoot and settling time depend upon a balance between the contributions from the integral and proportional modes. In general, the controller gain for maximum disturbance rejection can be used to minimize rise time, and the integral time can be increased to minimize overshoot and settling time.

The minimum rise time (T_r) can be approximated as the change in percent setpoint ($\Delta\%SP$) divided by the maximum rate of change of the process variable. For an integrating or "near

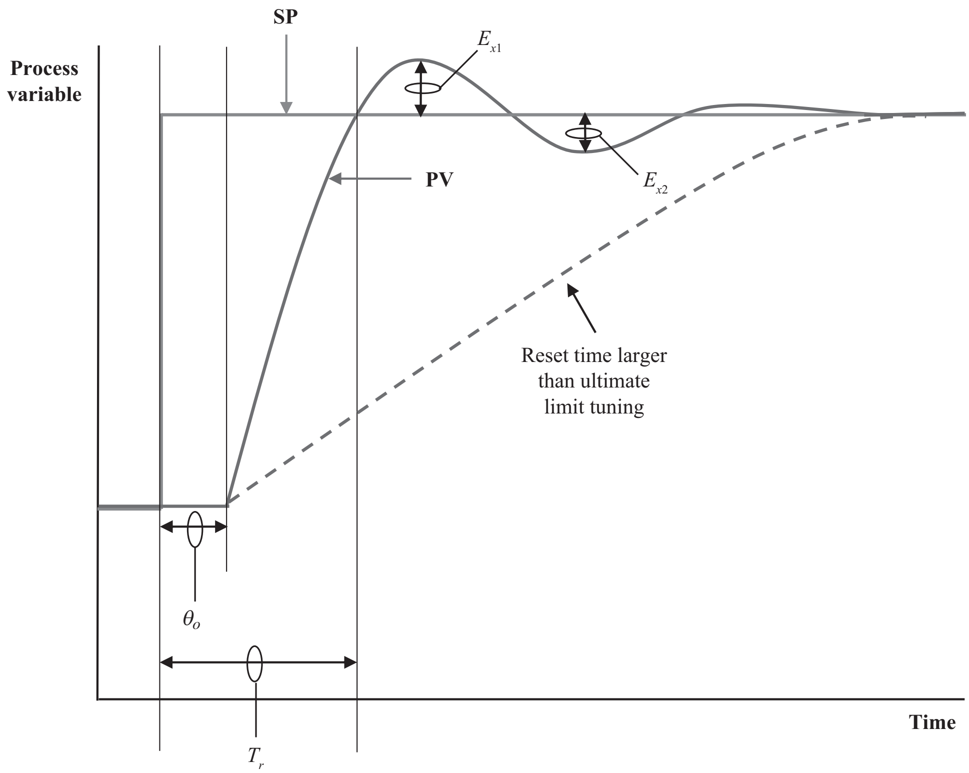


Figure C.2. Aggressive settings for load rejection minimizes rise time (T_r) but causes overshoot (E_{x1}) and undershoot (E_{x2}) in setpoint response.

integrating" process, the maximum PV ramp rate is the integrating process (K_i) gain multiplied by the change in controller output as detailed in the denominator of Equation C.15a. If the step change in controller output from the proportional mode for a structure of proportional action on error is less than the maximum available output change (difference between current output and output limit), Equation C.15a simplifies to Equation C.15b for feedback control. The output change must be corrected for methods used to make the setpoint response faster. For setpoint feedforward, the step change in output is a combination of the feedforward and feedback action. For smart bang-bang logic, the step output change is the maximum available output change.

$$T_r = \frac{\Delta\%SP}{K_i * \min(|\Delta\%CO_{\max}|, (K_c + K_{ff}) * \Delta\%SP)} + \theta_o \tag{C.15a}$$

For a maximum available output change larger than the step from the proportional mode ($|\Delta\%CO_{\max}| > K_c * \Delta\%SP$), the change in setpoint in the numerator and denominator cancel out yielding a simpler equation:

$$T_r = \frac{1}{(K_i * K_c)} + \theta_o \tag{C.15b}$$

For the *near-integrating* process response seen in vessel and column temperature loops where the process time constant is significantly larger than the total loop dead time, the integrating gain is the open loop gain (K_o) divided by the open loop time constant (τ_o) yielding Equation C.15c.

$$T_r = \frac{\tau_o}{(K_o * K_c)} + \theta_o \tag{C.15c}$$

The practical and ultimate limit to loop performance can be reconciled by realizing there is an implied dead time (θ_i) from the tuning. Equation C.16 shows the implied dead time that can be approximated as the original dead time (θ_o) multiplied by a factor that is 0.5 plus Lambda (λ). Lambda is the closed loop time constant for a setpoint change. For a PID tuned for maximum disturbance rejection, Lambda is set equal to the original dead time. The implied dead time is then equal to the original dead time.

$$\theta_i = 0.5 * (\lambda + \theta_o) \tag{C.16}$$

The peak and integrated errors for unmeasured step disturbances represents the worst case. Step disturbances originate from manual actions, safety, switches, and sequential operations. If discrete actions (e.g., the opening and closing of on-off valves and the starting and stopping of pumps) are replaced by control loops with modulated final control elements (throttling valves and variable speed drives) or are attenuated by intervening volumes, the step disturbances are smoothed. The attenuated load disturbance has a time constant (τ_L) that is the residence time of the volume or closed loop time constant of the upstream control loop. To include the effect of a load time constant, the process excursion in the first dead time, which is the key time for determining minimum peak error, can be computed by Equation C.17. The open loop error (E_o) in the equations for peak and integrated error can be replaced with a lagged load disturbance (E_L) that is the open loop error multiplied by the exponential response of the disturbance in one dead time The effect is mitigated by a reset time that is slow relative to the disturbance time constant.

$$E_L = (1 - e^{-\theta_o / \tau_L}) * E_o \tag{C.17}$$

PID controllers tuned too fast can introduce process variability from an oscillatory response. PID controllers tuned too slowly can make a loop with good dynamics perform as badly as a loop with poor dynamics. In other words, money invested to reduce process dead time or to get faster measurements and valves is wasted unless the PID controller tuning is commensurate with the speed of the process so that the practical limit approaches the ultimate limit to loop performance.

Nomenclature:

$\%CO_{t1}$ = controller output at time t1 before correction for load disturbance (%)
$\%CO_{t2}$ = controller output at time t2 before correction for load disturbance (%)
$\Delta\%CO$ = change in controller output converted to % of controller output scale (%)
$\Delta\%CO_{\max}$ = maximum available change in controller output (%)
$\Delta\%PV$ = change in process variable converted to % of controller input scale (%)
$\Delta\%PV_{\max}$ = maximum change in process variable in one dead time interval (%)
$\Delta\%SP$ = change setpoint converted to % of controller input scale (%)
$\%E_i$ = integrated error from load disturbance (% * sec)
E_L = open loop error corrected for load disturbance time constant (%)

E_o = open loop error (load disturbance error with controller in manual) (e.u.)
$\%E_t$ = error at time t from load disturbance converted to % of controller input scale (%)
$\%E_{t1}$ = error at time t1 from load disturbance converted to % of controller input scale (%)
$\%E_{t2}$ = error at time t2 from load disturbance converted to % of controller input scale (%)
E_x = peak error from load disturbance (e.u.)
K_c = controller gain (dimensionless)
K_i = integrating process gain (% per sec per % = 1/sec)
K_o = open loop gain (%/%) (dimensionless)
K_p = process gain (e.u./flow)
T_i = integral time (reset time) (sec)
T_r = rise time (time to first reach setpoint) (sec)
T_u = ultimate period (sec)
λ= closed loop time constant (sec)
θ_i = implied total loop dead time (sec)
θ_o = original total loop dead time (sec)
τ_o = open loop time constant in self-regulating process (sec) (negative feedback)
τ_f = signal filter or attenuating volume time constant (sec)
τ_L = load disturbance time constant (sec)
τ_p = primary process time constant (sec)
Δt_x = controller execution time (sec)

APPENDIX D

Discussion

The following are questions and answers posted on the Emerson Exchange 365 Mentoring Discussion website as part of the ISA Mentor program.

DANACA JORDAN'S QUESTION ON TUNING USUALLY INACTIVE LOOPS

I am trying to improve the control of a differential pressure controller (PDIC) that is only turned on when things go wrong on a particular product, so it is only active a couple unplanned hours at a time. Any suggestions for using limited historian or Insight data to improve tuning?

GREG MCMILLAN'S ANSWER

You can estimate the dead time from the time it takes the proportional-integral-derivative controller (PID) process variable (PV) to change direction after a change in PID output direction when the PDIC is active. For a large change in PID output, you can approximate an integrating process gain as the new rate of change minus the old rate of change (including sign) in %/sec divided by the % change in PID output yielding a gain in (%/sec)/% or 1/sec. You can then use Lambda tuning rules for integrating processes. I suggest a Lambda equal to 3× the dead time to help account for nonlinearities. Even if you cannot estimate the integrating process gain, you can check the reset setting based on just knowing the dead time. Most pressure and differential pressure loops have too much integral action (too small a reset time in seconds or too large a reset setting in repeats/minute).

I suggest you try using the auto tuner relay oscillation method for a quick better estimate of dynamics. You can set the change in output to be just 5× the valve dead band, check the integrating process option, and go for just two cycles. If the valve dead band is 0.5 percent (typical for a reasonably good valve) the amplitude in PID output for the relay method can be as small as 2.5 percent. The relay method should be able to keep the PV near the setpoint for small load upsets while you are running the test but the estimates will be affected. You can manually check the integrating process tuning settings using a lambda equal to the 3× the dead time.

The assumption here is that you have a process where the primary time constant is more than 4× the dead time making it a near-integrating process. If this is not the case, the reset time will be too large but this will be in the conservative direction.

HECTOR TORRES'S QUESTION ON RESOLUTION LIMITS

In different articles and books you will normally talk about resolution limits; I am not sure if I totally understand the concept and all it comprehends. What should we understand by such term; how will I identify I am experiencing such? What kind of control issues will I have? How can I solve a resolution limit problem?

GREG MCMILLAN'S ANSWER

A resolution limit is the smallest input change that will cause an output change. When the input change exceeds this limit, the output change is an integer multiple of the resolution limit. Thus, the output response is a series of steps with a size that is an integer multiple of the resolution limit. In contrast, when a change in input exceeds the threshold sensitivity limit, the change in output equals the change in input. The result is a step size that matches the input change. Resolution limits come from digital devices. However, the ISA standard for valve response testing expresses stick-slip as a resolution limit when really a threshold sensitivity limit is more at play where slip is about equal to the stick.

The 1980s vintage Distributed Control System (DCS) had only 12 bits in the input cards. Since one bit is for sign, there were only 11 bits for resolution. If you divide 100 percent by 2 to the 11th power, you get a resolution limit of 0.05 percent which doesn't seem bad until you consider the practice at the time was to use wide range thermocouple input cards rather than transmitters to save on field hardware costs. These cards caused a series of steps of about 0.4°F that caused such severe bumps in the PID output from rate action that derivative could not be used on reactors where derivative is most needed to compensate for secondary time constants from thermal lags. Today's DCS has 16 bit input and output cards.

The standard signal input card for Variable Speed Drives (VSD), the last time I checked (in 2008), has only eight bits which if there is one sign bit translates to a resolution of 0.78 percent which is really lousy when you consider the resolution limit of a sliding stem valve, digital valve controller, and diaphragm actuator is close to 0.1 percent. Users need to specify a VSD input card with at least 12 bits. Why this is a special card shows the lack of understanding of resolution limits.

When a step change in the PV occurs, the derivative mode thinks the entire change occurred in one PID execution. Exothermic liquid reactors have a slow temperature response and use a PID gain greater than 10 and a rate time greater than 120 seconds to prevent a runaway. Since the step from a resolution limit is divided by the PID execution time and multiplied by the product of the PID gain and rate time, a 0.4°F step can cause a full scale bump in the PID output.

Resolution and threshold sensitivity limits both cause limit cycles if there is any integral action in the process or the controller. The amplitude depends upon the process gain. For pH control on the steep part of a titration curve the process gain is so large, the limit cycle can violate pH limits. Such pH processes are very sensitive to valve resolution and threshold sensitivity limits.

Most sensors and analyzers have a threshold sensitivity limit larger than the resolution limit in today's smart transmitters. The wireless default trigger level is really a threshold sensitivity limit. The default update rate helps prevent a limit cycle if the trigger level is set too large.

The best solution is to use transmitters and valves with the best resolution and threshold sensitivity. Limit cycles and bumps in the PID output from rate action can be eliminated by the use of enhanced PID with external reset feedback.

HECTOR TORRES'S QUESTION ON SETPOINT LEAD–LAG

One of the options you recommend for eliminating overshoot at setpoint changes is using a setpoint lead-lag setup. I cannot visualize the concept in my mind. In a Two Degrees of Freedom (2DOF) PID Structure, you will have an idea of what beta and gamma do and where they are in the execution path.

For the Setpoint Lead-Lag setup I do not comprehend where the lead and where the lag act; in other words, what is led and what is lagged?

GREG MCMILLAN'S ANSWER

To get a setpoint Lead-Lag, the PID setpoint would be passed through a Lead-Lag function block. A Lead-Lag block is normally used for dynamic compensation of feedforward signals. For setpoint changes, we are using this block to moderate the step in the PID output from the proportional mode and the bump from the derivative mode and to prevent immediately driving the PID output past its final resting value. The tuning settings that reduce the peak and integrated errors from unmeasured step disturbances to the process input maximize the step and bump in the PID output for setpoint changes making the approach to setpoint faster but causing excessive setpoint overshoot. Preventing overshoot is important in many applications and reducing abrupt movements in the PID output that could upset other loops and systems is important in gas volumes and utility headers.

In the frequency domain, the lead is a zero and the lag is a pole. The Laplace transform shows that the numerator term for the zero has the same form as the denominator term for the pole. If the lead time equals the lag time, the numerator (zero) and denominator (pole) cancel out and the output equals the input. If the lead time is zero, the output is simply the filtered input where the lag time is the filter time. If the lead time is nonzero but less than the lag time, there is a small step in the block output followed by the exponential approach to the final value of the input set by the lag (filter) time. If the lead time is greater than the lag time, there is a spike in the output that is greater than the input followed by an exponential decay to the final value of the input set by the lag (filter) time.

If the lead time is zero and the lag time is set equal to the PID reset time, the response of the PID should be similar to a PD on PV and I on Error structure or a 2DOF structure with beta and gamma both equal to zero. This will prevent overshoot and provide a smooth and gradual change for both the PID output and PV. The time to reach setpoint (rise time) is much longer.

For bioreactors, gas furnaces and reactors, and plug flow systems (e.g., pipes, static mixers, extruders, sheet lines) a smooth approach with no overshoot is much more important than rise time. Bioreactor cells are very sensitive to abrupt changes and to a temperature or pH higher than optimum. Gas and plug flow volumes have no primary process time constant to filter out abrupt changes in PID output and reach a final value relatively quickly without having to make an abrupt change in the PID output.

For large well mixed liquid volumes, the increase in rise time for setpoint changes in startup, transitions, and batch operations can appreciably reduce process capacity. The response of these volumes for continuous operations is near-integrating and for batch operations is essentially true integrating. For both types of operations on these volumes, lambda integrating process tuning rules are used. The inherent approach to setpoint is very slow due to a large primary process time constant or slow integrating process gain making it necessary to have a large sudden change in the PID output to get to significantly different setpoint in minutes rather than hours.

If the lead time is set equal to 25 percent of the lag time, and the lag time is set equal to the reset time, you get a response with a good compromise between setpoint overshoot and rise time. Similar results can be obtained by a 2DOF structure with the beta equal to 0.5 and the gamma = 0.25. However, in my test results, the 2DOF structure with any gamma setting > than zero resulted in some short term higher frequency oscillations in the PID output whereas with the Lead-Lag there was simply one bump. The 2DOF oscillations disappear quickly and may not exist for less aggressive tuning than what I used.

Since adding a Lead-Lag function block to all setpoint changes requires more effort than simply changing the PID structure to 2DOF, the 2DOF option may be better. Since you can achieve many of the other PID structures that have integral action by a simple change of the beta and gamma, one could say for flexibility, the 2DOF structure may be the future way to go.

You can go to DeltaV Books Online to find out more about the Lead–Lag block and PID structures.

HECTOR TORRES'S QUESTION ON MEASUREMENT LAG

How important is it to determine the lag caused by my measurement system?

At a bump test I will note the associated delay time and the change in my process variable; any lag in my measurement will already be acting on the observed dynamics. Since the computed tuning parameters will be based on these, they will be already accounting for such measurement lag; will it make a difference determining how much is coming from such source?

When will it be important to determine the measurement lag?

GREG MCMILLAN'S ANSWER

Excellent question that really made me think. The motivation to identify the size of the measurement lag is in the knowledge whether the lag is causing a significant deception in the view or delay in the correction of process variability. Excessive lag in many cases can be fixed by an adjustment, replacement of the sensor, or redesign of the installation. The solution is often within the responsibility and capability of the automation engineer. Since the measurement is the window for analyzing and controlling the process, your knowledge and solution can be essential.

Deception occurs when the measurement lag is greater than the ultimate period of the loop. The large measurement lag acts as a filter attenuating process variability. The measurement amplitude of an oscillation is less than the actual amplitude. Doing the wrong thing (i.e., making the measurement lag larger), makes the trend chart look better in terms of less variability. If the measurement lag becomes the largest time constant in the loop, an increase in the lag enables an increase in the PID gain furthering the deception that the measurement lag is beneficial. Normally, the reset time cannot be decreased and may in fact need to be increased to stop a reset cycle. So the clue may be a slower more oscillatory response. The ratio of the measured to actual process variable amplitude is the period of the oscillation divided by 6.28 times the lag. This simple relationship enables one to estimate the degree of deception and the actual process variable amplitude from the observed amplitude for a given oscillation.

A noticeable effect of measurement delay occurs when the measurement lag is more than 10 percent of the PID reset time. The effect is seen in terms of an increase in the measured peak and integrated error for a load disturbance and the rise time for a setpoint change. The effect becomes intolerable if the additional delay from the lag increases the total loop dead time to the point where the PID must be re-tuned. This occurs when the additional dead time from measurement lag increases the total loop dead time to be greater than the implied dead time based on the tuning. The result is often an oscillatory response from excessive integral action. The implied dead time is half and one-fourth the sum of lambda plus the original loop dead time (dead time with no measurement lag) for self-regulating and near or true integrating processes, respectively. If the measurement lag is smaller than the process time constant, the additional dead time comes from the larger measurement lag itself. If the measurement lag is greater than the process time constant, the additional dead time comes indirectly from a greater fraction of other time constants becoming effectively dead time. A clue as to whether the measurement lag has increased is a process variable that oscillates with a larger period. Due to attenuation or deception, the amplitude may look less.

If the measurement lag becomes the largest time constant in a loop for an integrating process or even worse for a runaway process (e.g., highly exothermic reaction), the deterioration is huge. Setting the PID rate time equal to a large measurement lag is important. For runaway processes a large measurement lag can make the maximum allowable PID gain approach the minimum allowable PID gain needed to prevent the process from accelerating and reaching a point of no return. The window of allowable PID gains can close in a runaway process due to a large measurement lag (a serious safety risk).

How do you identify the measurement lag? How do you know that the primary or secondary time constant identified by software is in the measurement, process, or final control element (e.g., control valve or variable frequency drive)? There are few easy cases. The PID process variable filter time and transmitter damping should be checked and reduced if they are causing deception or deterioration as per above guidelines. The settings should be minimized if the settings are larger than 10 percent of the reset time. There is a chicken and egg possibility, in that the reset time may be large due to a large measurement lag.

If a step change in flow can be made to a process input, the open loop response of the measure variable will exclude the effect of the final control element lag from its rate limited exponential response and inherent time constant (another story). Also, the time constant for a small sliding stem valve or a variable frequency drive with insignificant rate limiting in the drive setup is negligible for small changes (e.g., <2 percent) in PID output. This still leaves us with the quandary of how to know if the primary or secondary time constant identified is in the process or the measurement.

For fast process where there is negligible back mixing, the process time constant is small. The process time constant is less than one second for gas flow reactor and furnace composition and temperature control by manipulation of feeds or fuel, liquid or polymer pressure and flow control, and inline pH control. A primary or secondary time constant much larger than one second is most likely a measurement lag. The electrode time constant is about two to six seconds for a new clean electrode and reasonable velocity (e.g., >5 fps). An aged or coated glass electrode can increase the measurement lag from a few seconds to minutes, even hours. A loosely fitted temperature sensor in a thermowell or even worse a thermocouple in a ceramic protection tube will cause the measurement lag to increase from seconds to minutes due to the air gap or ceramic tube acting as an insulator.

Transmitters and electrodes can be tested in the shop to determine the measurement lag excluding any installation effects. The transmitters and electrodes should not be cleaned, which would eliminate the effect of coatings. Electrodes are inserted into buffer solutions or even better process samples at operating conditions. Temperature sensors still in their thermowells can be inserted into temperature baths. Velocity affects the electrode and thermowell response, so a stirrer should give about the same velocity as in the process installation. The measurement lag is half the 86 percent response time assuming the measurement delay is negligible. Waiting for a more complete response (e.g., 95 percent response time) especially for pH electrodes causes less relevant and more inconsistent results.

APPENDIX E

Enhanced PID for Wireless and Analyzer Applications

INTRODUCTION AND OVERVIEW

It is widely recognized that wireless measurements considerably reduce installation and maintenance costs. However, it is less recognized that the portability of wireless offers the ability to optimize the measurement location for the fastest most representative process response and to demonstrate/prototype process control improvement for finding and justifying permanent solutions.

The wireless update time termed *default update rate* (refresh time) is the time interval for periodic reporting. The wireless update sensitivity termed *trigger level* is the minimum change in measurement value for exception reporting. To save power, the transmitter goes to sleep and wakes up periodically to check if the change in sensor value from the last value transmitted is larger than the trigger level. When the change exceeds the trigger level or the time since the last communication exceeds the default update rate, then the value is transmitted. The time interval between periodic checks of the sensor is the *triggered update rate* (wakeup time). If only periodic reporting is available, this time interval is the default update rate. Increases in the default update rate and trigger level settings reduce the number of transmissions, which increases battery life. The enhanced proportional-integral-derivative controller (PID) eliminates the ramps, limit cycles, and spikes from large values of these settings facilitating an increase in battery life.

For slow processes where a significant process response takes minutes or hours to develop, such as composition, level, and temperature in large volumes, wireless measurements are fast enough for closed loop control using traditional PID. The primary question for these applications is what is the potential detrimental effect of a loss of update and poor sensitivity? The enhanced PID protects against communication, sensor, valve, and analyzer failures and eliminates the cycling from wireless exception reporting, sensor, valve, or analyzer sensitivity limits for these and other loops. Thus, the enhanced PID improves the reliability and reduces the variability for all loops with and without wireless measurements. Additionally the enhanced PID extends the applicability of wireless measurements and analyzers to processes with a response time faster than the update time.

While the enhanced PID offers advantages for all loops, wireless measurements should not be used where the process can cause product degradation, shutdowns, equipment damage, or an unsafe condition faster than the triggered update rate (exception reporting) or default update rate (periodic reporting) of the wireless transmitter. Examples of loops that can get into trouble too fast for even the fastest wireless update time are sheet, compressor, and turbine speed control and incinerator and pipeline pressure control. Slow wireless default update rates, such as 30 and 60 seconds, can be a problem for vessel or header pressure control since the pressure may approach alarm, trip, or relief settings within the default update rate.

In a traditional PID, the integral and derivative modes are computed each execution of the PID block. The PID algorithm uses the execution time in the integral and derivative mode calculations. Thus, the reset and derivative contribution calculated by traditional PID may not be appropriate when used with a wireless measurement where the default update rate is significant compared to the process response time. In such cases, the traditional PID will ramp the controller output through continual integral action even though the actual measurement has not changed. The PID is acting on old information. The detrimental effect of integral mode acting on old information is greatest if the process response has largely responded before the next update, which occurs when the process response is faster than the update time or the measurement value has not been updated because of device or communication failure. The update time of wireless measurements on fast processes, such as flow and pipeline temperature and pH loops, and the update time for analyzers in many applications are slower than the process response. An update failure occurs for loss of transmission, sensor failure, a stuck valve, and a sample or multiplexer system failure. Furthermore, when the wireless measurement is communicated using exception reporting, the ramp of the controller output between updates due to a sensitivity limit results in a perpetual equal amplitude oscillation called a limit cycle even though there are no load disturbances.

When there is an update, the traditional PID considers the entire change in the measurement value occurred within the PID execution time for the derivative mode calculation. The result when the PID is executed faster than the wireless default update rate is a spike in controller output whose size increases with the movement of the process between updates, which is greatest for a fast process or for recovery from an update failure.

The ramp, limit cycle, and spike from a traditional PID inflict a disturbance upon the loop and possibly other loops. Thus, a traditional PID in automatic mode can cause process variability even if there are no disturbances.

The enhanced PID computes the integral and derivative mode contributions to the controller output when there is a measurement update and uses the elapsed time between updates in its calculations. Thus, the enhanced PID only acts on new information and considers the observed change in the measurement to have occurred not in just the last PID execution time but over the elapsed time. The key functional features of the enhanced PID are:

- PID responds immediately to setpoint, feedforward, and manual or remote output changes
- PID responds immediately to process variable changes
- PID action is suspended if there are no changes (integral action stops)
- PID action uses elapsed time instead of PID execution time in derivative calculation
- PID action computes positive feedback filter exponential response using elapsed time

If the dead time from a discontinuous signal whether a digital device or at-line or online analyzer is much greater than the open loop time constant, the enhanced PID will see the full process response in the next update. For these cases, the controller gain can be set equal to the inverse of the controller gain to provide considerably tighter control. Variation in the update time does not affect the tuning. Consequently, the enhanced PID is particularly valuable for closed loop control using off-line analyzers where the time between lab results is large and extremely variable.

When there is a loss of communication with the measurement or final control element (e.g., valve, damper, or variable speed drive), the enhanced PID provides no further reset action. The enhanced PID waits for new information whereas the traditional PID output ramps to an output limit. When communication is restored, the enhanced PID acts on the new information only. The traditional PID sees the effect of the ramp to the output limit that occurred during the loss of communication or loss of valve travel. Additionally, the traditional PID considers the entire observed change to have occurred within the last execution of the PID instead of the time duration of the failure. The result is a bump from gain action and spike from rate action in the traditional PID output.

The enhanced PID executes as fast as a traditional PID independent of the wireless default update rate or analyzer cycle and sample time. As a result, enhanced PID is ready to immediately act on changes in setpoint, mode, remote output, feedforward, and tuning. For a proportional gain that is the inverse of the process gain, the setpoint response is particularly impressive in that the controller output goes to the value needed to reach setpoint within the PID execution time. For a fast loop, the process is already at the setpoint before the next update.

Secondary loops are designed to be faster than primary loops for cascade control. If a wireless measurement is used in a secondary loop, the enhanced PID can prevent the secondary loop response from becoming too slow. The fast load response of the enhanced PID helps the secondary loop reject secondary disturbances. The fast setpoint response of the enhanced PID helps the primary loop reach its setpoint faster and reject primary loop disturbances.

Analyzers with sample systems offer a higher level of control. Ultimately what you want to know and control is the composition in a process stream. However, the use of analyzers for closed loop control is rather limited due to the long update time, excessive interruptions and extraneous values, and low sensitivity. Analyzer sensitivity is generally not as good as flow, pH, pressure, and temperature measurement sensitivity. The analyzer sensitivity limit is the result of interferences, analysis method, and sensor sensitivity. Analyzer update time, poor sensitivity, and noise results in a step response that precludes the use rate action even if the process time constant is large. The enhanced PID eliminates most of these concerns and enables more analyzers to go on closed loop control.

If there is a resolution instead of a sensitivity limit, where the response is quantized so that the increment or decrement in the measurement is an integer multiple of the resolution limit, the steps in the response never put the measurement at the setpoint. Analyzers with sample systems frequently have resolution limits due to analysis methodology and minimum dilution and reagent dose sizes. The enhanced PID will eliminate the cycling from reset action and the spikes from rate action for both resolution and sensitivity limits in the measurement. The enhanced PID can also eliminate the limit cycles from valve stick-slip and backlash (valve sensitivity limits).

The enhanced PID is available starting in DeltaV v11. The enhanced PID option can be enabled under the FRSIPID_OPTS for the PID block in control studio (Use PIDPlus) as shown in Figure E.1. This PIDPlus option turns on the Dynamic Reset Limit option.

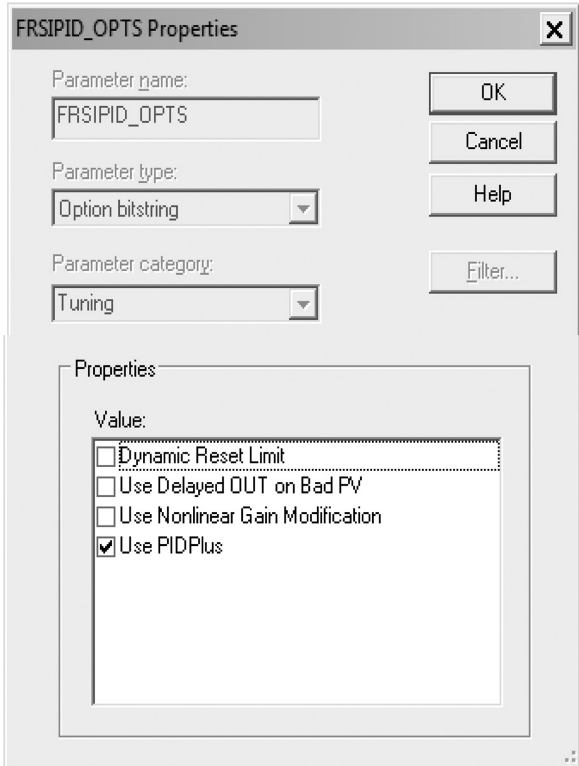


Figure E.1. Enhanced PID enabled in FRSIPID_OPTS for PID block.

TARGET APPLICATIONS

While the enhanced PID can be used in any application to reduce process variability and provide protection against measurement and valve failures, the greatest benefits are seen for loops where:

- Update time is larger than the process response time
- Update time is significant and variable
- Update sensitivity limit is larger than the desired process precision
- Update failures can occur (e.g., communication and electrode failures)
- Measurement resolution or sensitivity is poor causing limit cycles
- Valve exhibits significant stick-slip and backlash causing limit cycles
- Variable speed drive setup has significant deadband causing limit cycles

Since the enhanced PID takes into account the effect of update time, the benefit is greatest where the wireless or analyzer update time is much larger than the process response time (defined here as the process dead time plus time constant). The improvement in control loop response can be significant but diminishes as the update time decreases and approaches the process response time. For an update time that is less than the process response time, the larger benefit of enhanced PID is the protection against loss of communication or device failure or elimination of limit cycles from measurement and valve sensitivity limits.

Some examples of mainstream applications where enhanced PID provide greatest improvement in closed loop control are:

- Wireless flow control
- Wireless static mixer pH control
- Wireless desuperheater temperature control
- Pulp and paper stock consistency control
- Cross directional control of sheet thickness
- At-line analyzer control of pipeline composition
- At-line analyzer control of plug flow reactor product composition
- Off-line analyzer control of almost any unit operation

Examples of mainstream applications with less but still significant improvement in closed loop control are:

- Wireless vessel pressure control
- Wireless extruder temperature control
- Wireless jacket temperature control
- Wireless heat exchanger temperature control

TEST SETUP

A DeltaV v11 module with one second execution time was set up with two PID blocks in a virtual plant as shown in Figure E.2. The first PID block (PID1) had the enhanced PID option

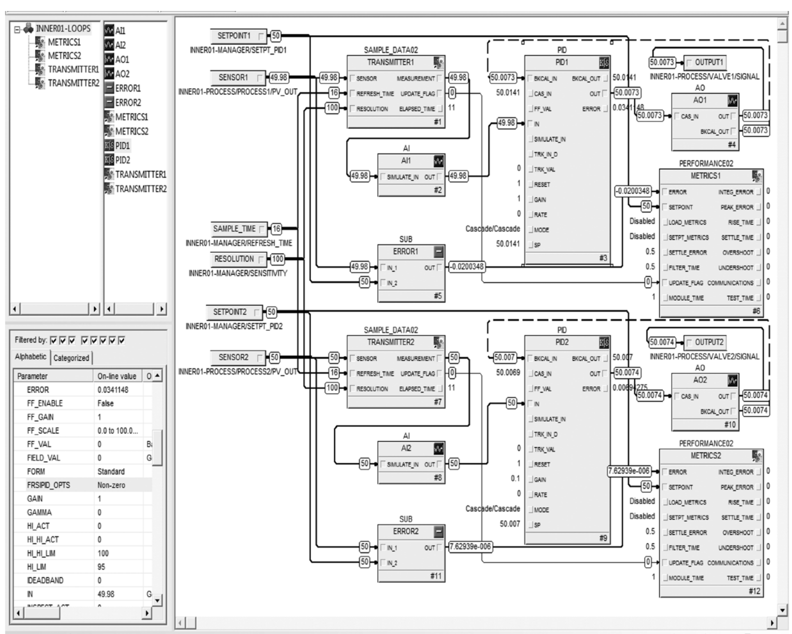


Figure E.2. Virtual plant test setup for comparison of enhanced and traditional PID.

enabled in the FRSIPID_OPTS parameter whereas the second PID block (PID2) did not. The PID analog inputs were provided by SAMPLE_DATA02 composite blocks that simulated the update time and sensitivity limits for wireless transmitters and analyzers (TRANSMITTER1 for PID1 and TRANSMITTER2 for PID2). A composite block (PERFORMANCE02) for performance metrics tallies the number of communications and computes the integrated absolute error (IAE) and peak error for load disturbances and the rise time, settling time, overshoot, and undershoot for setpoint changes

Many of the wireless devices on the market today support default update rates of 8, 16, 30, and 60 seconds. However, to maximize battery life it is assumed that the longest reasonable default update rate will be selected. The default update rates selected for the tests were 16 and 30 seconds for the fastest loops and 30 and 60 seconds for the slowest loops.

Lambda tuning was used for both PID controllers. Rate action was not used. In all of the cases studied except vessel temperature control, the PID2 gain had to be decreased to prevent excessive oscillations.

The output of each PID controller was the signal to an actuator composite block (ACTUATOR1 for PID1 and ACTUATOR2 for PID2) on a linear valve in a process module with a 0.2 seconds execution time. The actuator time response consisted of two one second time constants in series. The effect of actuator-valve backlash and stick-slip was not part of study, so the deadband and resolution of the actuator was set to 0.02 and 0.01 percent, respectively. The actuator block simulated the stroke and computed valve travel. The output of the valve for each loop was an input to a composite template block that simulated the dynamic response of a self-regulating and integrating process. The output of each process block is connected to the respective sensor of the transmitter block (PROCESS1 output is TRANSMITTER1 sensor input and PROCESS2 output is TRANSMITTER2 sensor input).

The setup of process dynamics and update time and sensitivity limit, the computation of tuning settings, and automation of tests for load steps, setpoint steps, and measurement failure were supplied by a test manager module with an execution time of two seconds. The manager module also turned on the performance metrics and the valve travel accumulation during the tests for load upsets and setpoint changes.

DISCUSSION OF TEST RESULTS

The variables in the trend charts are as follows:

1. Enhanced PID setpoint
2. Enhanced PID *measured* process variable (update set by wireless device or analyzer)
3. Enhanced PID *actual* process variable (instantaneous update seen by sensor)
4. Enhanced PID output
5. Traditional PID setpoint
6. Traditional PID *measured* process variable (update set by wireless device or analyzer)
7. Traditional PID *actual* process variable (instantaneous update seen by sensor)
8. Traditional PID output

In the trend charts, the actual process variable for each type of PID is denoted as the Sensor PV which is the input to the wireless transmitter.

The size of the step change in load or setpoint for each loop was 10 percent except for vessel pressure where the step size was reduced to 5 percent due to the small vessel size. The actual process response included two one second actuator lags besides the stated process delay (dead time), process lag (time constant), and process gain.

FLOW LOOP

Flow loops are the most common loops in the process industry. The process dynamics are fast with a response time less than two seconds. The use of a wireless measurement with a traditional PID will oscillate severely unless the loop is detuned. The approach taken in the tests was to severely reduce the controller gain approaching an integral-only type of control. The reset time was kept at its original value to help eliminate offsets for ratio control.

The enhanced PID tuning could be made more aggressive than the tuning for a flow loop with a wired measurement. The controller gain could be increased to be the inverse of the process gain. For step setpoint and load changes, this enabled the enhanced PID to make a single almost exact correction (Figures E.3 and E.4). For a setpoint change, the correction occurs within one module execution time. How soon the response fully appears in the actual flow measurement is limited only by the valve response time (two one second lags in series), the process dead time (0.5 seconds), and the process time constant (one second). However, this incredibly fast response is not apparent to the user because the measurement update is much longer. In the trend charts in the figures for the setpoint and load response test results, the actual process measured process variable is plotted to show the actual improvement. The enhanced PID load response is much faster than a traditional PID even though the correction is delayed on the average by about half of the update time after the upset. The exact delay in the recognition of the disturbance depends upon when the disturbance arrives in the update time interval.

For a 16 second pH update time and load upsets, the enhanced PID reduced the IAE by 45 percent (Table E.1). Since the flow loop has a loop dead time that is not significantly less

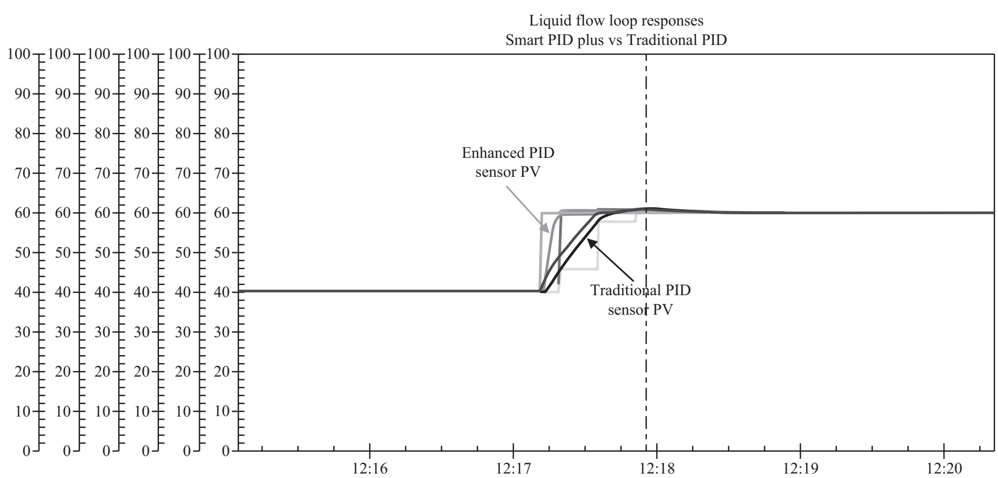


Figure E.3. Flow setpoint response of enhanced and traditional PID (update time = 16 seconds).

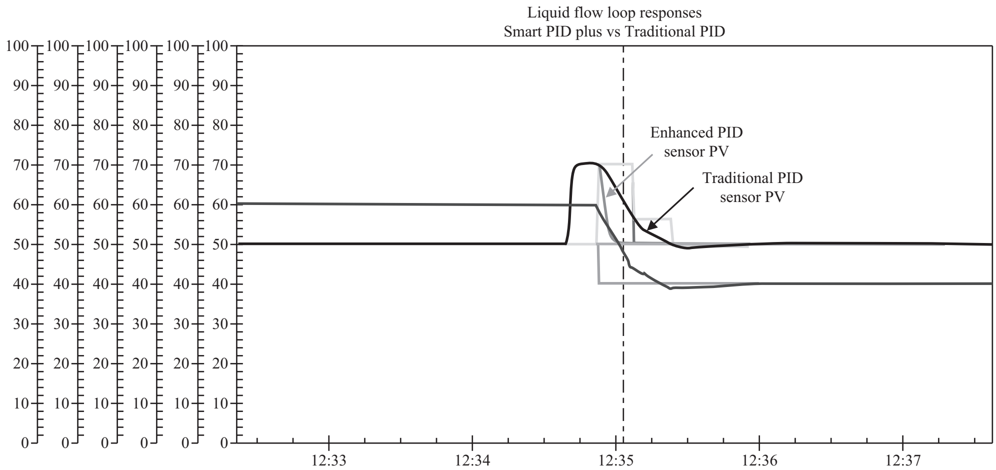


Figure E.4. Flow load response of enhanced and traditional PID (update time = 16 seconds).

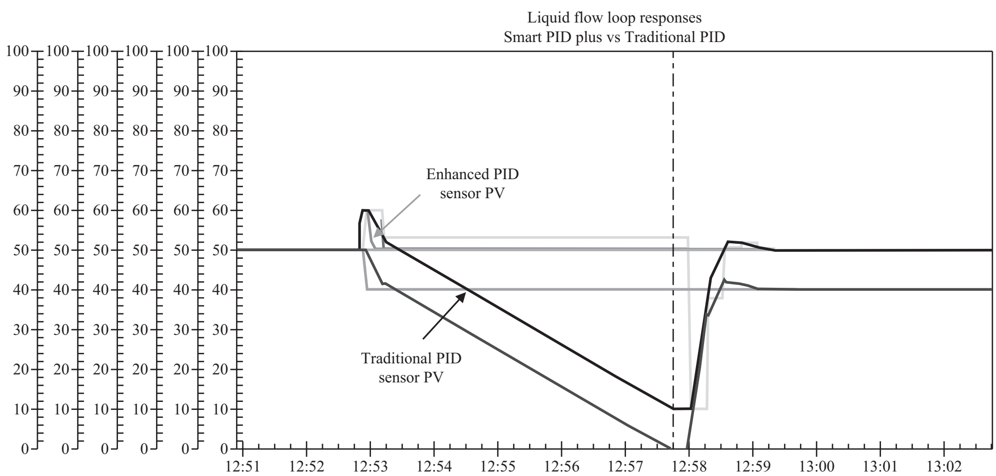


Figure E.5. Flow failure response of enhanced and traditional PID (update time = 16 seconds).

than the process time constant, the peak error of the enhanced PID and traditional loop for a load disturbance is the full open loop error (error with loop in manual). For a 16 second pH update time and setpoint changes, the enhanced PID reduced the rise time (time to reach setpoint) by 69 percent, the settling time (time to settle at setpoint) by 56%, and the overshoot by 55 percent (Table E.2).

If there is an update failure, the traditional PID will ramp towards an output limit. When the measurement is restored, the traditional PID sees the huge self-inflicted upset and overshoots the setpoint in its recovery. The enhanced PID corrects for upsets just before the failure and waits for the next update. The bump to the process is minimal if there are no upsets during the failure (Figure E.5).

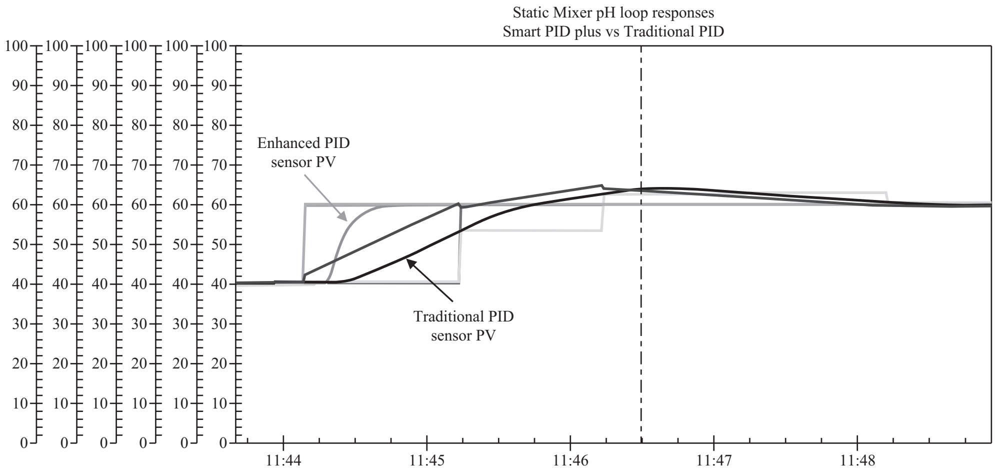


Figure E.6. pH setpoint response of enhanced and traditional PID (update time = 60 seconds).

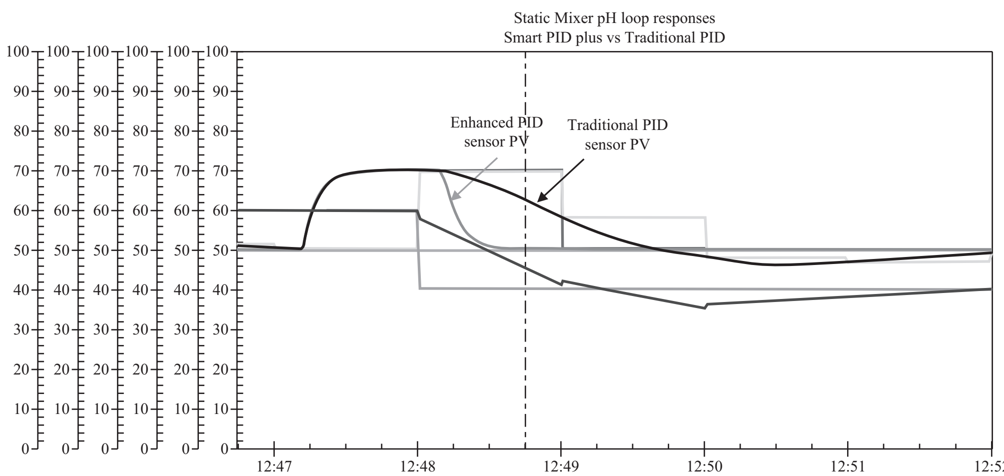


Figure E.7. pH load response of enhanced and traditional PID (update time = 60 seconds).

STATIC MIXER PH LOOP

The enhanced PID dramatically improves the response of cascade loops by enabling an immediate response of the secondary loop to changes in its setpoint. Since the static mixer process dead time (six seconds due to injection delay) and process time constant (six seconds due to electrode lag) is much less than the default update rate (60 seconds), the enhanced PID controller gain can be increased. With both the secondary flow and primary pH loop controller gain equal to the inverse of the process gain, the response to pH setpoint changes and load disturbances is impressive (Figures E.6 and E.7). In contrast, the

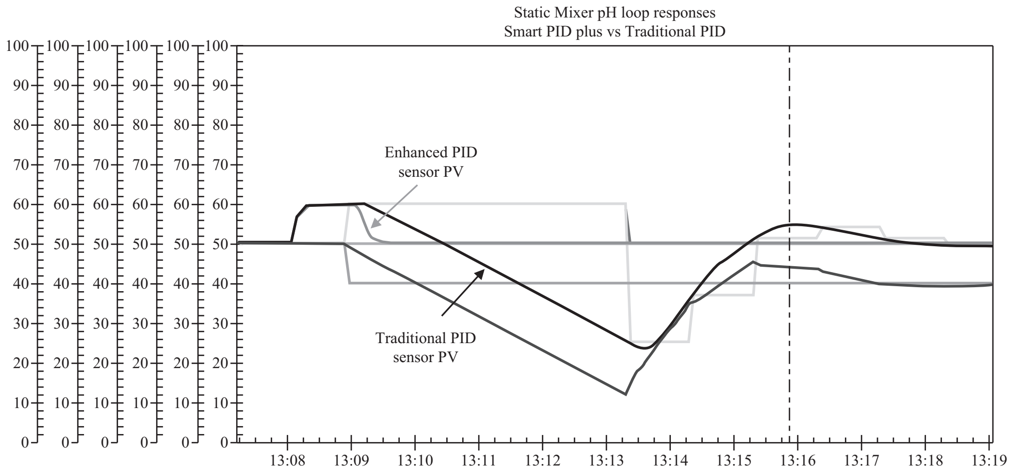


Figure E.8. pH failure response of enhanced and traditional PID (update time = 60 seconds).

traditional PID controller gains for both loops were a factor of ten lower to prevent oscillations. If the default update rate of the secondary loop is not about four times faster than the primary loop, the enhanced PID controller gain must be detuned to stabilize the loops since the oscillations from the secondary loop can be amplified by the primary loop. For a primary pH loop default update rate of 60 seconds, the secondary flow default update rate needs to be 16 seconds or faster.

For a 60 second pH update time and load upsets, the enhanced PID reduced the IAE by 40 percent (Table E.1). Since both the flow and pH loop have a process dead time that is not significantly less than the process time constant, the peak error of the enhanced PID and traditional loop for a load disturbance is the full open loop error (error with loop in manual). For a 60 second pH update time and setpoint changes, the enhanced PID reduced the rise time by 60 percent, the settling time by 84 percent, and eliminated the overshoot that was 40 percent of the setpoint change (Table E.2). For a pH measurement failure, the enhanced PID output remains constant at the last value before the failure whereas the traditional PID output ramps causing a disturbance to the actual process variable (Figure E.8).

VESSEL PRESSURE LOOP

The vessel pressure loop has an integrating response with a short process dead time (one second) and a relatively fast integrating process gain (0.1%/sec/%) due to a small vessel volume or narrow measurement span. Even with a relatively fast default update rate of 16 seconds, control is challenging. The setpoint and load change for this test was 5 percent, whereas for all other tests it was 10 percent.

For a 16 second update time and load upsets, the enhanced PID reduced the IAE by 52 percent and the peak error by 30 percent (Table E.1). The reduction in peak error is important to prevent activation of relief devices and safety instrument systems. For a 16 second update time and setpoint changes, the enhanced PID reduced the rise time by 39 percent and the settling time by 45 percent (Table E.2). However, the overshoot of the enhanced PID was 35 percent larger. Setpoint rate of change limits could be used to reduce the overshoot but this would increase the rise time.

EXTRUDER TEMPERATURE LOOP

For a 30 second update time and load upsets, the enhanced PID reduced the IAE by 18 percent (Table E.1). Since the extruder has a process dead time larger than the process time constant, the peak error of the enhanced PID and traditional loop for a load disturbance is the full open loop error (error with loop in manual). For a 30 second update time and setpoint changes, the enhanced PID reduced the rise time by 19 percent, the settling time by 42 percent, and the overshoot by 44 percent (Table E.2).

An extruder may have some radial mixing but there is negligible axial mixing (back mixing). As a result, the process dead time is larger than the process time constant and process fluctuations with time are not smoothed out and appear in the product. Consequently, the ability of the enhanced PID to eliminate process variability directly translates to improved product quality.

VESSEL TEMPERATURE LOOP

For a 30 second update time and load upsets, the enhanced PID offered no improvement in the IAE or peak error since the process time constant is larger than the default update rate (Table E.1). For a 30 second update time and setpoint changes, the enhanced PID increased the rise time by 19 percent, but reduced the settling time by 27 percent, and decreased the overshoot by 31 percent (Table E.2). The enhanced PID could have been tuned for a faster rise time by increasing the controller gain. The vessel studied was exceptionally small. The performance of temperature loops is excellent on the larger vessel sizes typically seen in industrial plants if noise, sensitivity, and resolution limits are not significant and the controller is tuned with the high gain permitted by the extremely large process time constant. Performance improvement is normally not needed to enable the use of wireless temperature transmitters on large volumes.

POLYMER LINE ANALYZER LOOP

For a 60 second update time and load upsets, the enhanced PID reduced the IAE by 30 percent and the peak error by 25 percent (Table E.1). For a 60 second update time and setpoint changes, the enhanced PID reduced the rise time by 52 percent, the settling time by 74 percent, and the overshoot by 79 percent (Table E.2).

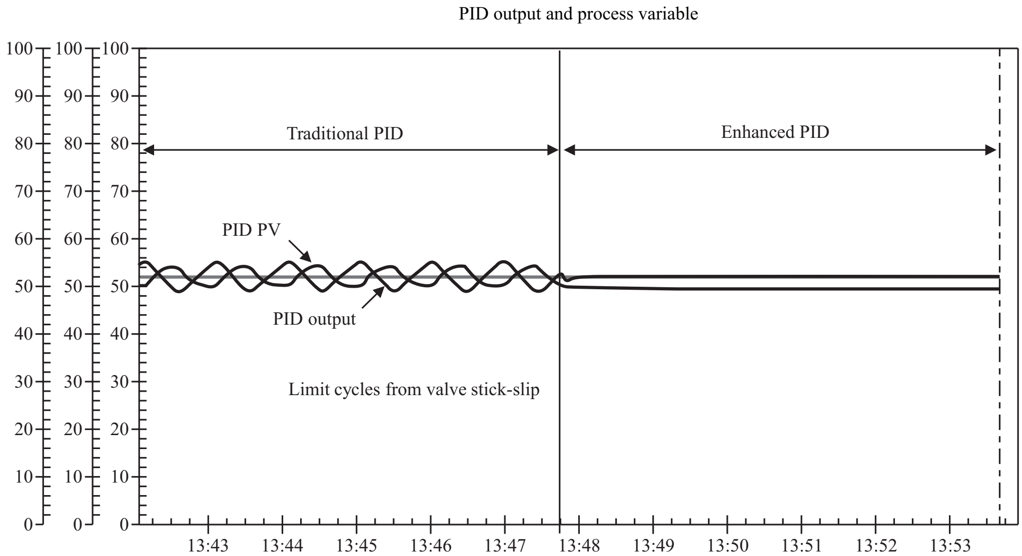


Figure E.9. Enhanced PID eliminates the limit cycle that persists in a traditional PID.

A polymer line analyzer may have some radial mixing but there is negligible axial mixing (back mixing). Thus, the enhanced PID offers an exceptional improvement in polymer and sheet line product quality by reducing variability.

SENSITIVITY LIMIT

In the future, most wireless devices will support exception reporting. When exception reporting is selected then the sensitivity limit is the minimum change that is reported. Once the change exceeds the trigger level (sensitivity limit) or the time since the last communication exceeds the default update rate, the full change and exact change is reported. For a traditional PID, a sensitivity or resolution limit whether due to wireless exception reporting, sensor, valve, or analyzer sample and processing system, will cause a continuous oscillation termed a limit cycle. The amplitude of the square wave in the measurement is the sensitivity or resolution limit. Because the valve or measurement did not respond, traditional PID reset action continues to drive the output even though there is no process disturbance. When there is an update, the jump in the measurement goes beyond the setpoint and the ramp goes in the opposite direction. This ramping of the traditional PID output up and down continues indefinitely. Load disturbances and setpoint changes will temporarily disrupt the pattern. Since the enhanced PID integral mode only makes a calculation when there is an update, there is no ramping of the controller output. If there are no setpoint changes or load disturbances, the measurement draws a straight line. The actual process variable is closer to the setpoint than indicated by the measurement. The benefits are less process variability and less valve travel (longer valve packing life).

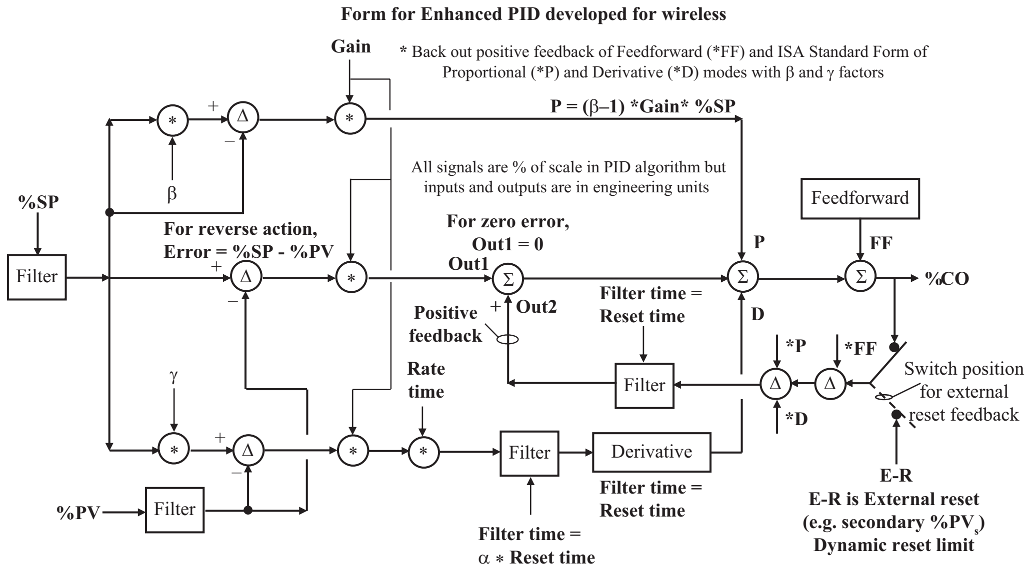


Figure E.10. The enhanced PID uses the positive feedback implementation of the integral mode as shown here for the ISA Standard Form.

The enhanced PID will also inherently eliminate limit cycles from control valve stiction and backlash and variable speed drive resolution limits (Figure E.9). When the final control element is not responding, there is no process variable update unless there is a disturbance and consequently the ramping action of the integral mode is suspended. To ensure measurement noise does not trigger an update, any signal filtering needed should be judiciously done by transmitter damping. The wireless update trigger level should also be increased to screen out noise. To absolutely prevent an unnecessary update that would trigger PID action, a simple small sensitivity limit should be added to the process variable used by the PID block.

BLOCK DIAGRAM

A key part of the enhanced PID uses the positive feedback implementation of the integral mode. When the enhanced PID capability is enabled, the output of the filter block in the positive feedback path of Figure E.10 is computed as a first order exponential response for the elapsed time. The derivative calculation uses elapsed time instead of the module execution time. The enabling of external reset feedback (e.g., dynamic reset limit) by the enhanced PID protects against a burst of oscillations when the primary PID output tries to change faster than a secondary PID, control valve, or variable speed drive can respond. The external reset feedback also allows the use of directional move suppression by the setting of up and down setpoint rate limits without having to retune the enhanced PID. The directional move suppression is useful for providing a slow approach to an optimum and a fast getaway for an abnormal condition and for slowing down movement to the split range point reducing unnecessary crossings of the split range point and the resulting cross neutralization of reagents or cycling of heating and cooling.

SUMMARY OF TEST RESULTS

Table E.1. Summary of load response test metrics

Process type	Process delay (sec)	Process lag (sec)	Process gain	Update times (sec)	IAE PID-Plus, PID (% × min)	Peak error PIDPlus, PID (%)	Value travel PIDPlus, PID (%)	Lambda PIDPlus, PID (sec)	Gain setting PIDPlus, PID	Reset setting PID-Plus, PID (sec)
Liquid flow inner loop 1	0.5 sec	1 sec	1	16 sec 30 sec	6.5, 11.9 9.2, 19.2	20, 20 20, 20	30, 33 30, 65	2, 2	1.0, 0.1	1, 1
Static mixer pH outer loop 1	6 sec	6 sec	1	60 sec	45.3, 75.5	20.0, 19.7	—	8.5, 8.5	1.0, 0.1	6, 6
Vessel pressure loop 2	1 sec	—	0.1 sec^{-1}	16 sec 30 sec	15.3, 32.0 34.8, 53.2	15.1, 21.5 31.5, 32.0	35, 40 82, 79	20, 20	0.8, 0.4	42.5, 42.5
Extruder temperature loop 3	24 sec	6 sec	1	30 sec 60 sec	34,9, 42.5 69.5, 67.9	20, 20 19.6, 18.8	33, 46 28, 60	25.5, 25.5	0.5, 0.1	6, 6
Vessel temperature loop 4	12 sec	60 sec	1	30 sec 60 sec	17.2, 16.6 27.4, 40.6	8.9, 8.0 13.9, 14.3	50, 97 59, 212	13.5, 13.5	2.2, 2.2	60, 60
Polymer line analyzer loop 5	1 sec	10 sec	1	30 sec 60 sec	13.4, 27.0 23.1, 33.0	11.8, 15.7 11.9, 15.9	31, 33 30, 40	2.5, 2.5	1, 0.2	10, 10

Table E.2. Summary of setpoint response test metrics

Process type	Process delay (sec)	Process lag (sec)	Process gain	Update times (sec)	Overshoot PIDPlus, PID (%)	Rise time PIDPlus, PID (sec)	Settling time PIDPlus, PID (%)	Lambda PIDPlus, PID (sec)	Gain setting PIDPlus, PID	Reset setting PIDPlus, PID (sec)
Liquid flow inner loop 1	0.5 sec	1 sec	1	16 sec 30 sec	0.5, 1.1 0.0, 2.3	9, 29 10, 23	25, 57 10, 98	2, 2	1.0, 0.1	1, 1
Static mixer pH outer loop 1	6 sec	6 sec	1	60 sec	0.0, 4.0	38, 94	38, 244	8.5, 8.5	1.0, 0.1	6, 6
Vessel pressure loop 2	1 sec	—	0.1 sec^{-1}	16 sec 30 sec	6.6, 4.9 11.0, 8.2	16, 26 16, 27	86, 156 233, 276	20, 20	0.8, 0.4	42.5, 42.5
Extruder temperature loop 3	24 sec	6 sec	1	30 sec 60 sec	2.2, 3.9 0.0, 7.9	74, 91 181, 85	167, 286 181, 293	25.5, 25.5	0.5, 0.1	6, 6
Vessel temperature loop 4	12 sec	60 sec	1	30 sec 60 sec	6.0, 8.7 7.3, 12.1	50, 42 51, 39	199, 272 276, 288	13.5, 13.5	2.2, 2.2	60, 60
Polymer line analyzer loop 5	1 sec	10 sec	1	30 sec 60 sec	2.5, 0.0 0.4, 1.9	25, 128 39, 81	63, 128 39, 147	2.5, 2.5	1, 0.2	10, 10

APPENDIX F

First Principle Process Relationships

GENERAL IMPLICATIONS

First principle relationships can define process cause and effects that can lead to improved controller tuning and performance by the selection of better tuning rules and process variables for scheduling of tuning settings. It also affects the choice of control valve trim and the feedforward design. Understanding these relationships does not require a degree in chemical engineering but presumes just some understanding of common terms (e.g., heat transfer coefficient and area), relationships (e.g., ideal gas law), and physical concepts (e.g., conservation of mass and energy).

Equations have been developed from first principle relationships for the process gains, dead times, and time constants of volumes with various degrees of mixing. The results show that for well mixed volumes with negligible injection delays, the effect of flow cancels out for the controller gain if one of the following methods is used: Lambda self-regulating rule where Lambda is set equal to the dead time, or the reaction curve method. The effect of flow also cancels out for the reset time besides the controller gain if the process is treated as a "near integrator" and the Lambda integrating tuning rule is used. This is because the flow rate cancels out in the computation of the ratio of process gain to time constant that is the "near integrator" gain. This ratio and "near integrator gain" are inversely proportional to the process holdup mass (e.g., liquid mass). However, for temperature control, the effect of changes in liquid mass cancels out because a change in level increases the heat transfer surface area covered. Several authors have mistakenly tried to schedule controller tuning based on liquid level for reactor temperature control. One author has reported being bewildered by its failure. This is not the case for gas pressure control. The equations show that liquid level has a profound effect on the process integrating gain for vessel pressure control because it changes the vapor space volume without any competing effect. To summarize, the integrator gain for composition and gas pressure is inversely proportional to liquid level (liquid mass). For temperature, the effect of level cancels out unless the level is above or below the heat transfer surface area, which is unusual but can occur at the beginning or end of a batch when coils instead of a jacket are used for heat transfer. For temperature, the integrator gain is nearly always proportional to the overall heat transfer coefficient that is a function of mixing, process composition, and fouling or frosting.

The equations also show that if the transport delay for flow injection is large compared to the time constant, which does occur for reagent injection in dip tubes for pH control, then the controller gain will be proportional to flow. Note that pH control is a class of concentration control.

For the control of temperature and concentration in a pipe, the process dead time and process gain are both inversely proportional to flow and the process time constant is essentially zero, which makes the actuator, sensor, transmitter, or signal filter time lag the largest time constant in the loop. Thus, the largest automation system lag determines the dead time to time constant ratio. For a static mixer, there is some mixing, and the process time constant is inversely proportional to flow but is usually quite small compared to other lags in the loop. The controller gain is generally proportional to flow for both cases.

Finally, the above has implications so far as whether a flow feedforward multiplier or summer and whether a linear or equal percentage trim should be used. A flow feedforward multiplier and equal percentage trim, which both have a gain proportional to flow, can help compensate for a process gain that is inversely proportional to flow provided the process time constant is not also inversely proportional to flow. This is generally the case for temperature and concentration control of plug flow volumes (pipelines, static mixers, and heat exchangers). For well mixed volumes, feedforward summers and an installed linear characteristic for valves is generally best. For control valves, this corresponds to a linear trim when the available pressure drop that is much larger than the system pressure drop or critical pressure drop so the installed is close to the inherent flow characteristic.

The results are also useful for determining the dead time to time constant ratio, which has a profound effect on the tuning factors used and the performance of dead time compensation, which has been discussed on the *Control* magazine voices Control Talk Blog site.

BATCH IMPLICATIONS

Plug flow volumes can always be considered as continuous because the volume is completely full and anything entering will be discharged after a transportation delay.

A back mixed volume is partially full. If the liquid discharge flow is zero, this volume can be considered to be in the batch mode rather than in the continuous mode. If the flows are all sequenced and charged based on time or totals, the vessel operation can be considered to be pure batch. If reactor flows are ratioed and manipulated by a control loop, the vessel operation can be classified as fed-batch.

Level has an integrating response whether in the batch or continuous mode. In the batch mode, it is a zero load integrator in that all the feed flows must be zero for the level to stop rising. Level has a one sided integrating response in the batch mode since the level can only rise and not drop. This type of response causes overshoot for any controller with reset action. Proportional plus derivative (PD) controllers with a zero or negative bias can be tuned for zero overshoot.

The temperature response of a back mixed volume remains self-regulating even for zero liquid discharge flow unless the liquid level is above or below the heat transfer surfaces. However, the temperature response does lose some self-regulation and behaves more like a “near integrator”.

The concentration response of a back mixed volume becomes integrating for a zero discharge flow. This is not obvious because the discharge flow cancels out of the differential equation from the application of the multiplicative rule of integration in the transition of the derivative from the rate of accumulation of component mass to the rate of accumulation of component concentration. The effect of zero discharge flow is more recognizable if we consider the case of a zero reaction rate so that the process time constant is simply the residence time (liquid mass divided by the liquid feed flow rate). The increase in mass for a fixed feed rate over the residence time is simply the feed rate multiplied by the residence time. The result is an increase in mass equal to the existing mass. This doubling of mass doubles the residence time and hence the process time constant. Consequently, the process never reaches a steady state because the process time constant is constantly increasing as the level is rising for a zero discharge flow. For the case of zero reaction, the integrator is a zero load integrator because the feed of the component must be zero for the concentration of the component to stop rising. The concentration here has a one sided integrating response in that the concentration can only increase and not decrease. This would also be the case for reaction products where there is only a forward reaction (no reverse or side reactions). As with level, overshoot is a problem unless PD controllers are used. Alternately, the controlled variable can be translated to a rate of change of concentration as noted in application of model predictive control for bioreactor biomass and product concentration.

Gas pressure is an integrator regardless of liquid discharge flow as long as the pressure in the vessel has a negligible effect on vent flow, which is the case for large or critical pressure drops. If this is not the case, the gas pressure response becomes self-regulating but for large volumes and small vent flows it behaves like a “near integrator”.

RESULTS

Here is a summary of process dynamics computed using equations derived from first principles.

The integrating process gain (K_{ip}) for the control of liquid level by the manipulation of a flow:

$$K_{ip} = 1/(\rho_o * A_o) \tag{F.4d}$$

The integrating process gain (K_{ip}) for the control of pressure per gas law by the manipulation of a flow:

$$K_{ip} = \left[(R_g * T_g)/V_g \right] \tag{F.5d}$$

For the manipulation of jacket temperature to control outlet temperature, the main process time constant (τ_p) is (positive feedback if heat of feed and reaction exceeds product of heat transfer coefficient and area):

$$\tau_p = (C_p * M_o)/\left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6g}$$

For the manipulation of jacket temperature to control outlet temperature, the integrating process gain (K_p) is:

$$K_p = (U * A)/\left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6h}$$

For the manipulation of jacket temperature to control outlet temperature, the near-integrating process gain (K_{nip}) is:

$$K_{nip} = (U * A) / (C_p * M_o) \tag{F.6i}$$

For the manipulation of feed temperature to control outlet temperature, the process gain (K_p) is:

$$K_p = (C_p * F_f) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6j}$$

For the manipulation of feed flow to control outlet temperature, the process gain (K_p) is:

$$K_p = (C_p * T_f) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6k}$$

For manipulation of jacket temperature, the additional small secondary process time constant associated with the heat capacity and mass of the jacket wall is:

$$\tau_{p2} = (C_w * M_w) / \left[U * A \right] \tag{F.6l}$$

The process dead time (θ_p) from the turnover time for temperature and concentration control in a well mixed volume is:

$$\theta_p = (M_o / \rho_o) / \left[(F_f + F_a + F_r) / \rho_o + F_v / \rho_v \right] \tag{F.6m}$$

The process dead time (θ_p) from injection delay for concentration control is:

$$\theta_p = V_1 / (F_1 / \rho_1) \tag{F.6n}$$

For the manipulation of feed flow to control reactant concentration (X_{Ao}), the main process time constant (τ_p) is:

$$\tau_p = M_o / (R_x + F_f) \tag{F.7g}$$

For the manipulation of feed flow to control reactant concentration (X_{Ao}), the process gain (K_p) is:

$$K_p = X_{Af} / (R_x + F_f) \tag{F.7h}$$

For the manipulation of feed flow to control reactant concentration, the near-integrating process gain (K_{nip}) is:

$$K_{nip} = X_{Af} / M_o \tag{F.7i}$$

For the manipulation of feed concentration to control reactant concentration, the process gain (K_p) is:

$$K_p = F_f / (R_x + F_f) \tag{F.7j}$$

For plug flow volumes where different streams are being combined, the process gain (K_p) for controlling the temperature of the mixture (T_f) by the manipulation of flow is:

$$K_p = dT_f / dF_1 = T_1 / \Sigma F_i \tag{F.8a}$$

For plug flow volumes where different streams are being combined, the process gain (K_p) for controlling the composition of component A (X_{Af}) in the mixture by the manipulation of flow is:

$$K_p = dX_{Af} / dF_1 = X_{A1} / \Sigma F_i \tag{F.8b}$$

For plug flow volumes where different streams are being combined, the process dead time for controlling the temperature or composition of the mixture by the manipulation of flow is:

$$\theta_p = V_1 / (F_1 / \rho_1) + V_p / \Sigma(F_i / \rho_i) \tag{F.8c}$$

The process time constant is essentially zero for true plug flow. For a static mixer there is some back mixing, the residence time in Equation F.8c is split between a dead time and time constant per Equations F.8d and F.8e.

$$\theta_p = V_1 / (F_1 / \rho_1) + x * V_p / \Sigma(F_i / \rho_i) \tag{F.8d}$$

$$\tau_p = (1 - x) * V_p / \Sigma(F_i / \rho_i) \tag{F.8e}$$

DERIVATIONS

There are three types of processes: self-regulating, integrating, and runaway as shown in Figures F.1, F.2, and F.3, respectively. A self-regulating process will decelerate to a new steady state. An integrating process will continually ramp. A runaway process will accelerate until hitting a relief or interlock setting.

Over 90 percent of the processes are self-regulating. However, many of the continuous and fed-batch processes in the chemical industry with the greatest direct economic benefits behave and can be best treated as *near-integrating* processes. The classic integrating process is a pure batch or level process. Less than 1 percent of the processes are runaway. When these exist, understanding the runaway response is critical in terms of safety and control because of the propensity to accelerate and reach a point of no return. Runaway responses are almost exclusively associated with highly exothermic reactors used in plastics and specialty chemical production.

This note develops the equations for process dynamics for back mixed volumes and plug flow volumes. The back mixed volumes section applies to volumes whenever an agitator pumping rate, an eductor or recirculation liquid flow rate, or gas evolution or sparge rate produces enough turbulence and back mixing to make the mixture more uniform in the axial besides the radial direction of the volume. Gas volumes generally have enough turbulence and a fast enough gas dispersion rates to be treated as a back mixed volume. The equations therefore hold relatively well for an evaporator and a single distillation stage due to turbulence from the vapor flow. The plug volume section applies to static mixers, pipelines, coil inlets, and jacket inlets where the turbulence from pipe fittings or internal mixing elements creates enough radial

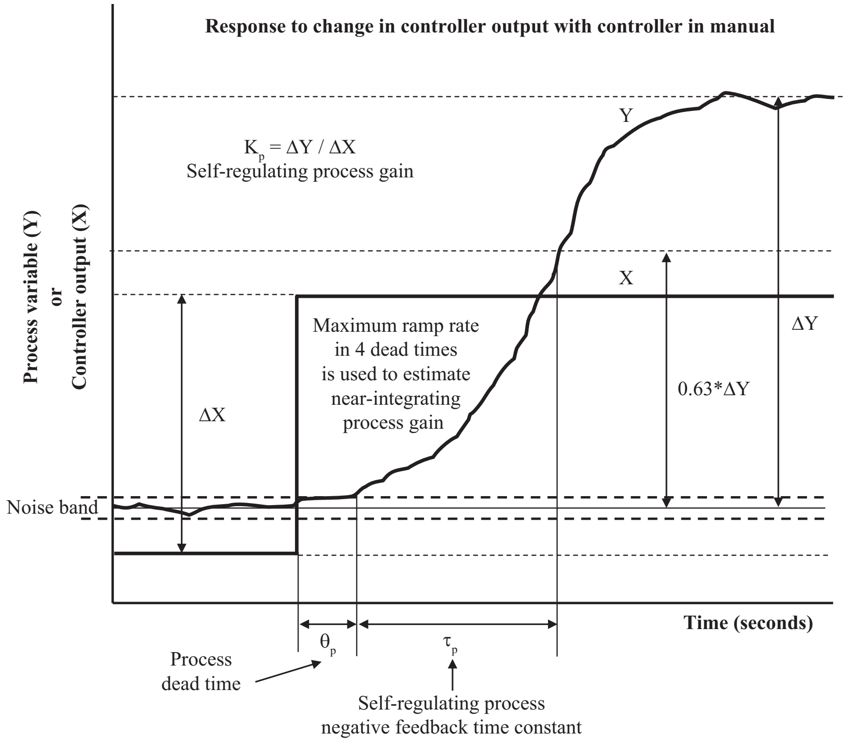


Figure F.1. *Self-regulating* (negative feedback) process.

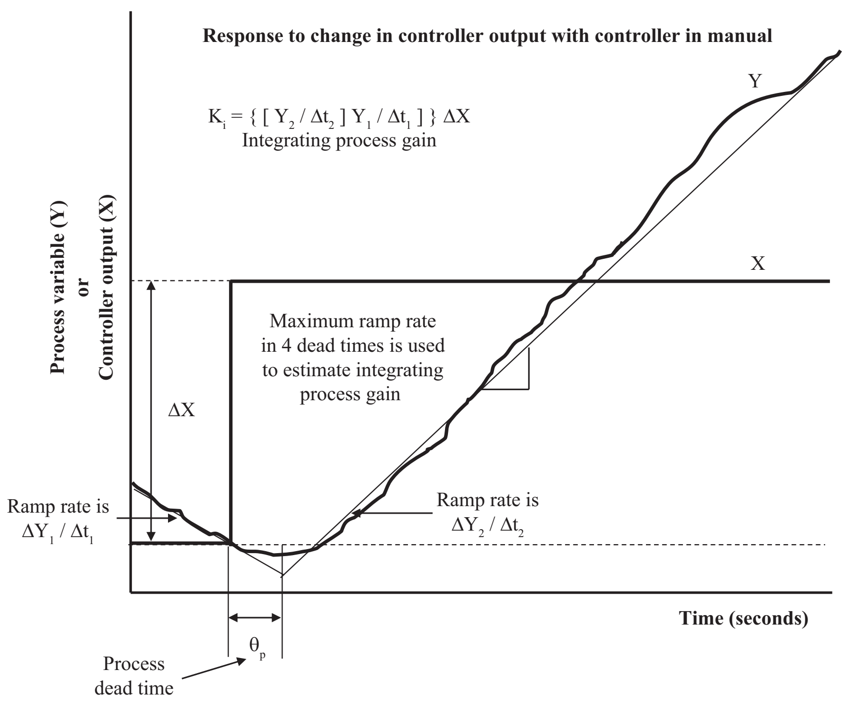


Figure F.2. *Integrating* (zero feedback) process.

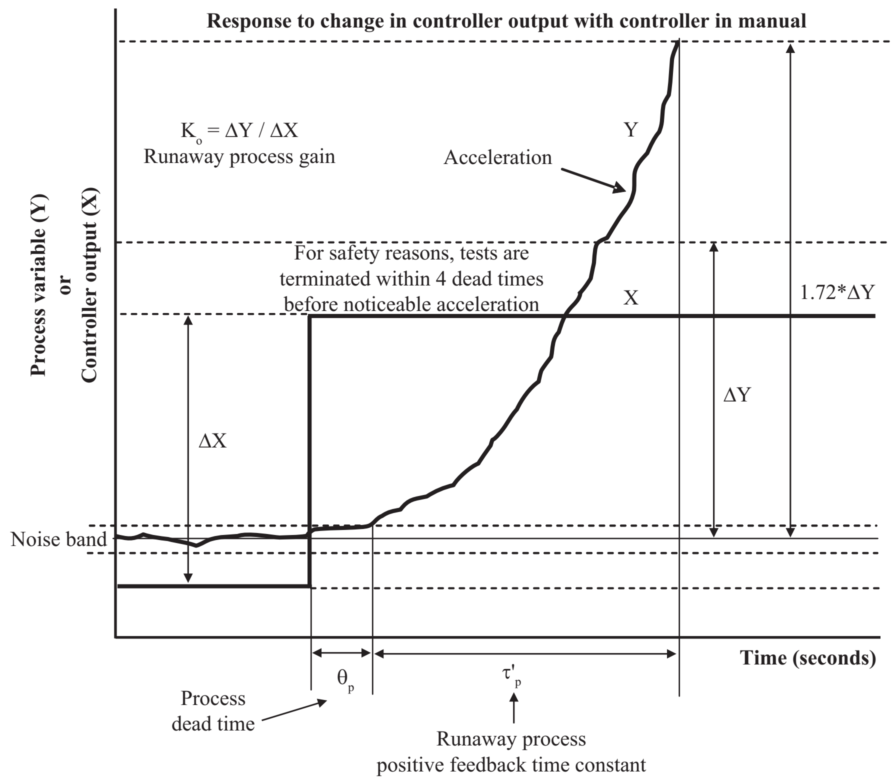


Figure F.3. *Runaway* (positive feedback) process.

mixing to make the mixture uniform over the cross section of the pipe or nozzle inlet but little axial mixing.

BACK MIXED VOLUMES

For a back mixed volume, the process gains and time constants can be readily identified if the ordinary differential equations (ODEs) for the rate of accumulation of energy or material in the volume are set up so that the process output of interest (Y) is on right side with a unity coefficient. From this simple generic form we can identify the process time constant (τ_p) as the coefficient of derivative of the process output (dY/dt) and the process gain (K_p) as the coefficient of the process input (X). The process output (Y) and input (X) can be viewed as the controlled and manipulated variables, respectively. Many other terms can exist but these are not shown in the following equations. These missing terms can be categorized as disturbances.

If the sign of the unity coefficient of the process output on the right side is negative (Equation F.1a), the process has negative feed back. As the process output changes, the negative feedback slows down and eventually halts the excursion of the process output at its new steady state when it balances out the effect of the process input and the disturbances.

$$\tau_p * dY / dt = K_p * X - Y \tag{F.1a}$$

The integration of this equation provides the time response of a change in the process output (ΔY) for a step change in the process input (ΔX). The step occurs at $t = 0$.

$$\Delta Y = K_p * (1 - e^{-t/\tau_p}) * \Delta X \tag{F.1b}$$

If the process output does not appear on the right side (Equation F.2a), there is no process feedback. As the process output changes, there is no feedback to slow it down or speed it up so it continues to ramp. There is no steady state. The ramping will only stop when X is zero or balances out the disturbances.

$$dY / dt = K_i * X \tag{F.2a}$$

$$\Delta Y = K_i * t * \Delta X \tag{F.2b}$$

Often in the more important loops for concentration, pressure, and temperature control of large volumes, the time constant in Equation F.1a is so large that the time to reach steady state is beyond the time frame of interest. Since these loops with small dead time to time constant ratios should be tuned with small Lambda factors (high controller gains) as per Advanced Application Note 3, the controller only sees the first part of the excursion before the inflection point and deceleration by negative process feedback. In this case, the response is best characterized by a near-integrating process gain calculated as per Equation F.2c.

$$K_{nip} = K_p / \tau_p \tag{F.2c}$$

If the sign of the unity coefficient of the process output on the right side is positive (Equation F.3a), the process has positive feed back. As the process output changes, the positive feedback speeds up the excursion unless disturbances counteract the effect of the process input and output.

$$\tau_p' * dY / dt = K_p * X + Y \tag{F.3a}$$

$$\Delta Y = K_p * (e^{t/\tau_p'} - 1) * \Delta X \tag{F.3b}$$

Consider a mixed volume with a jacket and vapor space. There are liquid reactant feeds, gas feeds (sparged through the liquid and added directly to the vapor space), an outlet liquid flow, a vent gas flow, and a jacket coolant flow. There is normally multiple components interest. For example, consider liquid or gas acid and base reagent or reactant components (a, b) to produce primary and secondary liquid or gas products (c, d, e). Consider also there are typically water and nitrogen gas components (w, n).

The ODE for the accumulation of liquid mass as shown in Equation F.4a includes inlet flows added directly to the liquid volume ($\sum F_i$), vapor flow rates from evaporation and vaporization ($\sum F_v$), and an outlet liquid flow rate (F_o). The liquid level depends upon density and cross section area of the liquid. Equation F.4a can then be reformulated to Equation F.4b to include the process variable of interest, liquid level (L_l), in the derivative.

$$dM_o / dt = \sum F_i - \sum F_v - F_o \tag{F.4a}$$

$$d(\rho_o * A_o * L_o) / dt = \sum F_i - \sum F_v - F_o \tag{F.4b}$$

If we consider the density (ρ_o) to be a weak function of composition and therefore constant like the cross sectional area (A_o) we can take these terms outside the derivative and divide through to get an equation for level (L_o) in the form of Equation F.2a. Now it is clearly evident that the integrating process gain (K_{ip}) for manipulation of flows in or flow out is simply the inverse of the product of the liquid density and cross section area (Equation F.4d).

$$dL_o / dt = \left[1/(\rho_o * A_o)\right] * \left[\Sigma F_i - \Sigma Fv - F_o\right] \tag{F.4c}$$

$$K_{ip} = 1/(\rho_o * A_o) \tag{F.4d}$$

The ODE for the accumulation of gas mass as shown in Equation F.5a includes inlet flows added directly to the gas volume (ΣF_i), vapor flow rates from gas sparging, evolution, and vaporization (ΣF_v), and an exit gas flow rate (F_g). Equations of state such as the ideal gas law can be used to express this relationship for a given composition. Equation F.5a can then be reformulated to Equation F.5b to include the process variable of interest, gas pressure (P_g), in the derivative.

$$dM_g / dt = \Sigma F_i + \Sigma F_v - F_g \tag{F.5a}$$

$$d\left[(P_g * V_g)/(R_g * T_g)\right]/ dt = \Sigma F_i + \Sigma F_v - F_g \tag{F.5b}$$

If we consider changes in the gas volume (V_g) and gas temperature (T_g) to be much slower than changes in the gas pressure (P_g) and therefore relatively constant during the integration step, we can take these terms outside the derivative and divide through to get an equation for pressure in the form of Equation F.2a. Now it is clearly evident that the integrating process gain (K_{ip}) for manipulation of flow in or flow out is simply the product of the universal gas coefficient (R_g) and the absolute gas temperature divided by the gas volume (Equation F.5d). This assumes that a change in pressure does not significantly change the gas glow out of the volume, which is normally the case for a pressure drop across the vent valve that is large or critical.

$$dP_g / dt = \left[(R_g * T_g)/V_g\right] * \left[\Sigma F_i + \Sigma F_v - F_g\right] \tag{F.5c}$$

$$K_{ip} = \left[(R_g * T_g)/V_g\right] \tag{F.5d}$$

The ODE for the accumulation of energy as shown in Equation F.6a includes the effects of feed temperature, heat of reaction as a function of temperature, heat of vaporization, and heat transfer to the jacket. If we consider the specific heat capacity relatively constant and use the multiplicative rule of integration, we can express the differential equation in the generic form of Equation F.1a in terms of temperature to show the process feedback in Equation F.6f, the final form of the ODE. The relative magnitude of the terms in the denominator of Equation F.6g determines the feedback sign.

$$\begin{aligned} dQ_o / dt = {} & C_p * \Sigma(F_i * T_i) - C_p * F_o * T_o + (\Delta Q_r / \Delta T_o) * T_o + H_x \\ & * R_x - H_v * F_v - U * A * (T_o - T_j) \end{aligned} \tag{F.6a}$$

$$dQ_o / dt = d(C_p * M_o * T_o)/ dt = C_p * (dM_o/dt) * T_o + C_p * M_o * (dT_o/dt) \tag{F.6b}$$

$$F_f = \Sigma F_i \tag{F.6c}$$

$$T_f = \Sigma(F_i * T_i) / \Sigma F_i \tag{F.6d}$$

$$C_p * (dM_o/dt) * T_o = C_p * (F_f - F_o) * T_o \tag{F.6e}$$

$$C_p * M_o * (dT_o/dt) = C_p * F_f * T_f + H_x * R_x - H_v * F_v + U * A * T_j - \left[C_p * F_f - DQ_r / DT_o + U * A \right] * T_o \tag{F.6f}$$

For the manipulation of jacket temperature to control outlet temperature, the main process time constant (τ_p) is (positive feedback if heat of feed and reaction exceeds product of heat transfer coefficient and area):

$$\tau_p = (C_p * M_o) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6g}$$

For the manipulation of jacket temperature to control outlet temperature, the process gain (K_p) is:

$$K_p = (U * A) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6h}$$

For the manipulation of jacket temperature to control outlet temperature, the near-integrating process gain (K_{nip}) is:

$$K_{nip} = (U * A) / (C_p * M_o) \tag{F.6i}$$

For the manipulation of feed temperature to control outlet temperature, the process gain (K_p) is:

$$K_p = (C_p * F_f) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6j}$$

For the manipulation of feed flow to control outlet temperature, the process gain (K_p) is:

$$K_p = (C_p * T_f) / \left[C_p * F_f - \Delta Q_r / \Delta T_o + U * A \right] \tag{F.6k}$$

For manipulation of jacket temperature, the additional small secondary process time constant associated with the heat capacity and mass of the jacket wall is:

$$\tau_p = (C_w * M_w) / \left[U * A \right] \tag{F.6l}$$

Any change in the temperature at the heat transfer surfaces or the feed inlet must be dispersed and back mixed into the volume. This process dead time (θ_p) is the turn over time that can be approximated as the liquid inventory divided by the summation of the feed flow rate (F_f), agitator pumping rate (F_a), recirculation flow rate (F_r), and vapor evolution rate or vapor bubble rate (F_v). Since this turn over time is computed in terms of volumetric flow rates, the liquid mass and the mass flow rates are divided by their respective densities as shown in Equation F.6m.

$$\theta_p = (M_o / \rho_o) / \left[(F_f + F_a + F_r) / \rho_o + F_v / \rho_v \right] \tag{F.6m}$$

If there is an injector (dip tube or sparger ring) volume, a change in composition at the nozzle must propagate by plug flow to the discharge points of the dip tube or sparger ring. The dead time for a feed flow (F_1) is the injector volume (V_1) divided by the injection mass flow (F_1) divided by its respective density (ρ_1).

$$\theta_p = V_1 / (F_1 / \rho_1) \tag{F.6n}$$

The ODE for the accumulation of liquid reactant mass (M_A) as shown in Equation F.7a includes the effects of feeds (F_i) with a reactant mass fraction (X_{Ai}), reaction rate (R_x), and outlet flow (F_o). The feeds can be from raw material, intermediate products, recycle streams, or multi-stage reactors. If we use the multiplicative rule of integration, we can express the differential equation in the generic form of Equation F.1a in terms of concentration to show the process feedback in Equation F.7f, the final form of the ODE.

$$dM_A / dt = \Sigma(F_i * X_{Ai}) - (R_x + F_o) * X_{Ao} \tag{F.7a}$$

$$dM_A / dt = d(M_o * X_{Ao}) / dt = (dM_o / dt) * X_{Ao} + M_o * (dX_{Ao} / dt) \tag{F.7b}$$

$$F_f = \Sigma F_i \tag{F.7c}$$

$$X_{Af} = \Sigma(F * X_{Ai}) / \Sigma F_i \tag{F.7d}$$

$$(dM_o / dt) * X_{Ao} = (F_f - F_o) * X_{Ao} \tag{F.7e}$$

$$M_o * (dX_{Ao} / dt) = F_f * X_{Af} - (R_x + F_f) * X_{Ao} \tag{F.7f}$$

For the manipulation of feed flow to control reactant concentration (X_{Ao}), the main process time constant (τ_p) is:

$$\tau_p = M_o / (R_x + F_f) \tag{F.7g}$$

For the manipulation of feed flow to control reactant concentration (X_{Ao}), the process gain (K_p) is:

$$K_p = X_{Af} / (R_x + F_f) \tag{F.7h}$$

For the manipulation of feed flow to control reactant concentration, the near integrator gain (K_i) is:

$$K_i = X_{Af} / M_o \tag{F.7i}$$

For the manipulation of feed concentration to control reactant concentration, the process gain (K_p) is:

$$K_p = F_f / (R_x + F_f) \tag{F.7j}$$

The process dead times from turnover time and from feed injection are the same as computed in the section for temperature control (Equations F.6m and F.6n).

PLUG FLOW VOLUMES

For plug flow volumes where different streams are being combined, the process gain for controlling the temperature (T_f) or composition (X_{Af}) of the mixture (often a feed to a downstream equipment) can be computed by taking the derivative of Equations F.6d and F.7d with respect to the manipulated flow stream 1 (F_1) to give Equations F.8a and F.8b, respectively. In both cases, the process gain is inversely proportional to total flow ($\sum F_i$).

$$K_p = dT_f / dF_1 = T_1 / \sum F_i \tag{F.8a}$$

$$K_p = dX_{Af} / dF_1 = X_{A1} / \sum F_i \tag{F.8b}$$

The process dead time for the manipulation of a flow for stream 1 (F_1) is the summation of the injection delay for steam 1 and the piping delay from the point of injection to the point of temperature or composition measurement. For plug flow the residence time, which is the second expression in Equation F.8c completely becomes dead time.

$$\theta_p = V_1 / (F_1 / \rho_1) + V_p / \Sigma(F_i / \rho_i) \tag{F.8c}$$

The process time constant is essentially zero for true plug flow. For a static mixer, there is some back mixing, the residence time in Equation F.8c is split between a dead time and time constant as per Equations F.8d and F.8e.

$$\theta_p = V_1 / (F_1 / \rho_1) + x * V_p / \Sigma(F_i / \rho_i) \tag{F.8d}$$

$$\tau_p = (1 - x) * V_p / \Sigma(F_i / \rho_i) \tag{F.8e}$$

It is obvious from the above that both the process gain and dead time are inversely proportional to total flow.

CONTROLLER TUNING

The implication of the results can be best seen if Lambda is set equal to the total loop dead time (θ_o) resulting in Equation F.9a for the PID gain. If the open loop time constant (τ_o) is large compared to the dead time, the process response is termed near-integrating and the ratio of the open loop self-regulating process gain (K_o) to the time constant is used to approximate the open loop integrating process gain (K_i) in Equation 5.10b. This PID gain is about half of the PID gain estimated by the Ziegler Nichols reaction curve method.

$$K_c = 0.5 * \frac{\tau_o}{K_o * \theta_o} \tag{F.9a}$$

$$K_c = 0.5 * \frac{1}{K_i * \theta_o} \tag{F.9b}$$

The open loop time constant (τ_o) in the numerator of Equation F.9a is the largest time constant in the loop wherever it occurs. Hopefully, the process is mixed well enough and the instrumentation is fast enough that the largest time constant is in the process ($\tau_o = \tau_p$) and not the automation system. A large time constant in the process slows down the disturbance and is desirable. A large time constant in the measurement and final element is detrimental because it slows down the ability of the controller to see and react to disturbance, respectively.

The open loop self-regulating process gain (K_o) in the denominator is dimensionless. The process gain is actually the product of the manipulated variable gain, the process gain (K_p), the gain of nonlinear process variables, such as pH (slope of the titration curve), and the controlled variable gain. For a loop that throttles a control valve, the manipulated variable gain is the slope of the valve's installed characteristic. For the primary loop of a cascade control system, the manipulated variable gain is the secondary loop set point span divided by 100 percent. The controlled variable gain is 100 percent divided by the process variable span. Thus, changes in calibration span affect the computed controller gain, which provides robustness and a less oscillatory response.

Finally, the dead time (θ_o) in the denominator is really the total loop dead time, which is summation of the process dead time (θ_p) plus all the small time lags and delays in the loop. While the names open loop time constant (τ_o), open loop self-regulating process gain (K_o), and total loop dead time (θ_o) for the parameters in Equation F.9a are more definitive, nearly all of the control literature uses the terms process time constant, process gain, and process dead time indiscriminately.

Nomenclature Process Parameters:

A_o = cross sectional area of liquid level (m^2)
A = heat transfer surface area (m^2)
C_P = heat capacity of process (kJ/kg*°C)
C_w = heat capacity of wall of heat transfer surface (kJ/kg*°C)
F_a = agitator pumping rate (kg/sec)
F_f = total feed flow (kg/sec)
F_g = gas flow (kg/sec)
F_i = feed stream i flow (kg/sec)
F_o = vessel outlet flow (kg/sec)
F_r = recirculation flow (kg/sec)
F_v = vaporization rate (kg/sec)
H_v = heat of vaporization (kJ/kg)
H_x = heat of reaction (kJ/kg)
L_o = liquid level (m)
M_A = component A mass (kg)
M_g = gas mass (kg)
M_o = liquid mass (kg)
M_w = mass of wall of heat transfer surface (kg)
P_g = gas pressure (kPa)
T_f = total feed temperature (°C)
T_g = gas temperature (°C)
T_i = feed stream i temperature (°C)
T_o = vessel outlet temperature (°C)
t = time (sec)
Q_o = total heat of liquid (kJ)

Q_r = heat from reaction (kJ)
R_x = reaction rate (kg/sec)
R_g = universal constant for ideal gas law (kPa m^3
ρ_g = gas density (kg/m^3)
ρ_i = stream i density (kg/m^3)
ρ_o = liquid density (kg/m^3)
ρ_v = density of vapor (kg/m^3)
U = overall heat transfer coefficient (kJ/m^2*°C)
V_g = gas volume (m^3)
V_i = injection (e.g. dip tube or sparger ring) volume (m^3)
V_p = piping volume (m^3)
x = fraction of volume that is plug flow
X_{Af} = total feed component A concentration (mass fraction)
X_{Ai} = feed stream i component A concentration (mass fraction)
X_{Ao} = vessel outlet component A concentration (mass fraction)

Generic Terms:

X = process input (manipulated variable) (eu)
Y = process output (controlled variable) (eu)

Dynamic Parameters:

K_c = PID controller gain (dimensionless)
K_i = open loop integrating process gain (1/sec)
K_{ip} = integrating process gain (eu/eu)
K_o = open loop self-regulating process gain (dimensionless)
K_p = process gain (eu/eu)
K_{nip} = near-integrating process gain (eu/eu/sec)
τ_p = negative feedback process time constant (sec)
τ'_p = positive feedback process time constant (sec)
τ_o = open loop time constant (sec)
θ_p = process dead time (sec)
θ_o = total loop dead time (sec)

APPENDIX G

GAS PRESSURE DYNAMICS

The interactive time constants of multiple gas volumes in series (e.g., vessels, columns, and headers) in series separated by flow resistances (e.g., valves and fittings) can be estimated by Equations G.1 through G.4 for turbulent and sonic flow in *Process Control Systems* (Harriot 1964).

$$\tau_i = \frac{C_i}{\sum_i \frac{1}{R_i}} \tag{G.1}$$

$$C_i = \frac{V_i}{P_a} \tag{G.2}$$

For turbulent flow:

$$R_i = \frac{2 * \Delta P_i}{F_i} \tag{G.3}$$

For sonic flow:

$$R_i = \frac{P_i}{F_i} \tag{G.4}$$

where

C_i = capacitance of gas volume i (scf/psi)
F_i = gas flow through resistance i (scfm)
P_a = atmospheric pressure at resistance i (psi)
P_i = inlet pressure at resistance i (psi)
ΔP_i = pressure drop across resistance i (psi)
R_i = resistance to gas flow into or out of volume i (psi/scfm)
V_i = gas volume i (scf)
τ_i = interactive time constant of gas volume i (minutes)

If the intermediate flow resistance pressure drops for turbulent non sonic flow between volumes in a gas system are less than 2 percent of the system inlet and outlet flow resistance pressure drops, then all the individual volumes can be summed and substituted into Equation G.2 to give a single process time constant for the gas system. The universal gas sizing equation can be used to estimate the Cv of the fitting or valve necessary to prevent the creation of multiple time constants. If the pressure drops between volumes is not negligible (flow coefficient of internal resistance does not satisfy Equation G.5), then the system should be partitioned into two major volumes and Appendix I used in conjunction with equations here to estimate the equivalent two non-interactive time constants for two volumes in series. The non-interactive time constants should be then used in Equations 4.1 and 4.2 in Chapter 4 to estimate the process dead time.

$$C_v > \frac{F}{P * \sqrt{\frac{0.02 * \Delta P}{P}}} * \sqrt{\frac{G * T}{520}} \tag{G.5}$$

where

C_v = gas flow coefficient of internal resistance (scfm/psi)
F = gas flow through system (scfm)
G = gas specific gravity in system (dimensionless)
P = pressure at system inlet (psi)
ΔP = pressure drop at system outlet (psi)
T = absolute temperature in system (°R)

APPENDIX H

Convective Heat Transfer Coefficients

1. For heating and cooling liquids flowing perpendicular to a single cylinder (e.g., thermowell), the convective heat transfer coefficient can be estimated as follows (note that the viscosity and velocity use hours for time units) "Response of Temperature Measuring Elements" (Kardos 1977):

$$h = [k/(2 * r)] * \{0.35 + 0.56 * [(2 * r * v * d)/u]^{0.52}\} * [(c * u)/k]^{0.3} \quad \text{(H.1)}$$

where

c = heat capacity of fluid (Btu/lb * °F)
d = density of fluid (lb/ft^3)
r = outside radius of thermowell (ft)
k = thermal conductivity of fluid (Btu/hr * ft * °F)
u = viscosity of fluid (lb/ft * hr)
v = velocity of fluid (ft/hr)

2. For the more general thermowell case that covers a wide range of Reynolds numbers and fluids, the convective heat transfer coefficient can be estimated as follows (note that viscosity is in centipoise and velocity is in ft/sec) "Thermowell Heat Conduction Error" (Crawford 1982):

$$h = 3960 * v * d * c * Cj * Re^{(Nj - 1)} * Pr^{(-0.69)} \quad \text{(H.2)}$$

$$Re = (248 * r * v * d)/u \quad \text{(H.3)}$$

$$Pr = (2.42 * u * c)/k \text{ (liquids)} \quad \text{(H.4)}$$

$$Pr = 0.735 \text{ (gases)} \quad \text{(H.5)}$$

where

c = heat capacity of fluid (Btu/lb * °F)
d = density of fluid (lb/ft^3)

r = outside radius of thermowell (ft)
k = thermal conductivity of fluid (Btu/hr * ft * °F)
Pr = Prandtl number (dimensionless)
Re = Reynolds number (dimensionless)
u = viscosity of fluid (cp)
v =velocity of fluid (ft/sec)

	Cj	Nj
Re < 4	0.891	0.330
4 < Re < 40	0.821	0.385
40 < Re < 4,000	0.615	0.466
4,000 < Re < 40,000	0.174	0.618
40,000 < Re	0.024	0.805

3. For the flow of fluids in tubes where the Reynolds number is between 10,000 and 120,000, the convective heat transfer coefficient (typical values shown in Table H.1) can be estimated as follows (note that viscosity and velocity use seconds for time units) (Kreith 1965):

$$h = 0.023 * v^{0.8} * (2 * r)^{-0.2} * k * (u/d)^{-0.8} \tag{H.6}$$

where

d = density of fluid (lb/ft^3)
r = inside radius of tube (ft)
k = thermal conductivity of fluid (Btu/hr * ft * °F)
u = viscosity of fluid (lb/ft * sec)
v = velocity of fluid (ft/sec)

Table H.1. The convective heat transfer coefficient range changes with fluid type, turbulence, and phase *Principles of Heat Transfer* (Kreith 1965)

Fluid and condition	h range[2] (Btu/hr * ft^2 * °F)
Air-free convection	1–5
Air- or superheated steam-forced convection	5–50
Oils-forced convection	10–30
Water-forced convection	50–2,000
Water boiling	500–10,000
Steam condensing	1,000–20,000

APPENDIX I

Interactive to Noninteractive Time Constant Conversion

Interactive time constants can be partitioned into two sub systems in series with their capacitance and resistance components in order to calculate the equivalent noninteractive time constants. The quadratic equation can then be used to find the two noninteractive time constants for two capacitances that have an inlet and outlet resistance and are separated by a resistance (Harriot 1964). Interactive time constants exist when the flow or transfer of mass or heat into, within, and out of a system depend upon the operating conditions as driving forces (e.g., pressure drop for flow, concentration differences for mass transfer, and temperature difference for heat transfer). For gas pressure, systems the flow resistance of valves and fittings into, within, and out of the volumes are important. For thermal systems such as vessel coils and jackets, heat exchanger tubes, and thermowells, the resistance of convective heat transfer of the inside and outside surfaces and the resistance of conductive heat transfer within the surface are important.

EQUIVALENT NONINTERACTIVE TIME CONSTANTS

$$\tau_{n1} = \frac{2 * A}{B - \sqrt{B^2 - 4 * A * C}} \tag{I.1}$$

$$\tau_{n2} = \frac{2 * A}{B + \sqrt{B^2 - 4 * A * C}} \tag{I.2}$$

$$A = R_1 * C_1 * R_2 * C_2 * R_3 \tag{I.3a}$$

$$B = R_1 * C_1 * R_3 + R_1 * C_1 * R_2 + R_2 * C_2 * R_3 + R_1 * C_2 * R_3 \tag{I.3b}$$

$$C = R_1 + R_2 + R_3 \tag{I.3c}$$

where

A = first term of quadratic equation
B = second term of quadratic equation
C = third term of quadratic equation

C_1 = capacitance of section 1
C_2 = capacitance of section 2
R_1 = resistance at inlet to section 1
R_2 = resistance between sections 1 and 2
R_3 = resistance at outlet of section 2
τ_{n1} = noninteractive time constant 1 (minute)
τ_{n2} = noninteractive time constant 2 (minutes)

ORIGINAL INTERACTIVE TIME CONSTANTS

The following equations can be used to compute the interactive time constants for a system partitioned into two sections with an inlet resistance (R_1) and outlet resistance (R_3) separated by a resistance (R_2). The capacitances and resistances for gas pressure systems are defined in Appendix G.

$$\tau_{i1} = \frac{C_1}{\dfrac{1}{R_1} + \dfrac{1}{R_2}} \tag{I.4a}$$

$$\tau_{i2} = \frac{C_2}{\dfrac{1}{R_2} + \dfrac{1}{R_3}} \tag{I.4b}$$

where

C_1 = capacitance of section 1
C_2 = capacitance of section 2
R_1 = resistance at inlet to section 1
R_2 = resistance between sections 1 and 2
R_3 = resistance at outlet of section 2
τ_{i1} = interactive time constant 1 (minute)
τ_{i2} = interactive time constant 2 (minutes)

APPENDIX J

JACKET AND COIL TEMPERATURE CONTROL

The heating and cooling system is critical for reactors since temperature plays such a huge role in determining reaction rate and selectivity (formation of desired product). *The limit to what a valve position controller (VPC) can do to increase production rate depends upon the capability of the heating and cooling system. The system can also be the source of discontinuities and nonlinearities. Most reactor temperature control problems beyond tuning can be traced back to deficiencies in the reactor coil or jacket temperature system design.*

Heating is required for endothermic reactions (reactions that consume energy), liquids that are being vaporized, or bringing a reactor up to operating temperature. Cooling is needed for exothermic reactions (reactions that product heat), vapors that are being condensed, and bringing a reactor down to operating temperature. Often both heating and cooling are required. Split ranged control is used to go back and forth between heating and cooling. In split range control for an exothermic reactor with a fail open cooling valve, as the temperature controller output increases from 0 percent to the split range point, the coolant valve goes from wide open to completely closed. As the temperature controller output increases from the split range point to 100 percent, the heating valve goes from closed to wide open. The split range point is traditionally set at 50 percent but would be better set so that the change in temperature for a change in flow is about the same for each valve. While cooling and heating may be done by coils besides jackets, many of the control considerations are the same. Jacketed reactors will be used for purposes of discussion and illustration. Cascade temperature control is the most prevalent strategy where the primary reactor temperature proportional-integral-derivative (PID) provides a setpoint to a secondary jacket temperature PID for the throttling of hot and cold fluids (Figures J.1a and J.1b). For highly exothermic reactors, boiler feedwater BFW is added under level control and the reactor temperature PID output is the jacket outlet steam pressure PID setpoint. These jacket control schemes are suitable for batch besides continuous operation.

The use of cascade control where the reactor temperature controller output is the setpoint of a coil or jacket temperature loop offers considerable performance improvements. The jacket temperature controller can correct for disturbances to the jacket before they affect the reactor temperature. The jacket temperature loop also isolates process and valve gain nonlinearity from the reactor loop. For a negligible increase in the heat of reaction compared to heat transfer capability, the process gain for the reactor temperature loop approaches unity.

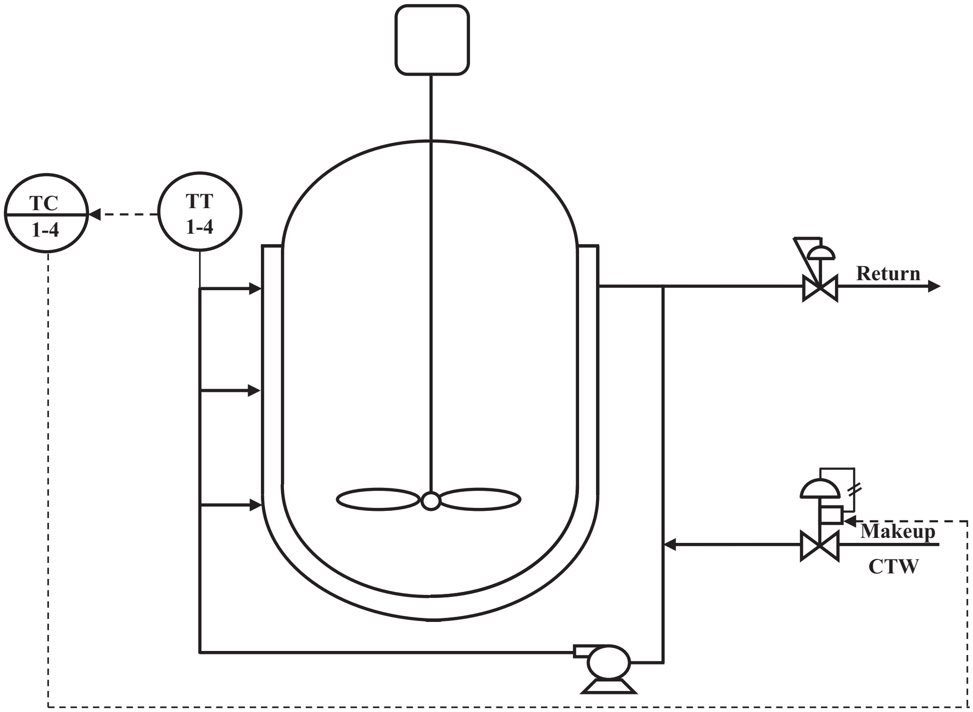


Figure J.1a. Jacket inlet temperature control offers faster correction of jacket disturbances.

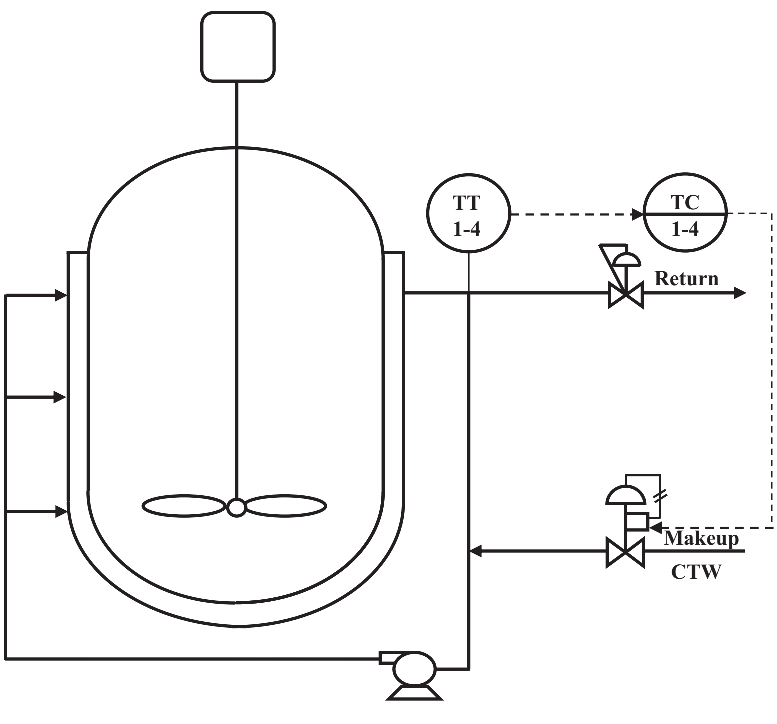


Figure J.1b. Jacket outlet temperature control offers moderation of mixing and discontinuities.

Coils generally offer a faster temperature response than a jacket by a decrease in the volume and an increase in the velocity. Both of these work to decrease the process dead time that is the coil volume divided by the utility flow rate. The increase in velocity increases the heat transfer coefficient but this is partially offset by a decrease in the surface area. An increase in the product of the heat transfer coefficient and surface area (UA) will decrease the secondary process lag in the thermal response. The transition in split range operation is faster, which is useful for a valid transition between hot and cold utility streams but can be problematic for inadvertent transitions from stick-slip and an integrating response in the process or controller.

Whether the secondary loop uses coil or jacket inlet or outlet temperature is often a matter of tradition for a particular company or process industry. The dynamic response of the cascade control system to reactor disturbances such as feed and reaction rate are the same for jacket inlet and outlet temperature control. *The coil or jacket inlet temperature control will correct for changes in cooling or heating utility supply temperature and pressure sooner than the transportation delay through the jacket.* The jacket loop process dead time is also less by the amount of this delay allowing a faster reset time setting and faster correction of valve nonlinearities. *At the coil or jacket inlet, the mixing of the recirculation with the hot or cold makeup flow may be incomplete and the discontinuity in the transition from hot to cold more abrupt.* The location of the temperature sensor on the jacket outlet offers time for mixing and volume for smoothing transitions. Less measurement temperature noise can translate to a higher controller gain and less overreaction to the discontinuity at a split range transition.

The difference between the reactor and the jacket outlet temperature (approach temperature) can provide an inferential measurement of the heat transfer coefficient (U) for a constant jacket circulation flow, a given production rate, and heat transfer area (A). The approach temperature increases as UA decreases. For residence time control, the increase in level will increase the heat transfer area covered by reactants, offsetting the increase in heat release with production rate eliminating the need for production rate and level correction. For fed-batch reactors and continuous reactors without residence time control, a correction for level is needed to compute U from the UA.

For the coil or jacket temperature control to provide rapid adjustments of cooling and heating for disturbances to the jacket and setpoint changes from the reactor temperature control, the coil or jacket PID response needs to be fast achieved by a fast sensor and fast tuning.

A tight fitting sensor bottomed in a thermowell with a high thermal conductivity metal and tapered tip provides a fast measurement. Spring loading can ensure the sensor sheath is bottomed. The clearance between the sheath outside diameter and the thermowell inside diameter must be minimized and the fluid velocity maximized. While grounded thermocouple sensors are a few seconds faster than resistance temperature detectors (RTD) sensors, the difference is insignificant compared to the effect of thermowell design and fluid velocity. The thermowell design details are also important for reactor temperature. The greater sensitivity and lower drift of the RTD is important for jacket besides reactor temperature measurements. The lower drift reduces maintenance and the higher sensitivity provides faster recognition. The use of RTDs also facilitates more accurate online heat transfer computation for process diagnostics and inferential measurements of reaction rate. Thermocouples are preferred for temperatures above 400°C where RTD insulation resistance and sensor integrity become problematic.

A high controller gain and a small reset time provide fast tuning. The jacket temperature controller is self-regulating process with a maximum PID gain that is about half of the open loop time constant divided by the product of the open loop gain and dead time as detailed in the section on controller tuning. The minimum reset setting is about four times the loop dead time for jacket temperature PID.

Most temperature loops tend to develop an oscillation across the split range point in making the transition between heating and cooling. The transition creates a discontinuity from the change in utility fluid and the increase in friction and loss of sensitivity in valve seating or sealing. The worst case is often the transition between steam and coolant due to the huge difference in temperatures and the creation of bubbles in coolant or water droplets in steam. Steam valves tend to have higher seating and sealing friction due to the higher temperatures and pressures. Hot and cold liquids and tempered water systems offer a much smoother transition than steam and coolant and reduce nonlinearities.

Steam does not provide uniform heating in coils or a jacket. Steam typically collects in the top of the jacket and condensates in the bottom. Hot spots can develop around inlets. Thermal shock and steam hammer can damage glass lined vessels. The time required to drive steam completely out of the jacket before cooling water introduces a considerably delay in the control system. Improper trap design or operation can cause condensate buildup.

The addition of hot water instead of steam directly into the jacket provides a more uniform heat distribution, a dramatically smoother transition between heating and cooling, and a more

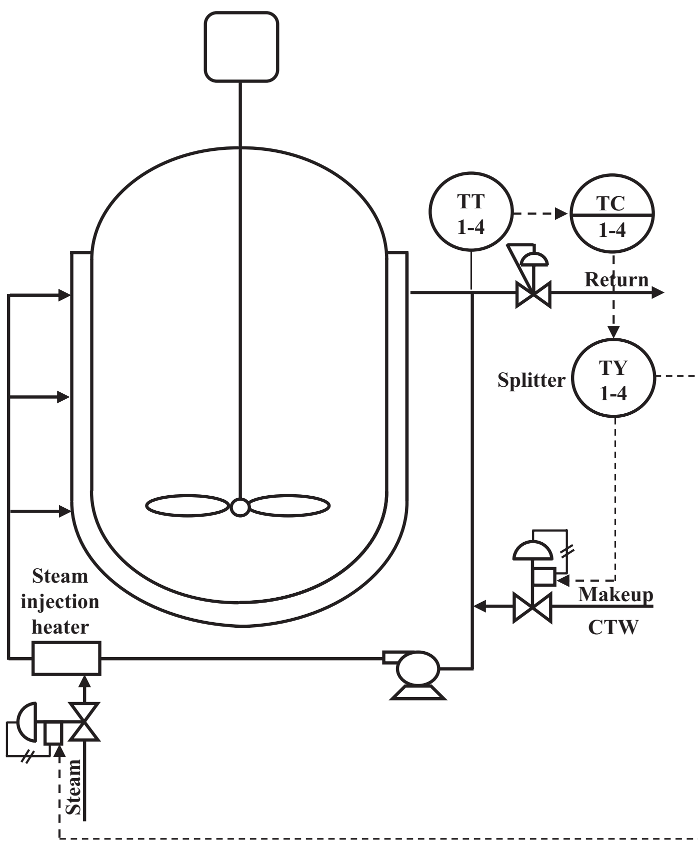


Figure J.2. Steam injection heaters to create hot water offer rapid heating and tight control of coil or jacket temperature with smoother split ranged transitions.

efficient and maintainable system. For rapid heating, the use of direct steam injection heaters and pressurized water can provide hot water temperatures well above 100°C (Figure J.2). If the injection heater has hundreds of small orifices, the bubbles are extremely small and are rapidly and quietly mixed into the water. Variable orifice steam injection heaters are not as quiet and the mixing is not as complete. The use of jacket outlet temperature reduces the possibility of bubbles hitting the temperature sensor.

APPENDIX K

PID Forms and Conversion of Tuning Settings

CONVERSION OF SETTINGS TO CORRECT UNITS AND FORM

The equations in this book are based on the ISA Standard Form with specific tuning setting units where the controller gain is dimensionless, the reset time is seconds (seconds per repeat), and the rate time is seconds. Proportional-integral-derivative (PID) tuning settings must first be checked for units and if necessary undergo a unit conversion before being used.

1. Convert proportional band (%) to controller gain (%/%) (dimensionless)
   Gain = 100%/Proportional Band
2. Convert reset setting in repeats per minute to reset time in seconds per repeat
   Seconds per repeat = 60/repeats per minute
3. Convert rate time in minutes to rate time in seconds
   Seconds = 60 * minutes
4. After the tuning setting units are verified, Series and Parallel Forms convert tuning settings to the ISA Standard Form tuning settings as per equations below. The primed tuning settings are for the Series Form, the double primed tuning settings are for the Parallel Form, and the unprimed tuning settings are for the ISA Standard Form used in this book.

To convert from Series to ISA Standard Form controller gain:

$$K_c = \frac{T_i' + T_d'}{T_i'} * K_c' \tag{K.1}$$

To convert from Series to ISA Standard Form reset (integral) time:

$$T_i = \frac{T_i' + T_d'}{T_i'} * T_i' = T_i' + T_d' \tag{K.2}$$

To convert from Series to ISA Standard Form rate time:

$$T_d = \frac{T_i^{'}}{T_i^{'} + T_d^{'}} * T_d^{'} \tag{K.3}$$

Note that if the rate time is zero, the ISA Standard Form (Figure K.1) and Series Form (Figure K.2) settings are identical. When using the ISA Standard Form, if the rate time is greater than one-fourth the reset time the response can become oscillatory. If the rate time exceeds the reset time, the response can become unstable from a reversal of action form these modes. The

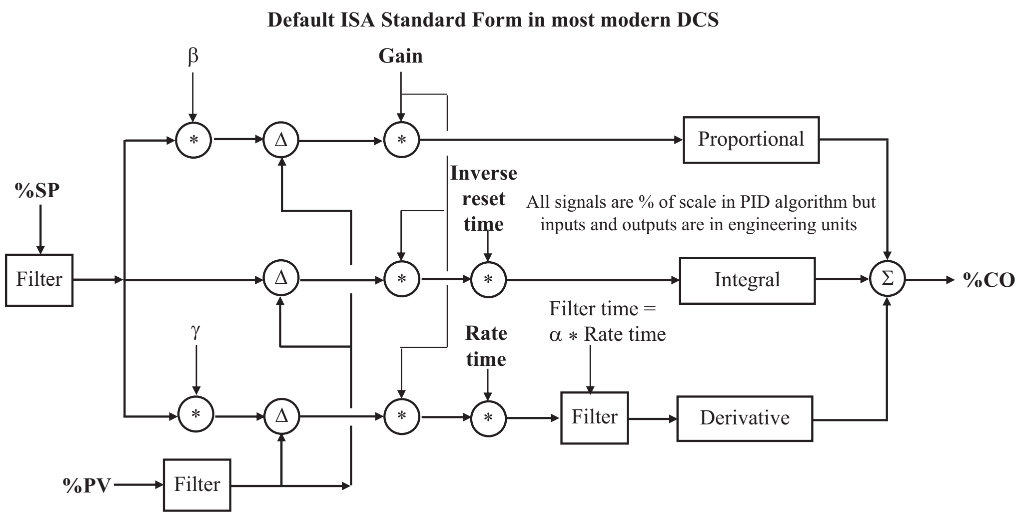


Figure K.1. ISA Standard Form has modes in parallel with proportional gain factor.

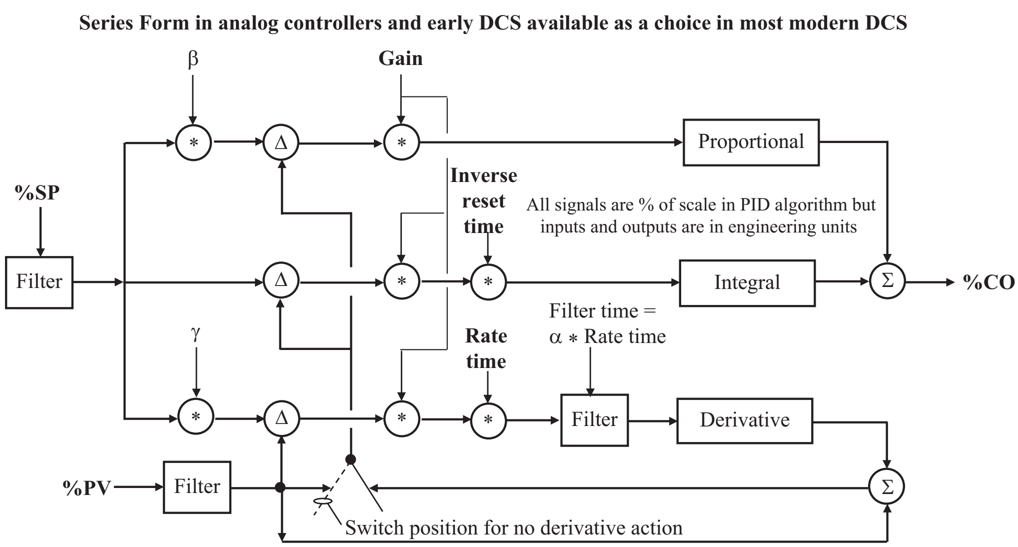


Figure K.2. Series Form has derivative mode computed in series with other modes.

Series Form inherently prevents this instability by increasing the effective reset time as the rate time is increased.

We can convert from ISA Standard Form to the Series Form using the following equations if the reset time is equal to or greater than four times the rate time ($T_i \geq 4 * T_d$).

$$K_c^{'} = \frac{K_c}{2} * \left[1 + \left(1 - 4 * \frac{T_d}{T_i} \right)^{0.5} \right] \tag{K.4}$$

$$T_i^{'} = \frac{T_i}{2} * \left[1 + \left(1 - 4 * \frac{T_d}{T_i} \right)^{0.5} \right] \tag{K.5}$$

$$T_d^{'} = \frac{T_i}{2} * \left[1 - \left(1 - 4 * \frac{T_d}{T_i} \right)^{0.5} \right] \tag{K.6}$$

The Parallel Form (Figure K.3) is depicted in many control theory text books but is rarely used in the process industries. The occasional use in some DCS and PLC to isolate the proportional mode tuning setting from the other modes is problematic in terms of creating dramatic differences in reset and rate settings causing significant confusion. The gain setting does not affect the contribution from the integral and derivative modes. Sometimes the integral and derivative mode settings use integral gain and derivative gain, respectively.

We can convert from the Parallel Form to the ISA Standard Form using the following equations:

$$K_c = K_c^{''} \tag{K.7}$$

If the parallel form integral mode tuning parameter is an integral *time:*

$$T_i = K_c^{''} * T_i^{''} \tag{K.8}$$

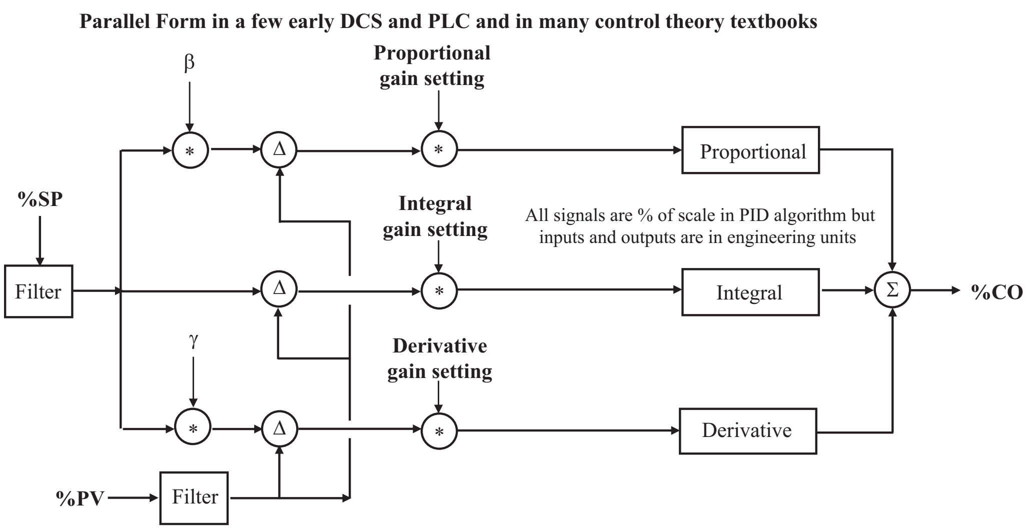


Figure K.3. Parallel Form has independent mode gain settings.

If the parallel form integral mode tuning parameter is an integral *gain*:

$$T_i = \frac{K_c''}{K_i''} \tag{K.8b}$$

If the parallel form derivative mode tuning parameter is a derivative *time*:

$$T_d = \frac{T_d''}{K_c''} \tag{K.9a}$$

If the parallel form derivative mode tuning parameter is a derivative *gain*:

$$T_d = \frac{K_d''}{K_c''} \tag{K.9b}$$

We can convert from the ISA Standard Form to the Parallel Form by using the following equations:

$$K_c'' = K_c \tag{K.10}$$

If the parallel form integral mode tuning parameter is an integral *time*:

$$T_i'' = \frac{T_i}{K_c} \tag{K.11a}$$

If the parallel form integral mode tuning parameter is an integral *gain*:

$$K_i'' = \frac{K_c}{T_i} \tag{K.11b}$$

If the parallel form derivative mode tuning parameter is a derivative *time*:

$$T_d'' = K_c * T_d \tag{K.12a}$$

If the parallel form derivative mode tuning parameter is a derivative *gain*:

$$K_d'' = K_c * T_d \tag{K.12b}$$

where

K_c =controller gain for ISA Standard Form (%/%) (dimensionless)
K_c' =controller gain for Series Form (%/%) (dimensionless)
K_c'' =controller gain for Parallel Form (%/%) (dimensionless)
K_i'' = integral gain for Parallel Form (1/seconds)
K_d'' = derivative gain for Parallel Form (seconds)
T_i = integral time (reset time) for ISA Standard Form (seconds)
T_i' = integral time (reset time) for Series Form (seconds)
T_i'' = integral time (reset time) for Parallel Form (seconds)
T_d = derivative time (rate time) for ISA Standard Form (reset time) (seconds)
T_d' = derivative time (rate time) for Series Form (seconds)
T_d'' = derivative time (rate time) for Parallel Form (seconds)

Note that the parallel form derivative *time* and derivative *gain* have the same value.

DIFFERENCE EQUATIONS FOR ISA STANDARD AND SERIES FORM

The difference equations for a reverse acting ISA Standard Form PID with rate limiting are as follows where the filter time on the derivative mode is a fraction α of the rate time (typical default value of PID alpha parameter is 0.125):

$$\%CO_n = P_n + I_n + D_n + \%CO_i \tag{K.13}$$

$$P_n = K_c * (\beta * \%SP_n - \%PV_n) \tag{K.14}$$

$$I_n = \frac{K_c}{T_i} * (\%SP_n - \%PV_n) * \Delta t_x + I_{n-1} \tag{K.15}$$

$$D_n = \frac{K_c * T_d * \left[\gamma * (\%SP_n - \%SP_{n-1}) - (\%PV_n - \%PV_{n-1})\right] + \alpha * T_d * D_{n-1}}{\alpha * T_d + \Delta t_x} \tag{K.16}$$

The difference equations for a reverse acting Series Form PID with rate limiting are as follows where the filter time on the derivative mode is a fraction α of the rate time (typical default value of PID alpha parameter is 0.125):

$$\%CO_n = PD'_n + ID'_n + \%CO_i \tag{K.17}$$

$$PD'_n = K'_c * (\beta * \%SP_n - \%PV_n - D'_n) \tag{K.18}$$

$$ID'_n = \frac{K'_c}{T'_i} * (\%SP_n - \%PV_n - D'_n) * \Delta t_x + ID'_{n-1} \tag{K.19}$$

$$D'_n = \frac{K'_c * T'_d * \left[\gamma * (\%SP_n - \%SP_{n-1}) - (\%PV_n - \%PV_{n-1})\right] + \alpha * T'_n * D'_{n-1}}{\alpha * T'_d + \Delta t_x} \tag{K.20}$$

If the structure used does not have integral action, the term $\%CO_i$ is an adjustable bias and integral term is zero (Equations K.15 and K.19) are not used.

POSITIVE FEEDBACK IMPLEMENTATION OF INTEGRAL MODE

The positive feedback implementation of the integral mode as shown in Figure K.4 effectively yields the equations above as seen in the following derivation for a PI controller using Laplace transforms. Instead of an integrator, a filter whose input is the controller output and whose output is added to the contribution from the proportional mode in the positive feedback implementation. The filter time is the integral time setting. When external reset feedback is enabled, the input to the filter is switched from the controller output to the external reset feedback signal.

The use of the positive feedback implementation of the integral mode is the basis for the enhanced PID developed for wireless that has been shown to be so effective for analyzer besides wireless devices especially when the update time is much larger than the process 63 percent response time.

$$O(s) = K_c * E(s) + \frac{1}{1 + T_i * s} * O(s) \tag{K.21}$$

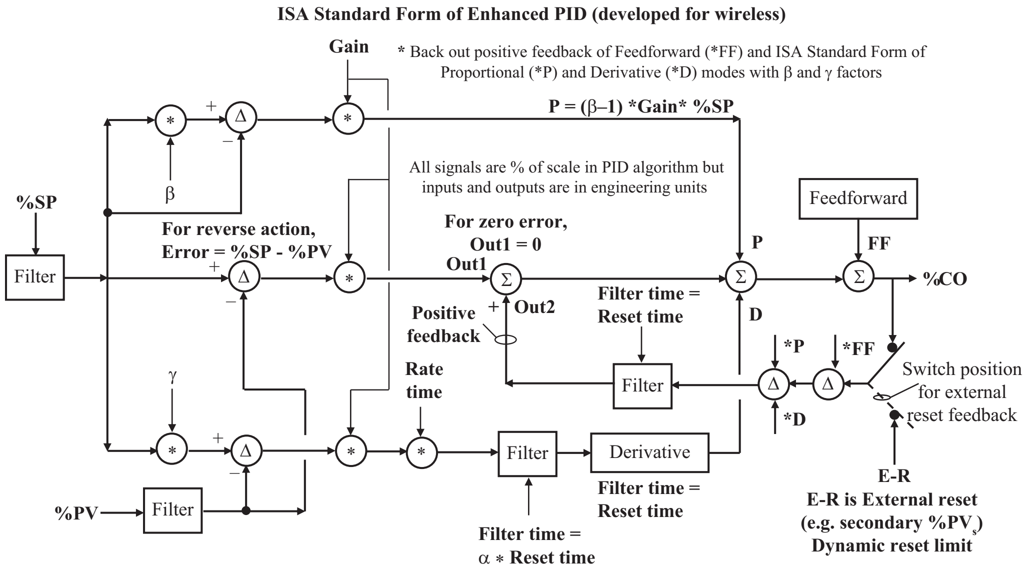


Figure K.4. ISA Standard Form with positive feedback implementation of integral mode and external reset feedback.

$$O(s) * \left(1 - \frac{1}{1 + T_i * s}\right) = K_c * E(s) \tag{K.22}$$

$$O(s) * \left(\frac{T_i * s}{1 + T_i * s}\right) = K_c * E(s) \tag{K.23}$$

$$\frac{O(s)}{E(s)} = K_c * \left(\frac{1 + T_i * s}{T_i * s}\right) = K_c + \frac{K_c}{T_i * s} \tag{K.24}$$

where

$\%CO_i$ =controller output at the transition from MAN or ROUT modes (%)
$\%CO_n$ =controller output for current scan n (%)
$\%CO_{n-1}$ =controller output for last scan n−1 (%)
$\%PV_n$ =process variable for current scan n (%)
$\%PV_{n-1}$ =process variable for last scan n−1 (%)
$\%SP_n$ =setpoint for current scan n (%)
$\%SP_{n-1}$ =setpoint for last scan n−1 (%)
K_c =controller gain for ISA Standard Form (%/%) (dimensionless)
K_c' =controller gain for Series Form (%/%) (dimensionless)
T_i = integral time (reset time) for ISA Standard Form (seconds)
T_i' = integral time (reset time) for Series Form (seconds)
T_d = derivative time (rate time) for ISA Standard Form (reset time) (seconds)
T_d' = derivative time (rate time) for Series Form (seconds)
P_n = proportional mode contribution for ISA Standard Form for current scan n (%)
P_{n-1} = proportional mode contribution for ISA Standard Form for last scan n−1 (%)
PD_n' = proportional-derivative mode contribution for Series Form for current scan n (%)

PD'_{n-1} =proportional-derivative mode contribution for Series Form for last scan n−1 (%)
I_n = proportional mode contribution for ISA Standard Form for current scan n (%)
I_{n-1} = proportional mode contribution for ISA Standard Form for last scan n−1 (%)
ID'_n = proportional-derivative mode contribution for Series Form for current scan n (%)
ID'_{n-1} =proportional-derivative mode contribution for Series Form for last scan n−1 (%)
D_n = derivative mode contribution for ISA Standard Form for current scan n (%)
D_{n-1} = derivative mode contribution for ISA Standard Form for last scan n-1 (%)
D'_n = derivative mode contribution for Series Form for current scan n (%)
D'_{n-1} =derivative mode contribution for Series Form for last scan n−1 (%)
Δt_x = execution time of PID block (seconds)
β=beta setpoint weight factor for proportional mode (0 to 1) (dimensionless)
γ=gamma setpoint weight factor for derivative mode (0 to 1) (dimensionless)
α = alpha factor for derivative mode filter (fraction of rate time) (0 to 1) (dimensionless)
$E(s)$ =Laplace transform of controller error (%)
$O(s)$ =Laplace transform of controller output (%)

APPENDIX L

Liquid Mixing Dynamics

The dead timer and time constant of tanks, static mixers, and pipelines for liquid composition control can be estimated by Equations L.1 and L.2 for an estimated or computed equivalent agitation flow (F_a). For static mixers and pipelines, the agitation flow for back mixing is minimal and rarely exceeds 25 percent of the feed flow (F_f). These volumes are characterized as plug flow and most of the residence time (volume divided by feed flow) becomes dead time. The lack of a process time constant translates to no slowing down of a fast disturbance. The saving grace is that the residence time of these systems is relatively short (e.g., three seconds) so that the dead time introduced is minimal and may be less than the electrode time constant or signal filter that provides some smoothing and prevents the control loop from becoming dead time dominant.

$$\theta_p = 0.5 * \frac{V}{F_f + F_r + F_a} \tag{L.1}$$

$$\tau_p = \frac{V}{F_f} - \theta_p \tag{L.2}$$

For axial agitation where liquid height is between 50 percent and 150 percent of vessel diameter:

$$F_a = 7.48 * N_q * N_s * D_t^3 \tag{L.3a}$$

$$N_q = \frac{0.4}{\left[\frac{D_i}{D_t}\right]^{0.55}} \tag{L.3b}$$

The volume cancels out in the ratio of the process dead time to 63 percent response time (dead time plus time constant) important for peak error per Equation C.3 in Appendix C if the recirculation plus agitation flow is much larger than the feed flow:

$$\frac{\theta_p}{\tau_p + \theta_p} = 0.5 * \frac{F_f}{F_f + F_r + F_a} \tag{L.4}$$

where

D_i = impeller diameter for axial (back) mixing (feet)
D_t = tank diameter for axial (back) mixing (feet)
N_s = impeller speed for axial (back) mixing (rpm)
N_q = impeller discharge coefficient for axial (back) mixing (gpm per rpm*ft^3)
F_a = agitation flow for axial (back) mixing (gpm)
F_f = total feed flow (gpm)
F_r = recirculation flow (gpm)
V = liquid volume (gallons)
θ_p = process dead time (minutes)
τ_p = process time constant (minutes)

For self-regulating processes, the peak error is proportional to the dead time to time constant ratio for a dead time much smaller than the time constant. If the recirculation flow and agitation flow is much greater than the feed flow, the volume cancels out in the ratio of Equation L.2 to L.1 giving Equation E.4, showing that the size of the recirculation and agitation flow relative to feed flow is important. Since the integrated (accumulated) error is proportional to the ratio of the dead time squared to time constant, the volume does not cancel out and will increase the amount of product off-spec. The recirculation flow may tend to provide more radial mixing than axial agitation unless the entry point is vertical instead of side entry. An educator can magnify the recirculation flow and improve the mixing pattern.

For pipelines, there is no recirculation flow and the agitation flow for back mixing is about 20 percent of the feed flow. For static mixers and laminar flow, the back mixing is less resulting in more of the residence time becoming dead time and making the volume more plug flow with only radial mixing (no axial mixing). Plug flow is desirable for reactions because the residence time is constant for each infinitesimal volume of reactants as touted in the paper "The Static Mixer as a Chemical Reactor" (Bor 1971). Plug flow is undesirable for in-line systems where reaction time is not a consideration (e.g., static mixers for blending, stock consistency, and neutralization) because there is no smoothing of axial fluctuations.

Equations L.1 and L.2 lead to some interesting conclusions about the use of well-mixed tanks for composition control. If the tank volume is increased but the feed, recirculation, and agitation flows are kept constant, the peak error is the same but the integrated (accumulated) error is increased in proportion to the increase in tank volume. If the tank volume is made exceptionally large, the dead time may be so large that the minimum reset time setting may cause a reset cycle. If the recirculation and agitation flows were increased in proportion to the tank volume, the peak error would be reduced and the integrated error would be about the same as before the increase in volume and recirculation and agitation flows. The increased tank volume will serve to slow down upstream disturbances and attenuate upstream oscillations for downstream control systems. Thus, exceptionally large tanks improve composition control performance if the loops are upstream or downstream but not on the tank, unless the agitation flow is scaled up with the volume. If the large tank is upstream of the loop, the reagent or coolant usage will be reduced when there are short term fluctuations in reagent or cooling demand.

APPENDIX M

Measurement Speed Requirements for SIS

As this is a book on control loops, the speed of a Safety Instrumented System (SIS) is so important that some guidance is needed. The same methodology used to get the time constants and dead times for control loop analysis can provide the data necessary to check whether an interlock system is too slow. Designs to date have concentrated on redundancy to make interlocks safer, but they have largely ignored the dynamic requirements due to a lack of an approach to an assessment of the performance problem. The following method consists of estimating the allowable time delay in actuation of an interlock and checking Tables M.1 through M.4 to see if the instrumentation is slow or OK. The response of various sizes of a bare thermocouple (TC) and resistance temperature detector (RTD) sensor elements are compared. Note that to provide a safer operation and removal, temperature sensors in corrosive and hot streams are installed in a thermowell which has a much larger measurement lag time than a bare element. Process and installation conditions can cause a measurement to be too slow even if the rating in the table is OK. When speed of response is critical, the 98 percent response time of the installed measurement should be measured in the plant.

All these tables assume minimum damping setting, no signal filtering, and no digital update delay. Tables M.5a through M.5c offer typical digital delays for various models of smart transmitters and various types of communication.

$$\theta_m < \frac{\left|\%PV_{unsafe} - \%PV_{trip}\right|}{\Delta\%PV_{\max} \big/ \Delta t} \tag{M.1}$$

where

$\%PV_{unsafe}$ = unsafe value of process variable (%)
$\%PV_{trip}$ = trip value of process variable (%)
$\%PV_{\max} / \Delta t$ = maximum rate of change of process variable (%/sec)
θ_m = Allowable measurement delay (sec)

Table M.1. Temperature measurements speed ratings for given allowable SIS delay (Table assumes no signal filtering of measurement and no digital update delay)

Time delay (sec)	Thermowell* (>50 fps gas)	Thermowell* (>5 fps liquid)	Bare ¼″ RTD	Bare ¼″ TC	Bare ⅛″ RTD	Bare ⅛″ TC	Optical Pyrometer†
180	OK	OK	OK	OK	OK	OK	OK
120	OK	OK	OK	OK	OK	OK	OK
60	Slow	OK	OK	OK	OK	OK	OK
6	Slow	Slow	Slow	OK	OK	OK	OK
4	Slow	Slow	Slow	Slow	Slow	OK	OK
2	Slow	Slow	Slow	Slow	Slow	Slow	OK
0.5	Slow	Slow	Slow	Slow	Slow	Slow	OK

*TC and RTD are single element sheathed with tight fit (less than 0.05 inch annular clearance).
†No noise reduction technique in optical pyrometer software that would introduce time delay.

Table M.2. Pressure measurements speed ratings for given allowable SIS delay

Time Delay (sec)	Capacitance	Resonant wire	Diffused silicon	Strain gauge
0.4	OK	OK	OK	OK
0.2	Slow	Slow	OK	OK
0.05	Slow	Slow	Slow	OK

Table M.3. Flow measurements speed ratings for given allowable SIS delay

Time delay (sec)	Doppler meter	Vortex meter	Mag meter DC	Mag meter AC	Coriolis meter U-tube	Turbine or swirl meter
6	OK	OK	OK	OK	OK	OK
4	Slow	OK	OK	OK	OK	OK
2	Slow	Slow	OK	OK	OK	OK
1	Slow	Slow	Slow	OK	OK	OK
0.05	Slow	Slow	Slow	Slow	OK	OK

Table M.4. Level measurement speed ratings for given allowable SIS delay

Time delay (sec)	Ultrasonic	Nuclear	Admittance	D/P
20	OK	OK	OK	OK
10	Slow	OK	OK	OK
5	Slow	Slow	OK	OK
1	Slow	Slow	Slow	OK

Table M.5a. Digital time delays of smart DP transmitter (Rosemount 3051C)

Output signal	Inherent dead time	Signaling dead time	Total dead time
4–20 mA	70 ms	0 ms	70 ms
Digital	70 ms	675 ms	745 ms
Field Bus	70 ms	100 ms	170 ms

Table M.5b. Digital time delays of smart DP transmitter (Rosemount 1151S)

Output signal	Inherent dead time	Signaling dead time	Total dead time
4–20 mA	95 ms	0 ms	95 ms
Digital	95 ms	675 ms	770 ms
Field Bus	95 ms	100 ms	195 ms

Table M.5c. Digital time delays of smart DP transmitter (Honeywell ST 3000)

Output signal	Inherent dead time	Signaling dead time	Total dead time
4–20 mA	250 ms	150 ms	400 ms
Digital	250 ms	670 ms	920 ms
Field Bus	250 ms	100 ms	350 ms

References

Bergeron, L. 1961. *Water Hammer in Hydraulics*. New York: Wiley.

Blevins, T., G. McMillan, W. Wojsznis, and M. Brown. 2003. *Advanced Control Unleashed*. ISA.

Bohl, A., and T. McAvoy. 1976. "Linear Feedback vs. Time Optimal Control II - The Regulator Problem." *Industrial & Engineering Chemistry Process Design and Development* 15, no. 1.

Bor, T. 1971. "The Static Mixer as a Chemical Reactor." *British Chemical Engineering* 16, no. 7.

Crawford, C. L. January, 1982. "Thermowell Heat Conduction Error Versus" Texas A&M Symposium, Appendix I, p. 2.

Harriott, P. 1964. *Process Control*. McGraw-Hill.

Kardos, P. August 29, 1977. "Response of Temperature Measuring Elements." *Chemical Engineering*.

Kreith, F. 1965. *Principles of Heat Transfer*, 376. 2nd ed. International Text-book Company.

Luyben, W., and M. Melcic. 1978. "Consider Reactor Control Lags." *Hydrocarbon Processing*. 57, no. 3.

McMillan, G. 1999. *Process/Industrial Instruments and Controls Handbook*. 5th ed. McGraw-Hill.

McMillan, G. 2004. *Advanced pH Measurement and Control*. 3rd ed. McGraw-Hill. International Society of Automation

McMillan, G. 2006. "Full Throttle Batch and Startup Response." *Control.*

McMillan, G. March–April, 2011. "Feedforward Control Enables Flexible, Sustainable Manufacturing." *InTech*.

McMillan, G. January–February, 2012. "PID Tuning Rules." *InTech*.

Shinskey, G. 1994. *Feedback Controllers for the Process Industries*. McGraw-Hill., 1994.

Shinskey, G. 1996. *Process Control Systems*. 4th ed. McGraw-Hill.

Skogestad, S. 2003. "Simple Analytic Rules for Model Reduction and PID Controller Tuning." *Journal of Process Control* 13, no. 4.

Vilanova, R., and A. Visioli, eds. 2012. *PID Control in the Third Millennium: Lessons Learned and New Approaches*. London: Springer.

Ziegler, J., and N. Nichols. 1943. "Process Lags in Automatic Control Circuits." *ASME Transactions* 65, no. 5.

Bibliography

Al-Khalifah, M., and G. McMillan. 2012. Control Valves versus Variable Speed Drive for Flow Control. *ISA Automation Week*. Orlando, FL.

Arant, J. November, 1993. "Control Valve Notes and Positioner Myths - Part I." *Control.*

Arant, J. December, 1993. "Control Valve Notes and Positioner Myths - Part II." *Control.*

Astrom K., and T. Hagglund. 2006. *Advanced PID Control*. ISA.

Bailey, J., and D. Ollis. 1977. *Bio chemical Engineering Fundamentals*. New York: McGraw-Hill.

Bialkowski, W. 1990. "Process Control Related Variability and the Link to End-Use Performance." *Control Systems 90*. Helsinki, Finland.

Bialkowski, W., and B. Janvier. 2005. On Lambda Tuning—How and Why. *PAPTAC 91st Annual Meeting*.

Boudreau, M., and G. McMillan. 2006. *New Directions in Bioprocess Modeling and Control*. ISA.

Bradner, M. July, 1949. "Pneumatic Transmission Lag." *Instruments* 22.

Brez, C. 1975. "Predicting Limit Cycling in Control Valves." *Control Engineering* 22, no. 7.

Buchwald, W. October, 1974. Control Performance with Reduced Valve Pressure Drop for Energy Conservation. Advances in Instrumentation. *ISA International Conference*, Vol. 29, Part 2, Paper 610. New York.

Buckley, P. 1979. "Dynamics of Temperature Measurement." In Vol. 34 of *Process Annual Symposium*. Wilmington, DE: Instrumentation Process Industry.

Coughran, M. October, 1994. "Valves: Testing for Peak Performance." *InTech* 41, no. 10.

Coughran, M. August, 1996. "Performance Influences in Globe Control Valves." *InTech* 43, no. 8.

Coughran, M. 2014. "Lambda Tuning—the Universal Method for PID Controllers in Process Control." White Paper, *Control*.

Fitzgerald, B. 1995. *Control Valves for the Chemical Process Industries*. McGraw-Hill Inc.

Giusti, A., and J. Hougen. 1961. "Dynamics of pH Electrodes." *Control Engineering* 8, no. 4.

Green, R., and W. Field. June, 1964. Evaluation of pH Measurement Response. *AID Symposium Proceedings*. San Francisco, CA.

Hagglund, T. 1995. "A Control-Loop Performance Monitor." *Control Engineering Practice* 3, no. 11. doi: http://dx.doi.org/10.1016/0967-0661(95)00164-P.

Heider, R. June, 1982. How Digital Controllers Can Cause Process Cycling. *ACC Conference Proceedings*. Arlington, VA.

Heider, R., and L. Clay. January, 2009. "Make the Most of Variable Frequency Drives." *Control.*

Hopkins B., and G. Alford. May, 1993. "Temperature Control of Polymerization Reactor." *Instrumentation Technology*.

Keith, G. November, 1977. "Linear Installed Characteristic Simplify Control Valve Selection." *Control Engineering*.

Kerlin, T., R. Shepard, H. Hashemian, and K. Petersen. 1982. Time Response of Installed Temperature Sensors. *Sixth International Symposium on Temperature*. Paper 19-2.

Latour, L., J. Sharpe, and M. Delaney. 1986. "Estimating Benefits from Advanced Control." *ISA Transactions* 25, no.4.

Liptak, B. 2010. *Instrument Engineers Handbook*. Chilton.

Luyben, W. 2007. *Chemical Reactor Design and Control*. Wiley.

Luyben, W., and Y. Cheng-Ching. 1986. "Analysis of Valve-Position Control for Dual Input Processes." *Industrial & Engineering Chemistry Fundamentals* 25, no. 3. doi: http://dx.doi.org/10.1021/i100023a007

Luyben, W., and T. Cheung. October, 1978. Proportional-Derivative Control of Exothermic Reactors. *Joint Automatic Control Conference*, Vol. 3. Philadelphia, PA.

Luyben, W., and M. Luyben. 1997. *Essentials of Process Control*. McGraw Hill.

Luyben, W., and B. Tyreus. 1993. "Dynamics and Control of Recycle Systems. 4. Terneary Systes with One or Two Recycle Streams." *Industrial & Engineering Chemistry Research* 32, no. 6. doi: http://dx.doi.org/10.1021/ie00018a021

Mamzic, C. August, 1958. "Improving the Dynamics of Pneumatic Positioners." *ISA Journal* 5, no. 8.

McMillan, G. October, 1981. The Effect of Process Non-Self-Regulation on Control Loop Performance. *ISA Conference Proceedings*. Anaheim, CA.

McMillan, G. June, 1982. The Effect of Cascade Control on Loop Performance. *ACC Conference Proceedings*. Arlington, TX.

McMillan, G. 1999. *Process/Industrial Instruments and Controls Handbook*. 5th ed. McGraw Hill.

McMillan, G. 2004. *Advanced pH Measurement and Control*. 3rd ed. International Society of Automation.

McMillan, G. May, 2004. "What's Your Flow Control Valve Telling You?" *Control Design*.

McMillan, G. July, 2008. "Unlocking the Secret Profiles of Batch Reactors." *Control*.

McMillan, G. May, 2009. "Is Wireless Process Control Ready for Prime Time." *Control.*

McMillan, G. 2010. *Centrifugal and Axial Compressor Control*. Momentum Press.

McMillan, G. 2010. *Essentials of Modern Measurements and Final Elements in the Process Industries.* ISA.

McMillan, G. 2010. *Process Control Case Histories*. Momentum Press.

McMillan, G. January–February, 2010. "Advances in Flow and Level Measurements Enhance Process Knowledge, Control." *InTech* 57, no. 1.

McMillan, G. July–August, 2010. "Wireless - Overcoming Challenges of PID Control, Analyzer Applications." *InTech*.

McMillan, G. November–December, 2010. "Temperature Measurement, Control Key to Plant Performance." *InTech* 57, no. 6.

McMillan, G. 2011. *Advanced Temperature Measurement and Control*. 2nd ed. ISA.

McMillan, G. 2011. "Corn Analysis Modeling and Control for Ethanol Yield Improvement." *ISA Automation Week*. Mobile, AL: ISA.

McMillan, G. 2011. "pH Measurement and Control Opportunities." *ISA Automation Week*. Mobile, AL: ISA.

McMillan, G. 2011. "Reactor Control Opportunities." *ISA Automation Week*. Mobile, AL: ISA.

McMillan, G. 2011. "Wireless Measurement and Control Opportunities." *ISA Automation Week*. Mobile, AL: ISA.

McMillan, G. November, 2011. "Don't Overlook PID in APC." *Control*.

McMillan, G. 2012. *Good Tuning: A Pocket Guide.* 3rd ed. ISA.

McMillan, G. September, 2012. "Get the Most Out of Your Batch." *Control* 25, no. 9, pp. 51–55.

McMillan, G. February, 2013. "Maximize pH Response, Accuracy and Reliability." *Control* 26, no. 2.

McMillan, G., and R. Baril. August, 2010. "pH Measurement and Control." *Chemical Engineering*.

McMillan, G., T. Benton, Y. Zhang, and M. Boudreau. March, 2008. "PAT Tools for Accelerated Process Development and Improvement." Process Design Supplement, *BioProcess International.*

McMillan, G., and M. Boudreau. 2007. *New Directions in Bioprocess Modeling and Control*. ISA.

McMillan, G., M. Boudreau, and G. Wilson. November, 2006. "Maximizing Process Benefits form Bioprocess Modeling and Control." *Pharmaceutical Technology.*

McMillan, G., R. Reedy, and J. Moulis. December, 2009. "Virtual Plant Provides Real Insights." *Chemical Processing*.

McMillan, G., and M. Sowell. November, 2007. "Virtual Control of Real pH." *Control*.

McMillan G., and H. Torres. 2012. "Effective Use of Key PID Features." *ISA Automation Week*. Orlando.

McMillan, G., and H. Vegas. 2012. 101 Tips for a Successful Automation Career. ISA.

Moore, R. December, 1990. "Good pH Measurements in Bad Process Streams." *Instruments and Control Systems* 63, no. 12, pp. 39–43.

Nelson, D. March, 2004. "Better Damper Control Nixes NOx." *Chemical Processing* 67, no. 1, pp. 32–37.

Nino, S. November, 2012. Improve Material Balance in High-Purity Distillation Control. *Hydrocarbon Processing*, Process Control and Information Systems Bonus Report.

O'Dwyer, A. 2006. *Handbook of PI and PID Controller Tuning Rules*. 2nd ed. Imperial College Press.

Rinehart, N.,and N. Ingram. 2004. "Smart Control Valve Diagnostics: Predictive Maintenance and Beyond." *Valve Magazine* 16, no. 4.

Ruel, M. November, 2000. "Sticktion: The Hidden Menace." *Control* 13, no. 11.

Schork, E. et al. 1993. *Control of Polymerization Reactors*. Marcel Dekker Inc.

Schuder, C. October, 1971. "Control Valve Rangeability and the Use of Valve Positioners." *Advances in Instrumentation, ISA 26th Annual Conference*, Vol. 26, Part 4, Paper 817. Chicago.

Shinskey, G. 1973. *pH and pION Control in Process and Waste Streams*. McGraw-Hill.

Shinskey, G. 1978. *Energy Conservation Through Control*. McGraw-Hill.

Shinskey, G. 1981. *Controlling Multivariable Processes*. ISA.

Shinskey, G. June, 1993. "The Effect of Scan Period on Digital Control Loops." *InTech*.

Shinskey, G. 1996. *Process Control Systems*. 4th ed. McGraw-Hill.

Shinskey, G. 2001. "PID-Deadtime Control of Distributed Processes." *Control Engineering Practice* 9, no. 11, pp. 1177–1183. 2001

Shinskey, G. May, 2006. "The Power of External-Reset Feedback." *Control*.

Shinskey, G. May, 2009. "Multivariable Control of Distillation." *Control*.

Shinskey, G. May, 2011. "Meditating on Disturbance Dynamics." *Control* 24, no. 5, pp. 51–55.

Skogestad, S., and J. Downs. April, 2011. "An Industrial and Academic Perspective on Plantwide Control." *Annual Reviews in Control* 35, no. 1, pp. 99–110. Elsevier

Skogestad, S., and M. Shamsuzzoha. 2010. "The setpoint overshoot method: A Simple and Fast Closed-Loop Approach for PID Tuning." *Journal of Process Control* 20, no. 10, pp. 1220–1234. Elsevier

Smith, C. 2008. "Watch Out with Variable Speed Pumping." *Chemical Processing* 75, no. 5, p. 33.

Smith, C. 2009. *Practical Process Control – Tuning and Troubleshooting.* Wiley.

Smuts, J. 2011. *Process Control for Practitioners*. OptiControls.

Spitzer, D. 2008. *Variable Speed Drives: Principles and Applications for Energy Cost Savings*. 3rd ed. ISA.

Spitzer, D. September, 2008. "Getting up to Smart Speed." *Control*.

St. Clair, D. 1989. *Controller Tuning & Control Loop Performance – A Primer*. Straight-Line Control Company.

Trevathan, V. 2006. *A Guide to the Automation Book of Knowledge*. ISA.

Ziegler, J. and N. Nichols. 1942. "Optimum Settings for Automatic Controllers." *ASME Transactions* 64, no. 11.

About the Author

Greg is a retired Senior Fellow from Solutia Inc. During his 33 year career with Monsanto Company and its spin off Solutia Inc., he specialized in improving loop performance, controller tuning, valve dynamics, opportunity assessments, dynamic simulation, fermentor control, pH control, and reactor control. Greg is the author of more than 80 articles and papers and 20 funny and serious books. Greg's most recent humorous books are *The Funnier Side of Retirement for Engineers and People of the Technical Persuasion* and *The Life and Times of an Automation Professional*. His latest technical books are *New Directions in Bioprocess Modeling and Control*, *Essentials of Modern Measurements and Final Elements in the Process Industry*, *Advanced Temperature Measurement and Control* (2nd edition), and *101 Tips for a Successful Automation Career.* He is also the author of the ISA best sellers titled *Advanced Control Unleashed*, *Good Tuning: A Pocket Guide* (3rd edition), *Advanced Control Unleashed*, and *The Funnier Side of Retirement for Engineers and People of the Technical Persuasion.* He has contributed to several handbooks, is the editor of the *Process/Industrial Instrumentation and Controls* handbook, and has written a monthly "Control Talk" column for *Control* and has written a dozen articles for *Control* and *InTech* since 2002. Greg posts blogs on the Control Global site and tips on the International Society of Automation (ISA) Interchange site.

As an Affiliate Professor of Washington University in Saint Louis, Missouri from 2001 to 2003, he developed a virtual plant and course on modeling and control as practiced in the process industry for chemical engineers using the book *Advanced Control Unleashed*.

Greg is an ISA Fellow and received the ISA Kermit Fischer Environmental Award for pH control in 1991, the *Control* magazine's Engineer of the Year Award for the Process Industry in 1994, and was one of the first inductees into the *Control* magazine's Process Automation Hall of Fame in 2001. Greg was honored as one of InTech's 50 most influential innovators for advancing automation and control technologies and received the ISA Life Time Achievement Award in 2010. He received a BS from Kansas University in 1969 in Engineering Physics and an MS from Missouri University of Science and Technology (Rolla) in 1976 in Electrical Engineering (Control Theory). Greg contracts as a consultant at Emerson Process Management through CDI Process and Industrial and is a part time consultant at MYNAH Technologies. Greg was the program committee chair for ISA Automation Week 2011.

Index

Q

R

*Only the first occurrence of these terms in a chapter are listed here.